

Convolutional neural network for simultaneous prediction of several soil properties using visible/near-infrared, mid-infrared, and their combined spectra



Wartini Ng^{a,*}, Budiman Minasny^a, Maryam Montazerolghaem^b, Jose Padarian^a, Richard Ferguson^c, Scarlett Bailey^c, Alex B. McBratney^a

^a School of Life and Environmental Sciences & Sydney Institute of Agriculture, The University of Sydney, NSW 2006, Australia

^b Sydney Informatics Hub, The University of Sydney, NSW 2006, Australia

^c USDA-NRCS, Kellogg Soil Survey Laboratory, National Soil Survey Center, 100 Centennial Mall North Lincoln, NE 68508-3866, United States of America

ARTICLE INFO

Handling Editor: Morgan Cristine L.S.

Keywords:

Deep learning
Multi-task learning
Near-infrared spectroscopy
Mid-infrared spectroscopy

ABSTRACT

No single instrument can characterize all soil properties because soil is a complex material. With the advancement of technology, laboratories have become equipped with various spectrometers. By fusing output from different spectrometers, better prediction outcomes are expected than using any single spectrometer alone. In this study, model performance from a single spectrometer (visible-near-infrared spectroscopy, vis-NIR or mid-infrared spectroscopy, MIR) was compared to the combined spectrometers (vis-NIR and MIR). We selected a total of 14,594 samples from the Kellogg Soil Survey Laboratory (KSSL) database that had both vis-NIR and MIR spectra along with measurements of sand, clay, total C (TC) content, organic C (OC) content, cation exchange capacity (CEC), and pH. The dataset was randomly split into 75% training ($n = 10,946$) and the remaining ($n = 3,648$) as a test set. Prediction models were constructed with partial least squares regression (PLSR) and Cubist tree model. Additionally, we explored the use of a deep learning model, the convolutional neural network (CNN). We investigated various ways to feed spectral data to the CNN, either as one-dimensional (1D) data (as a spectrum) or as two-dimensional (2D) data (as a spectrogram). Compared to the PLSR model, we found that the CNN model provides an average improvement prediction of 33–42% using vis-NIR and 30–43% using MIR spectral data input. The relative accuracy improvement of CNN, when compared to the Cubist regression tree model, ranged between 22 and 36% with vis-NIR and 16–27% with MIR spectral data input. Various methods to fuse the vis-NIR and MIR spectral data were explored. We compared the performance of spectral concatenation (for PLSR and Cubist model), two-channel input method, and outer product analysis (OPA) method (for CNN model). We found that the performance of two-channel 1D CNN model was the best ($R^2 = 0.95\text{--}0.98$) followed closely by the OPA with CNN ($R^2 = 0.93\text{--}0.98$), Cubist model with spectral concatenation ($R^2 = 0.91\text{--}0.97$), two-channel 2D CNN model ($R^2 = 0.90\text{--}0.95$) and PLSR with spectral concatenation ($R^2 = 0.87\text{--}0.95$). Chemometric analysis of spectroscopy data relied on spectral pre-processing methods: such as spectral trimming, baseline correction, smoothing, and normalization before being fed into the model. CNN achieved higher performance than the PLSR and Cubist model without utilizing the pre-processed spectral data. We also found that the predictions using the CNN model retained similar correlations to the actual values in comparison to other models. By doing sensitivity analysis, we identified the important spectral wavelengths variables used by the CNN model to predict various soil properties. CNN is an effective model for modelling soil properties from a large spectral library.

1. Introduction

In recent years, infrared spectroscopy has been used extensively to

characterize soil because of its advantages, mainly speed, in comparison to conventional analytical techniques. Conventional soil analysis is tedious and time-consuming, and its high cost makes it unattractive to

* Corresponding author at: Sydney Institute of Agriculture & School of Life and Environmental Sciences, The University of Sydney, Suite 105, Biomedical Building, 1 Central Avenue, Australian Technology Park, Eveleigh, NSW 2015, Australia.

E-mail address: wartini.ng@sydney.edu.au (W. Ng).

analyse a large enough number of samples to accurately characterize large areas. Diffuse reflectance spectroscopy in the visible-near-infrared (vis-NIR; 400–700–2,500 nm) and mid-infrared (MIR; 2,500–25,000 nm) is a promising technology as it is rapid, cost-effective, and non-destructive. With the help of multivariate statistics, multiple soil properties can be derived from the same spectral data (Islam et al., 2003; Nocita et al., 2015). Various soil properties have been predicted successfully, such as organic matter content, total nitrogen content, pH, cation exchange capacity (CEC) and soil texture components (Chang et al., 2001; Shepherd and Walsh, 2002; Islam et al., 2003; Stenberg et al., 2010).

Every spectrometer has its strengths and weaknesses, and information contained in a specific spectral range does not contain enough information to predict all soil properties. One plausible way to improve the prediction of various soil properties is to fuse the data from multiple spectrometers. Data fusion is a process of integrating data from multiple sources. Data fusion can be divided into various levels from low-, mid- to high-level data fusion (Borras et al., 2015). Low-level data fusion simply concatenates data from multiple instruments at the measurement stage. In the mid-level, data fusion corresponds to the fusion of feature extracted from the original instruments (e.g., retaining scores from the principal component analysis), while high-level data fusion occurs at the decision level where the different outputs from each instrument are combined to achieve the final predictions.

Examples of low data fusion are described in Wang et al. (2015) and Terra et al. (2019). Wang et al. (2015) fused data from an x-ray fluorescence (XRF) and a vis-NIR spectrometer to improve the prediction of total carbon and total nitrogen using random forest regression and penalized spline regression. Terra et al. (2019) fused the vis-NIR and MIR spectral data using the outer product analysis (OPA) method to improve the soil organic carbon predictions. O'Rourke et al. (2016) demonstrated high-level data fusion where vis-NIR and XRF spectroscopy model ensemble techniques were used to determine soil geochemical properties. In this study, we will be mainly focusing on low-level data fusion.

Each digital spectrum contains thousands of wavelength variables. Various chemometric methods, such as support vector machine (Devos et al., 2009), artificial neural network (Daniel et al., 2003), random forests (Lee et al., 2013), Cubist regression tree, and partial least squares regression (McCarty et al., 2002) had been explored to assess the capability of handling large number of input variables. Methods such as variable selection approaches have also been explored (Xuemei et al., 2010; Zou et al., 2010; Sarathjith et al., 2016). The performance of these models depends on the use of spectral pre-processing techniques.

Recently, Padarian et al. (2019) demonstrated the successful use of the convolutional neural network (CNN) for soil property prediction without the use of pre-processed spectra. CNN is one of the most popular learning architectures used for high dimensional data. CNN had been proven to be very successful in ImageNet Large Scale Visual Recognition Challenge (ILSVRC) which evaluates the best algorithms for object detection and image classification at large scale (approximately 10,000,000 images). Successful applications of the convolutional neural network have included object classification (Krizhevsky et al., 2012), object detection (Sermanet et al., 2013), facial expression recognition (Matsugu et al., 2003) and age prediction (Niu et al., 2016) among other applications. Padarian et al. (2019) proposed the representation of raw spectral data as a two-dimensional (2D) spectrogram and showed its superior performance over conventional techniques such as partial least squares regression (PLSR). A spectrogram is generated by decomposing the spectral signal into overlapping signals using short-time fast Fourier transformation (Griffin and Lim, 1984).

To improve the prediction accuracies of images using CNN model, the use of various multiple image channels (such as colour (grayscale, RGB, LUV), gradient magnitude and gradient histograms) have been proven to be successful (Dollar et al., 2009; Ribeiro et al., 2017).

Inspired by these works, we explore the use of two-channel input to fuse the spectral data. The use of outer product analysis (OPA), another data fusion method that combines two or more spectral data by calculating an outer-product matrix from each pair of spectral data (Jaillais et al., 2005; Vesela et al., 2007; Terra et al., 2019), will also be explored. This method considers all possible combinations of the two spectral variables.

In this study, we compared the performance of the CNN model using raw spectral data (vis-NIR or MIR) as one-dimensional (1D) and as a two-dimensional (2D) spectrogram. We also evaluated the performance of CNN in fusing two spectral data (vis-NIR and MIR) using a two-channel input method and OPA. The performance of these CNN models was then compared to the commonly used regression algorithms: PLSR and Cubist. We analysed how well the CNN model retained the correlation of the predicted properties and identified important spectral wavelengths used by the CNN model to predict various soil properties.

2. Materials and methods

2.1. Dataset

The soil samples used in this study were obtained from the Kellogg Soil Survey Laboratory (KSSL) database. This dataset contained measurements of > 17,000 pedons from the USA with well-documented and precise standard operating procedures. This large spectral database contained measurements using visible-near (vis-NIR) and mid-infrared (MIR) spectra. Such a large soil and spectral database with highly reliable soil information enabled the testing of new models such as CNN using vis-NIR, MIR, and fused vis-NIR-MIR spectral data input.

The detailed measurement methods were described in the soil survey laboratory methods manual (Soil Survey Staff, 2014). Properties used in this study were sand, clay, total C (TC) content, organic C (OC) content, cation exchange capacity (CEC), and pH in H₂O. These properties had been shown in the literature to be predicted well by the vis-NIR and MIR spectra (Janik et al., 1998; Chang et al., 2001; Shibusawa et al., 2001; McCarty et al., 2002; Shepherd and Walsh, 2002).

The wide range and non-normal distribution of the soil properties were expected for large area coverage database. Samples from histosols and andisols were not used in this study as they behaved differently from mineral soils. Only samples that had a complete measurement of the soil properties mentioned above, along with both vis-NIR and MIR spectral data were utilized. In addition, this study only focused on samples with OC content < 10%, common for agricultural soils and excluding organic soils. This resulted in a total of 14,594 samples from various soil depths. Prior to any model development, the dataset was randomly divided into a training set (75%; 10,946 samples) and a test set (25%; 3648 samples). The summary statistics of the soil properties are given in Table 1. The spatial distribution of the observations is included in Supplementary Material Fig. S1. Soil properties that are highly skewed (see Table 1) compared with others are subjected to log transformation prior to PLSR and Cubist modelling.

2.2. Spectral data acquisition

All the samples were air-dried, ground, and sieved to < 2-mm. Two spectral measurements (vis-NIR and MIR) were collected separately for each sample. Vis-NIR spectra were obtained using a LabSpec 2500 spectrometer (Analytical Spectral Devices, Boulder, CO) in the spectral range from 350 to 2500 nm with 1 nm increments. Soils were scanned from below using a Muglight (Analytical Spectral Devices, Boulder, CO), and Spectralon (LabSphere, North Sutton, NH) was scanned as the white reference standard every 15 min.

For the MIR spectra, samples were scanned with Vertex 70 (Bruker Optics, Ettlingen, Germany). The instrument covered a spectral range between 7498 and 600 cm⁻¹ with a resolution of 4 cm⁻¹. The spectrometer used an MCT (mercury cadmium telluride) detector cooled by

Table 1
Summary statistics of soil properties.

Dataset	Soil property					
	TC (%)	OC (%)	CEC (cmol(+) / kg)	Clay (%)	Sand (%)	pH
Training	Min.	0.03	0.60	0	0.1	2.86
	1st Qu.	0.46	0.29	7.40	9.9	5.06
	Median	1.13	0.77	14.70	20.7	5.92
	Mean	1.63	1.25	15.94	22.61	6.11
	3rd Qu.	2.05	1.59	21.80	32.5	7.42
	Max.	9.99	9.95	134.10	92.9	10.32
	Skewness	2	2.43	1.3	0.75	0.12
Test	Min.	0.03	0.03	0.60	0.00	3.29
	1st Qu.	0.42	0.27	7.40	10.20	5.01
	Median	1.09	0.75	14.50	21.10	5.93
	Mean	1.58	1.22	15.81	22.65	6.10
	3rd Qu.	2.02	1.58	21.80	32.53	7.40
	Max.	9.99	9.86	78.60	96.10	100
	Skewness	1.92	2.37	1.11	0.78	0.53

liquid nitrogen. The samples were further ground to 180 µm before being loaded into a 96 well plate. Wells were 6 mm wide and 1.3 mm deep. Every sample was loaded into four wells, resulting in four replicate scans; average spectra were used for modelling. An empty well with anodized aluminium at the bottom was scanned as a reference before every sample was scanned.

2.3. Model development

2.3.1. Partial Least Square Regression (PLSR)

PLSR is a linear chemometric regression model that projects spectral data into latent variables that explain the variances within the spectral data. Given predictors matrix \mathbf{X} and response variable \mathbf{Y} , the PLS algorithm decomposes \mathbf{X} and \mathbf{Y} into scores (\mathbf{T} and \mathbf{U}) of l latent components and loadings (\mathbf{P} and \mathbf{Q}) plus error matrix (\mathbf{F}) as follows:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{F}_x \quad \mathbf{Y} = \mathbf{UQ}' + \mathbf{F}_y \quad (1)$$

The relation in terms of original space can be expressed as:

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E} \quad (2)$$

where \mathbf{B} is the matrix of the regression coefficient, and \mathbf{E} is the matrix of residuals. The optimal number of latent variables (l) used in the PLS regression that results in the smallest root mean squared error (RMSE) using the cross-validation approach was used to create the linear models.

2.3.2. Cubist

Cubist is a rule-based regression model developed by Quinlan (1993). The Cubist model was introduced by Minasny and McBratney (2008) as an alternative data-mining tool for multivariate spectral calibration. The algorithm creates partitions of data with similar spectral characteristics and creates one or more rules for each partition. If the partition rules are satisfied, then the linear regression of that partition is used to create the prediction. Each rule can be defined as:

$$\text{if [condition is true], then [regression], else [apply next rule]} \quad (3)$$

as described by Quinlan (1993).

2.3.3. Convolutional neural network (CNN)

A typical CNN model would have an input layer, several hidden layers (convolution layers, pooling, and fully connected) and an output layer. The representation of the CNN architecture using vis-NIR spectral input as one-dimensional input is included in Fig. 1.

2.3.3.1. Input. The input layer is used to accept multi-dimensional raw data for processing in the network. It is usually specified with the width, height, and several channels. The number of channels is often set

to three if the input data are images to account for the colour channels (red, green, and blue). If the input data is one-dimensional, the number of channels is set to one. In this study, we evaluate two types of input: one-dimensional (as spectral data) and two-dimensional (as a spectrogram).

2.3.3.2. Convolution. The convolution layer is usually used as the layer to extract features from the input feature map. The convolution layer consists of a filter that slides over the input. At each location, dot products between the filter and the corresponding receptive field are performed to produce an output feature map (convolved feature). The size of the output feature map depends on:

- Filter size (k). Filter is a weight matrix used for features detection.
- Depth. Depth corresponds to the number of filter used in the network.
- Strides (s). Strides is the number of pixel shifts over the input feature map. If the stride is one, then the filters is moved one pixel at a time.
- Padding (p). Padding refers to the addition of appropriate number of width and height on each side of the input feature map to ensure that the output feature map has the same dimensions as the input. Given the input image of size ($win \times hin$), the size of output feature map can be calculated using the formula:

$$w_{out} = \left\lceil \frac{w_{in} - f_w + 2p}{s_w} \right\rceil + 1 \quad h_{out} = \left\lceil \frac{h_{in} - f_h + 2p}{s_h} \right\rceil + 1 \quad (4)$$

w = width size of the feature map (input = win ; output = $wout$).

h = height size of the feature map (input = hin ; output = $hout$).

f = convolution filter size (width = fw , height = fh).

p = padding size.

s = stride size (width = sw , height = sh).

The convolution operation is illustrated in Fig. 2. The filter (yellow square) is sliding over the input (green square). The area where the filter is overlapping with the input is referred to as a receptive field. The dot products between the filter and the corresponding receptive field from the input go into the feature map (pink square). The filter is then slid by the stride value (which is one in this case). Each unit in the convolution layer is connected to the local patches in the feature maps of the previous layer through a set of weights called filter bank. This local weighted sum is then passed to the activation layer. The purpose of the activation layer is to allow non-linearity transform neurons within the network layer. Several activation functions exist, such as sigmoid: $(x) = (1 + e^{-x})^{-1}$, tanh: $f(x) = \tanh(x)$ and rectified Linear Units (ReLU): $f(x) = \max(0, x)$. We utilized ReLU in the all hidden layers because the network was able to converge faster compared with sigmoid or tanh functions (Krizhevsky et al., 2012).

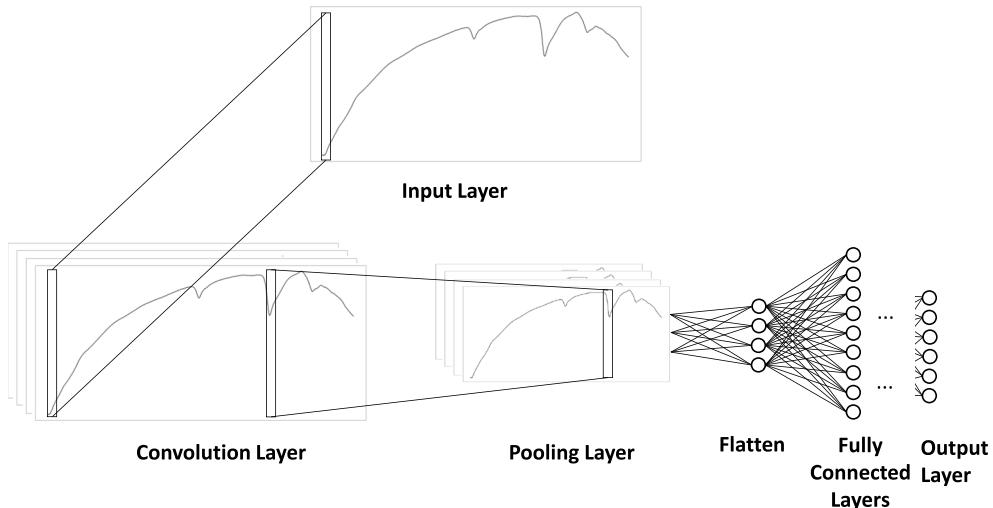


Fig. 1. Representation of the one-dimensional Convolutional Neural Network (CNN) architecture with visible near-infrared (vis-NIR) spectral data input connected to the convolution, pooling, flatten, fully connected and the output layer.

2.3.3.3. Pooling. The pooling layer is used to reduce the dimensionality of the feature map while retaining information from the input feature map. It reduces the computational cost by reducing the number of parameters and prevents overfitting, therefore increasing the overall performance and accuracy of the network. The type of pooling utilized in this study is max-pooling. The max-pooling operation selects the maximum value from the pooled field. The pooling operation is illustrated in Fig. 3.

2.3.3.4. Flatten. Flatten is a layer used to convert the output of pooling layer (a multi-dimensional data) into a one-dimensional vector which can be fed into the fully connected layer.

2.3.3.5. Fully connected. The fully connected layer contains numerous neurons that are connected to all node in the preceding layers. This layer is always found at the end of the CNN network. The fully connected layer is illustrated in Fig. 4.

2.3.3.6. Output. The output layer has a number of neurons, which is equal to the number of target variables. Depending on the type of output, the output layer utilizes a different type of activation. For regression, the linear activation is used.

Each of the layers contains units, called neurons. The behaviour of neurons depends on the previous layer neurons.

CNN enables the possibility of multi-task learning. Though generally acceptable performance can be achieved when a model focussed on a single-task, sharing information from related tasks might improve the model (Ruder, 2017). Ramsundar et al. (2015) and Padarian et al. (2019) observed that the model performance improved significantly when multi-tasks were assigned in comparison to the single-task setting. This improvement is particularly important when the target variables are correlated (Xu et al., 2017). Creating multiple models for the multiple outputs prediction is not optimal because the extracted data from the input may be redundant, and the correlation between inputs may not be preserved. Multi-task learning can be achieved by having some shared layers within the neural networks and keeping several unshared task-specific output layers.

Furthermore, CNN enables the instance of fusing inputs from different sources in various ways, such as the OPA method and two-channel method (which could not be done in PLSR and Cubist model).

Because of the huge number of hyperparameters (number of weights and biases in each layer) used in the network, overfitting often becomes a problem. The training dataset is further split into 75% calibration and 25% validation. Only the calibration dataset is used during the training of the CNN, and the remaining data is used to tune the hyperparameters. Batch normalization and dropout layers are also introduced to the network to prevent overfitting. Batch normalization layer is used to prevent internal covariate shift by standardizing each element in the

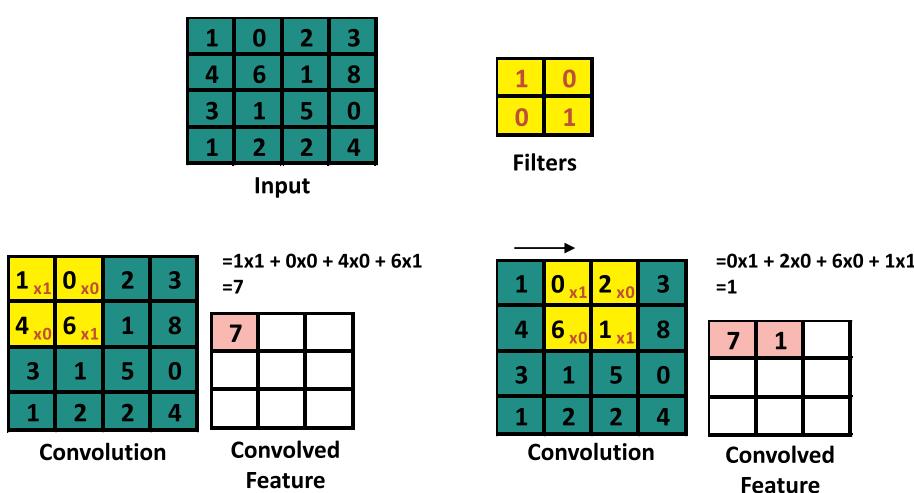


Fig. 2. Illustration of the convolution operation to produce convolved feature (pink colour) is calculated by taking the dot product of the filters (yellow colour) and the corresponding receptive field from the input layer using 2×2 filter and stride 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

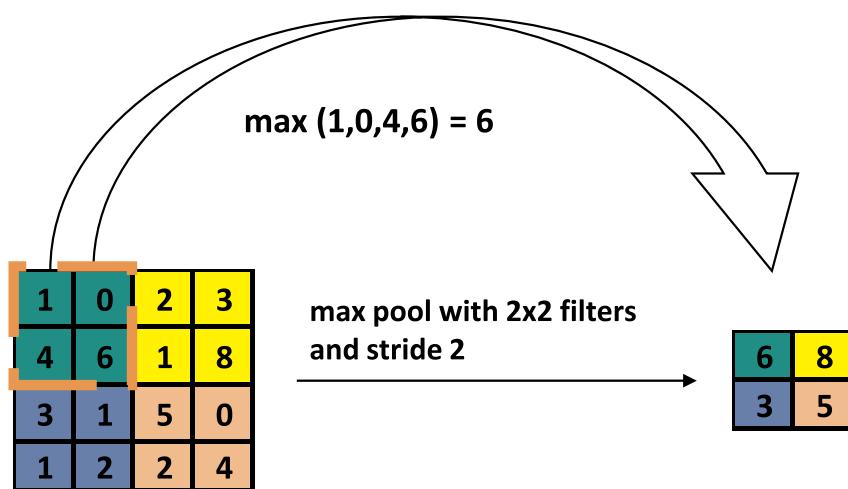


Fig. 3. Illustration of the max-pooling operation using 2×2 filter and stride 2.

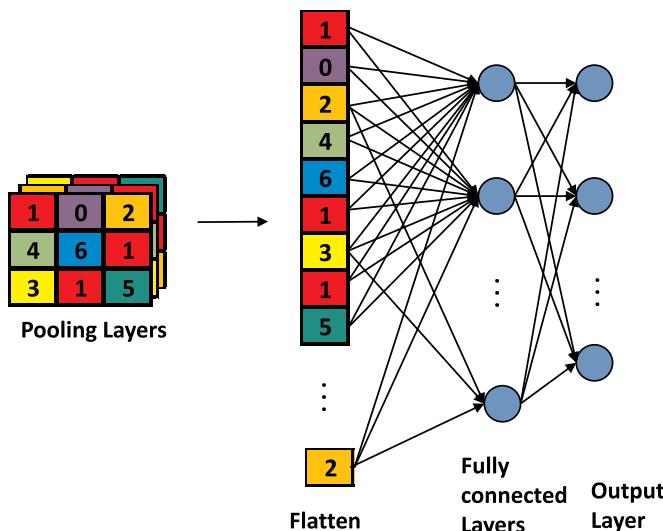


Fig. 4. Illustration of flatten layer that is connecting the pooling layers to the fully connected layers for the various targets in the output layer.

layer to zero mean and unit variance (Ioffe and Szegedy, 2015). Batch normalization is added after each convolution layer. A dropout is a form of regularization that randomly deactivates neurons and connection during network training. The dropout layer is added to the network at the last pooling layer as well as in between the fully connected layers and the output layer.

We created two different convolution networks architectures for the two different inputs (1D and 2D):

1. One-dimensional input CNN (as spectra)

This model was implemented when we used the spectra as one-dimensional input. The input of one-dimensional data was fed into a Gaussian noise layer before training. Gaussian noise layer acted as a regularization layer to mitigate overfitting.

This network consisted of seven trainable layers - four convolutional layers and three fully connected layers. The output of the convolutional layer was passed to the pooling layer. The convolution layer contained 32 filters with a filter size of 20, stride as one and zero paddings. The number of filters at each subsequent convolutional layer was increased by a factor of two, while other parameters were kept the same. The pooling size of the first layer was set to two, and the following ones to five. The last output feature maps from the pooling layer were then

flattened and connected to fully connected layers as input. Some of the layers in the network were shared to allow multi-task setting. The detailed architecture of the 1D model is summarized in Table 2.

2. Two-dimensional input CNNs (as spectrograms)

Two-dimensional input was obtained by transforming the 1D spectra into spectrogram using Hann window with a segment length of 100 with 50% overlap and sampling frequency of one. The output was then adjusted to a log scale to promote faster convergence (Padarian et al., 2019). Vis-NIR spectra with a vector length of 2151 were converted to a matrix with a size of 51×43 while the MIR spectra were resampled to the same length as the vis-NIR spectra with a total of 2151 wavelength variables so that it resulted in the same matrix size as the vis-NIR spectrogram.

This two-dimensional network consisted of several combinations of convolution and pooling layers. The convolution layer comprised of 3×3 size filters. After several convolution and pooling layers, the last output feature maps were flattened and connected into fully connected layers. We also set-up several shared layers in this network to allow multi-task setting. The 2D model architecture is included in Table 3.

All the observed soil properties were standardized using the mean and standard deviation of the training dataset before being used in the CNN model. The architecture of both 1D and 2D networks is included in Fig. 5.

The weights of the neurons in the network were initialized with the He initialization (He et al., 2015) that account for non-linearities activation. This initialization enabled models to easily converge (He et al.,

Table 2
One-dimensional Convolutional Neural Network (CNN) architecture.

Type	Shared	Filter size	# Filters	Activation
Convolutional	Yes	20	32	ReLU
Max-pooling	Yes	2	–	–
Convolutional	Yes	20	64	ReLU
Max-pooling	Yes	5	–	–
Convolutional	Yes	20	128	ReLU
Max-pooling	Yes	5	–	–
Convolutional	Yes	20	256	ReLU
Max-pooling	Yes	5	–	–
Dropout (0.4)	Yes	–	–	–
Flatten	Yes	–	–	–
Fully-connected	No	–	~100	ReLU
Dropout (0.2)	No	–	–	–
Fully-connected	No	–	~1	linear

Table 3
Two-dimensional Convolutional Neural Network (CNN) architecture.

Type	Shared	Filter size	# Filters	Activation
Convolutional	Yes	3 × 3	32	ReLU
Max-pooling	Yes	2 × 2	—	—
Convolutional	Yes	3 × 3	64	ReLU
Convolutional	Yes	3 × 3	64	ReLU
Max-pooling	Yes	2 × 2	—	—
Convolutional	Yes	3 × 3	128	ReLU
Convolutional	Yes	3 × 3	128	ReLU
Max-pooling	Yes	2 × 2	—	—
Dropout (0.4)	Yes	—	—	—
Flatten	Yes	—	—	—
Fully-connected	No	—	100	ReLU
Dropout (0.2)	No	—	—	—
Fully-connected	No	—	1	linear

2015). These weights were then updated through training processes. The network was trained with a batch size of 50 with a maximum epoch of 500. Batch size referred to a smaller subset from the whole training set. An epoch referred to one cycle through the whole training dataset. The network was trained with early stopping on the validation set with a patience of 30 epochs using Adam optimizer with an initial learning rate of 0.001 (Kingma and Ba, 2014) and a learning rate decay of 0.5 with a patience of three epochs.

2.4. Implementation

The PLSR and Cubist models were implemented in R (v3.5.3; R Core Team, 2019), using the pls (v2.7.0; Mevik et al., 2018) and Cubist (v0.2.2; Kuhn and Quinlan, 2018) package, respectively. The CNN was implemented in Python (v3.5.1; Python Software Foundation, 2017) using Keras library (v2.1.2; Chollet, 2015) and Tensorflow (v1.4.1; Abadi et al., 2015) backend.

2.5. Spectral pre-processing

Spectral pre-processing was used to correct for sample variations and noisy spectra. The two spectra were pre-processed separately.

For the vis-NIR data, the spectra were first converted from reflectance (R) to absorbance ($A = \log(1/R)$). Only regions that contain a high signal to noise ratio were kept. Absorbance spectra between the wavelength of 500–2450 nm and wavenumbers of 4000–700 cm⁻¹ were kept for vis-NIR and MIR spectral data respectively. For PLSR and Cubist modelling, the spectra were then pre-processed using Savitzky-Golay (SG) smoothing with a window size of 11 and a second order polynomial, followed by Standard Normal Variate (SNV) transformation. SG is an averaging algorithm that fits polynomial to the data points (Savitzky and Golay, 1964). SNV is used to standardize the spectra, scaling it to zero mean and unit standard deviation (Rinnan et al., 2009). Illustrations for the pre-processed vis-NIR and MIR spectra were included in Supplementary Material Fig. S2 and S3. For the CNN model, the raw spectra from both vis-NIR and MIR were standardized using SNV transformation before being fed into the model.

Calibration of NIR and MIR data using PLSR and Cubist models require pre-processing of the spectra to enhance prediction. As a comparison, the performance of PLSR and Cubist model using spectral data without pre-treatment is included in Supplementary Materials Table S1 and S2.

2.6. Spectral data fusion

We analysed several methods of fusing spectral data, i.e., simple

concatenation for the PLSR and Cubist models; two-channel method and outer product analysis (OPA) method for the CNN model. Concatenation was simply performed by stacking the spectral data from vis-NIR to MIR range. For the CNN model, the architecture of the network was modified to accept the spectral data as two different channels input. To enable the CNN model in accepting input using different channels, the dimensions of the input data must be the same. Therefore, the MIR spectra were resampled down to achieve the same number of spectral variables as the vis-NIR spectra. Each spectrum was then fed separately into different channels of the CNN model.

For OPA analysis, the absorbances of the MIR range (3578 variables) were multiplied by all the absorbances in the vis-NIR range (2151 variables) which resulted in a matrix with 7,696,278 variables (3578 × 2151). The OPA matrix was further resampled down to a dimension of 358 × 216 by selecting a value for every ten variables because it was computationally expensive to analyse such a large dimension matrix,

2.7. Accuracy evaluation

To evaluate the predictive ability of the models, the coefficient of determination for prediction (R^2) and root mean square error of prediction (RMSE), bias, and the ratio of performance to the interquartile range (RPIQ; Bellon-Maurel et al., 2010) were calculated.

The R^2 is calculated as:

$$R^2 = 1 - \frac{RSS}{TSS} \quad (5)$$

where RSS is the residual sum of squares and TSS is the total sum of squares. The RMSE is defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (obs_i - pred_i)^2}{n}} \quad (7)$$

where obs is the observed value, $pred$ is predicted value, and n is a total number of observations i . Bias is defined as:

$$bias = \frac{\sum_{i=1}^n (pred_i - obs_i)}{n} \quad (8)$$

RPIQ can be defined as:

$$RPIQ = \frac{IQ}{RMSE} \quad (9)$$

where IQ = inter quantile distance of the observed values.

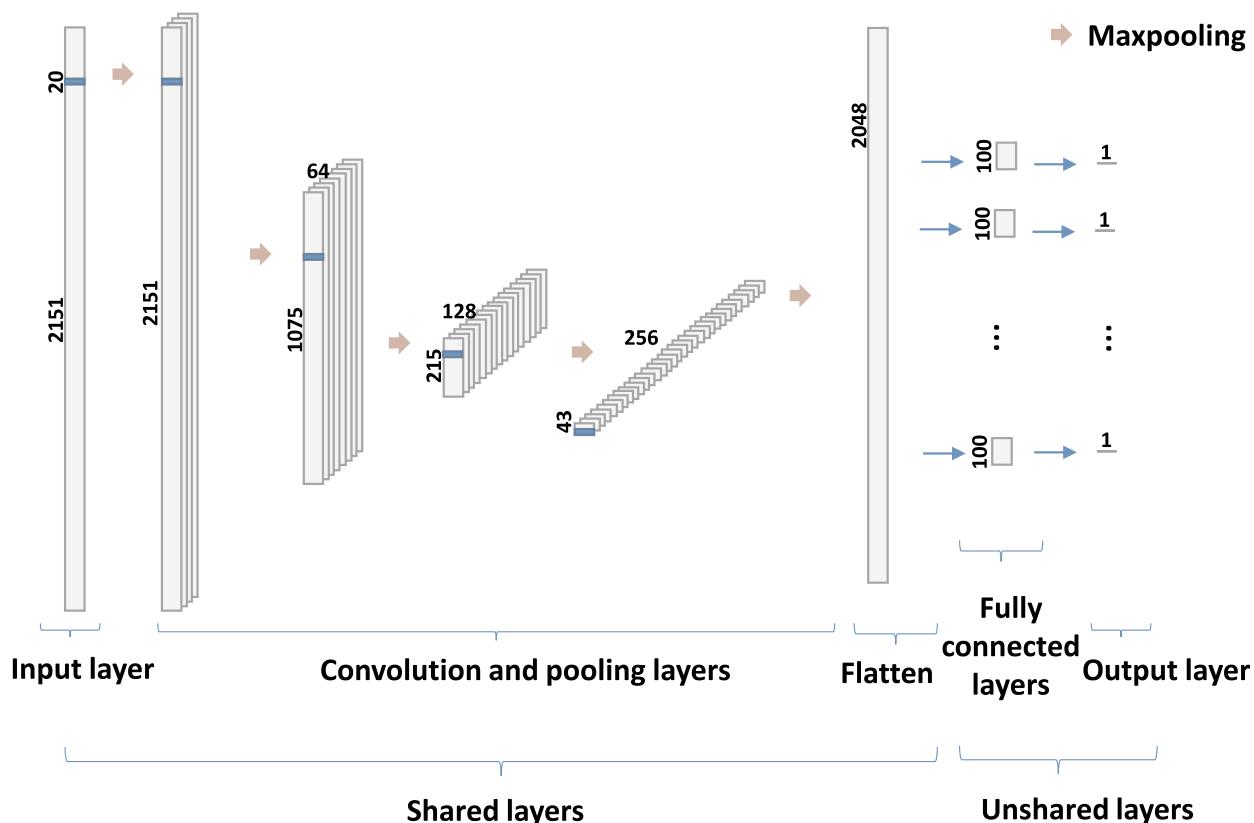
$$IQ = Q_3 - Q_1 \quad (10)$$

where Q_3 is the third quartile, and Q_1 is the first quartile.

2.8. Exploring the correlation between predicted soil properties

The capability of the CNN model to predict the various soil properties simultaneously could enable it to inherit the correlation among the predicted soil properties, despite multi-tasking CNN does not model output correlation explicitly (Xu et al., 2017). This is due to the deterministic behaviour of the output soil properties that are reproduced by the CNN model. We determined the Pearson's linear correlation coefficient (r) between the six soil properties from the observed data and compared it with the correlation of the properties derived from predicted values. We calculated the absolute relative difference of the correlation coefficient. The model that provided lower absolute relative difference was able to retain the correlation of the properties better. To create an easy comparison among all the models, we calculated the mean of the absolute relative difference.

a.



b.

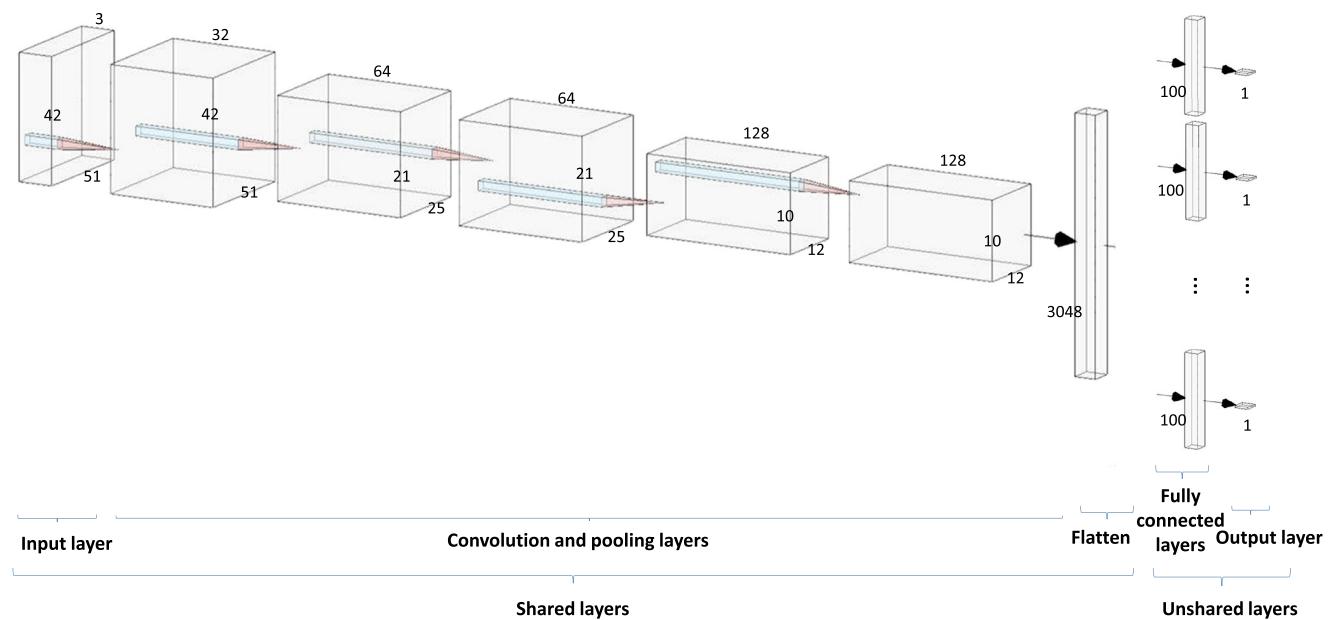


Fig. 5. Convolutional Neural Network (CNN) architecture for (a) one-dimensional and (b) two-dimensional inputs.

Table 4

Goodness of fit of various soil properties prediction in the test set using Partial Least Squares Regression (PLSR), Cubist and Convolutional Neural Network (CNN) model with near-infrared (NIR), mid-infrared (MIR) and combined vis-NIR-MIR spectral data as one-dimensional input. R^2 – coefficient of determination, RMSE – root mean squared error on prediction, RPIQ – ratio of performance to inter-quartile distance, TC – Total Carbon, OC – Organic Carbon, CEC – Cation Exchange Capacity.

Model	Properties	Vis-NIR				MIR				Combined (vis-NIR + MIR)			
		R^2	RMSE	bias	RPIQ	R^2	RMSE	bias	RPIQ	R^2	RMSE	bias	RPIQ
PLSR	TC (%)	0.54	1.11	0.00	0.60	0.95	0.37	0.00	1.83	0.95	0.37	0.00	1.82
	OC (%)	0.49	1.02	-0.01	0.47	0.94	0.35	0.01	1.37	0.94	0.36	0.00	1.33
	CEC (cmol(+) / kg)	0.61	6.73	-0.04	1.05	0.93	2.85	0.12	2.49	0.93	2.85	0.09	2.49
	Clay (%)	0.67	9.03	-0.03	1.21	0.90	4.97	0.04	2.19	0.91	4.73	-0.01	2.30
	Sand (%)	0.44	21.79	0.11	0.95	0.90	9.04	0.01	2.28	0.90	9.09	-0.02	2.27
	pH	0.66	0.78	0.01	1.18	0.87	0.49	-0.01	1.87	0.87	0.48	-0.01	1.91
Cubist	TC (%)	0.70	0.90	-0.03	0.75	0.97	0.28	0.00	2.41	0.97	0.29	-0.01	2.33
	OC (%)	0.66	0.83	-0.06	0.58	0.97	0.24	0.00	2.02	0.97	0.25	0.00	1.92
	CEC (cmol(+) / kg)	0.68	6.15	-0.10	1.15	0.95	2.37	-0.02	2.99	0.96	2.26	0.05	3.14
	Clay (%)	0.76	7.84	0.10	1.39	0.93	4.22	-0.01	2.58	0.92	4.34	0.00	2.51
	Sand (%)	0.60	19.00	-0.18	1.09	0.91	8.54	-0.09	2.42	0.91	8.50	-0.19	2.43
	pH	0.76	0.66	0.00	1.38	0.91	0.39	-0.01	2.35	0.93	0.36	-0.02	2.53
CNN	TC (%)	0.83	0.67	-0.03	0.99	0.98	0.21	0.01	3.12	0.98	0.22	-0.01	3.01
	OC (%)	0.83	0.60	-0.02	0.81	0.98	0.20	0.01	2.37	0.98	0.21	-0.01	2.27
	CEC (cmol(+) / kg)	0.87	3.92	-0.20	1.81	0.97	1.98	0.02	3.58	0.97	1.98	-0.20	3.59
	Clay (%)	0.87	5.73	-0.44	1.90	0.96	3.27	-0.22	3.33	0.95	3.45	-0.34	3.16
	Sand (%)	0.80	13.03	-0.36	1.58	0.95	6.26	-0.35	3.30	0.95	6.46	-0.01	3.20
	pH	0.85	0.52	0.01	1.76	0.95	0.30	-0.01	3.10	0.95	0.31	-0.02	3.00

2.9. Uncovering the CNN black box

Aside from comparing the performance of the CNN model (both 1D and 2D input) to Cubist and PLSR model, we also explored more deeply into the CNN model.

- We determined how well the CNN model retained the correlation among the predicted soil properties using Pearson correlation and attempted to derive indirect predictions on some properties, such as inorganic carbon content and silt content.
- We determined which wavelengths are utilized within the CNN model to predict certain properties by keeping most of the wavelengths variable constant as mean values while allowing only five variables to change at any given time.

3. Results

3.1. Effect of spectral input

The results obtained using the various spectra (vis-NIR, MIR, and vis-NIR-MIR) as one-dimensional input were tabulated in Table 4. Note that all results presented here pertain to the test dataset ($N = 3648$). We first evaluated the accuracy of prediction of the six soil properties using conventional PLSR and Cubist regression techniques.

The overall performance using various spectral data inputs are between fair to excellent. Regression models with vis-NIR spectra yield the lowest accuracy in comparison to MIR and the combined spectra. The lower accuracy achieved is due to the weak spectral signature in the vis-NIR region which only relates to the overtones and combinations of the fundamental molecular bands. The model performance for the predictions of all soil properties improves with the use of MIR spectra. These results are in line with findings of Terra et al. (2015) and Clairotte et al. (2016) among others.

The modelling results show that prediction using vis-NIR for all properties are good with R^2 between 0.44 and 0.67 for PLSR, and 0.60–0.76 for Cubist. Prediction with MIR is significantly better with R^2 of 0.87–0.95 for PLSR and between 0.91 and 0.97 for the Cubist model. The high R^2 values (> 0.9) for all properties on the validation dataset using MIR is rarely reported in the literature. For example, Terra et al.

(2015) reported R^2 values of 0.77, 0.88, 0.90, 0.80 and 0.54 for the predictions of organic carbon, clay and sand content, CEC and pH respectively. A review by Soriano-Disla et al. (2014) reported median R^2 values of 0.93, 0.93, 0.80, 0.83, 0.85 and 0.75 for the predictions of total carbon, organic carbon, clay, and sand content, CEC, and pH, respectively, using MIR spectra.

This high performance found in the KSSL dataset could be due to the precise and accurate laboratory analysis of the soil samples. The uniformity and standard laboratory procedures conducted at the KSSL made it a highly reliable soil and spectral dataset.

When combined spectral inputs (vis-NIR and MIR) were used with PLSR or Cubist model, no significant improvement is observed in comparison with the use of only MIR spectra as input with R^2 ranging between 0.87 and 0.95 for the PLSR, and 0.91–0.97 for Cubist. The results are similar to those obtained using the MIR spectra. The lack of model improvement is due to the already high accuracy achieved by using MIR spectra; and hence little room for further improvement. Terra et al. (2019) reported an R^2 of 0.81 for the prediction of organic carbon using the combined vis-NIR and MIR spectra, 0.69 for vis-NIR, and 0.77 for MIR. For the KSSL dataset, since the MIR already produced R^2 of 0.97 for Cubist, the addition of vis-NIR spectra could not further enhance the accuracy.

3.2. Effect of model types

Given the same spectral data input, CNN achieved a higher model performance in comparison to the PLSR and Cubist models. Also, the spectral data input used in the CNN model did not undergo spectral pre-processing as in the PLSR and Cubist models. The performance of all models is shown in Figs. 6–8.

To gauge the accuracy enhancement of CNN, we calculated the relative improvement of soil properties prediction in terms of %RMSE and summarized in Table 5. Compared to the PLSR model, the relative accuracy improvement achieved by CNN model range from 33.33 to 41.75% for vis-NIR, 30.53–43.24% for MIR, and 27.06–41.67% for the combined vis-NIR-MIR. The relative accuracy improvement when the CNN model compared to the Cubist model ranged between 21.21 and 36.26% for vis-NIR, 16.46–26.70% for MIR, and 12.39–24.14% for the combined vis-NIR-MIR.

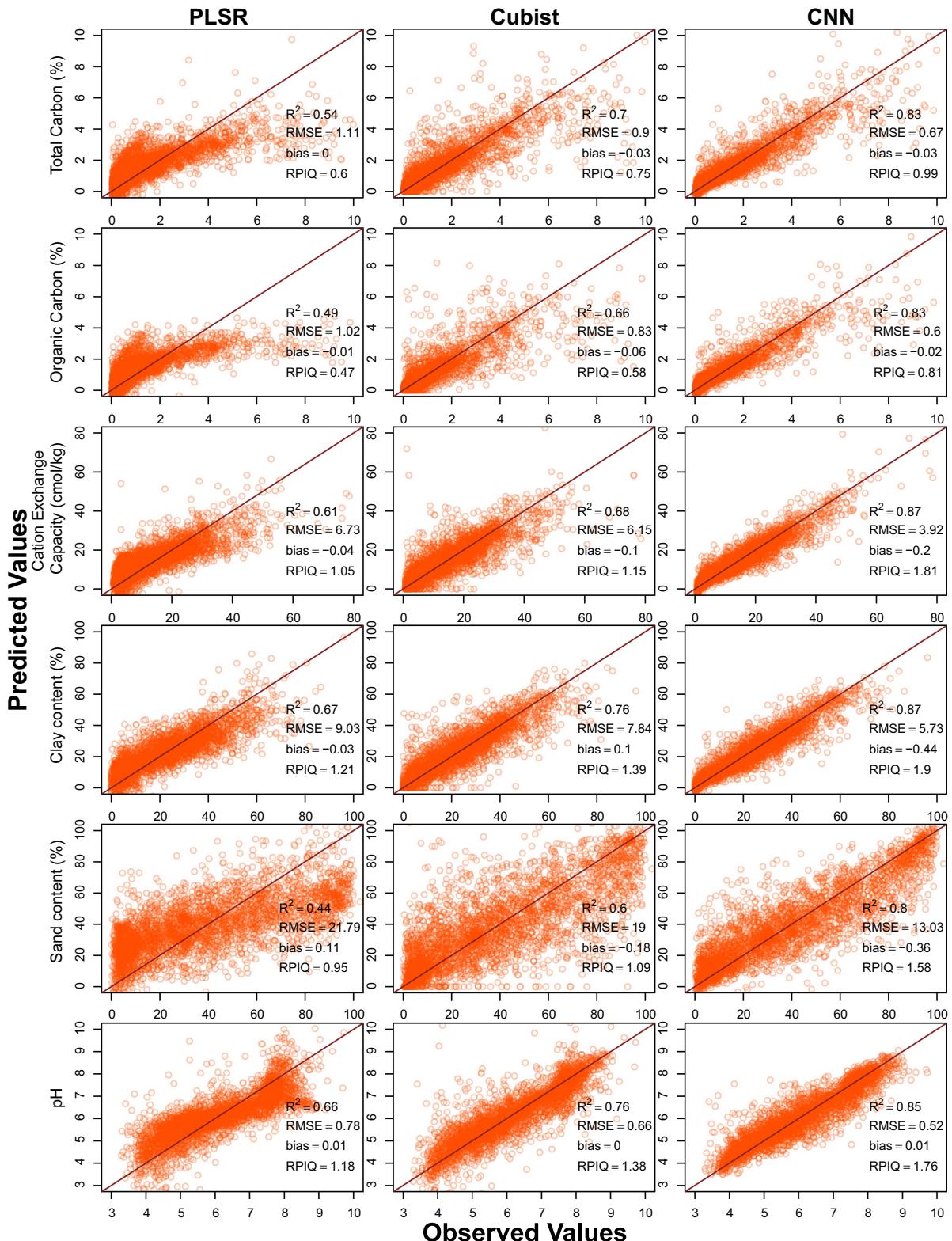


Fig. 6. Goodness of fit showing the relationship of observed and predicted soil properties (pH, sand content, clay content, cation exchange capacity, organic carbon and total carbon) using vis-NIR spectral data input with Cubist, PLSR and CNN model. Vis-NIR – visible-near-infrared, PLSR – Partial Least Squares Regression, CNN – Convolutional Neural Network, R^2 – coefficient of determination, RMSE – root mean squared error, RPIQ – ratio of performance to inter-quartile distance.

Note that the R^2 improvement of the CNN model in comparison to Cubist using MIR was small, as the Cubist model already produced excellent results ($R^2 > 0.9$). However, the improvement over vis-NIR data was significant, with CNN achieving R^2 of 0.80–0.87 in comparison to 0.44–0.67 for the PLSR, and 0.60–0.76 for the Cubist model.

3.3. Effect of multidimensional input

Out of the three models utilized, only the CNN model could accept multi-dimensional input. Hence, the performance of the CNN model using a spectrogram (short time Fourier transformed spectral data) as a

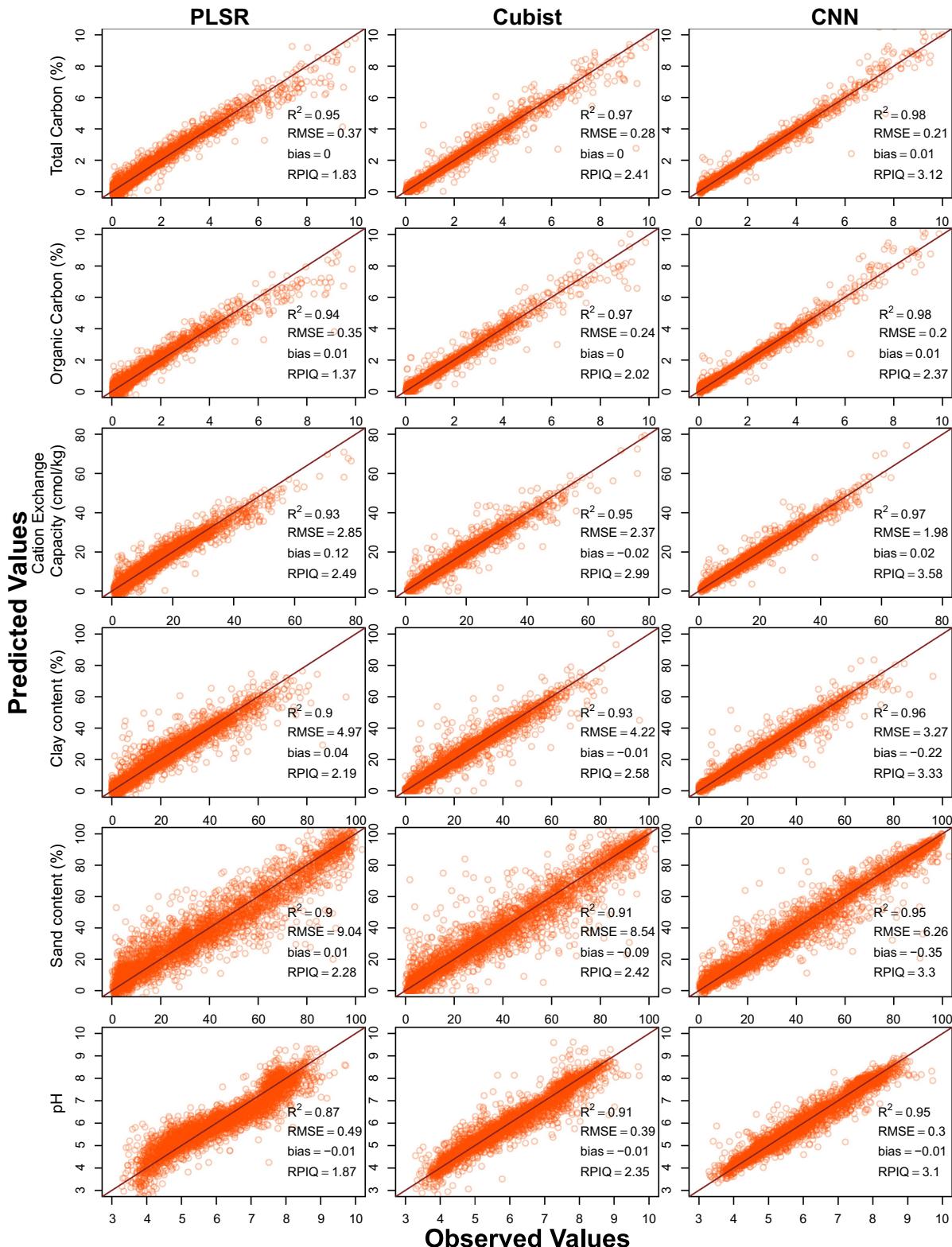


Fig. 7. Goodness of fit showing the relationship of observed and predicted soil properties (pH, sand content, clay content, cation exchange capacity, organic carbon and total carbon) using MIR spectral data input with Cubist, PLSR and CNN model. MIR – mid-infrared, PLSR – Partial Least Squares Regression, CNN – Convolutional Neural Network, R^2 – coefficient of determination, RMSE – root mean squared error, RPIQ – ratio of performance to inter-quartile distance.

two-dimensional input was tested. The comparison of one-dimensional and two-dimensional spectral input was tabulated in Table 6. In comparison to the one-dimensional spectral data input, the performance using the two-dimensional spectral data input was lower. Padarian et al. (2019) only tested the spectrogram as input but did not test the

use of one-dimensional spectral input.

The CNN models were further tested for improving the accuracy of prediction using combined vis-NIR-MIR spectra. For the fusion, the vis-NIR and MIR spectra were fed into the model as two-channel of one-dimensional input and two-dimensional input. The use of OPA as a

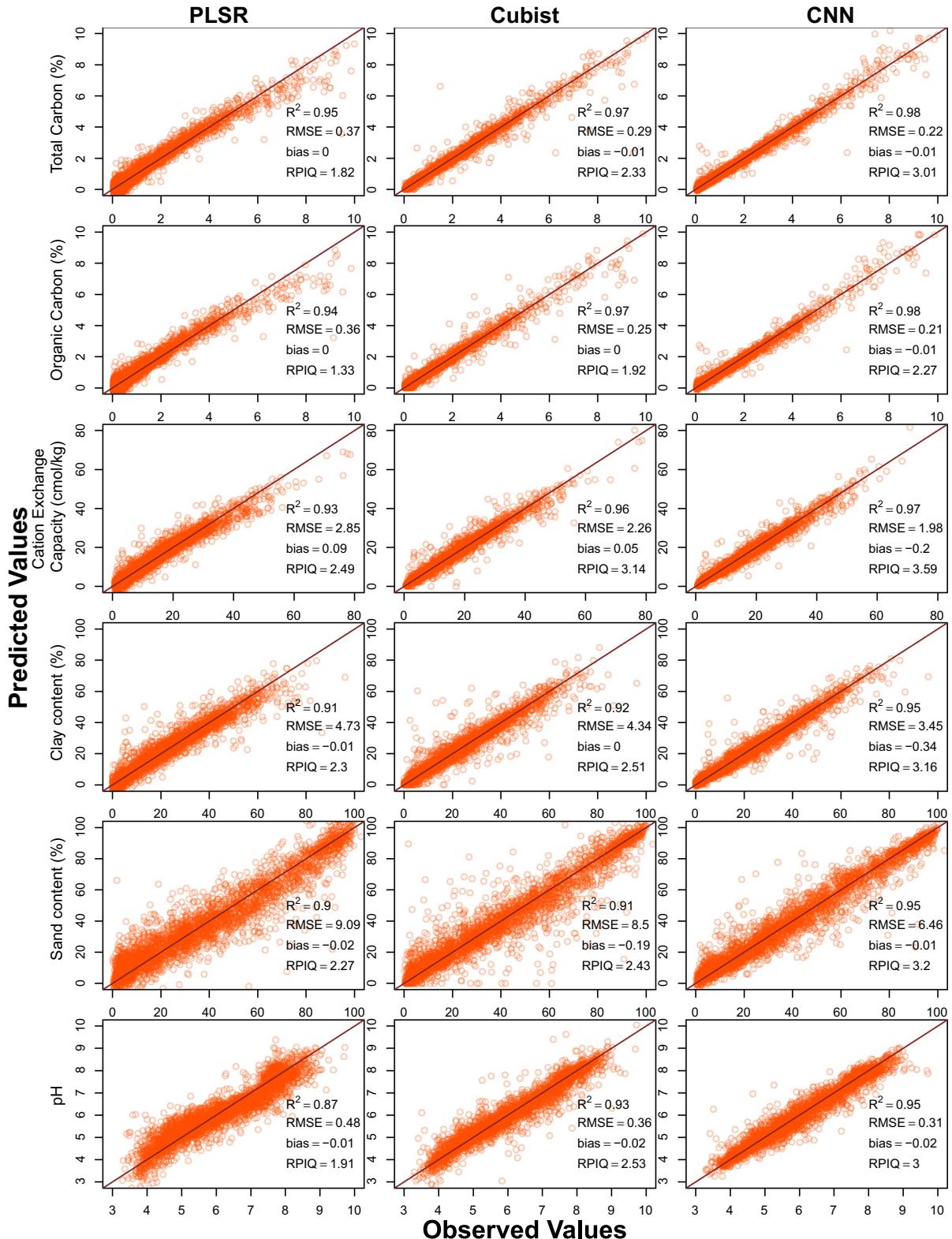


Fig. 8. Goodness of fit showing the relationship of observed and predicted soil properties (pH, sand content, clay content, cation exchange capacity, organic carbon and total carbon) using combined spectral data input (vis-NIR and MIR) with Cubist, PLSR and CNN model. Vis-NIR – visible-near-infrared, MIR – mid-infrared, PLSR – Partial Least Squares Regression, CNN – Convolutional Neural Network, R^2 – coefficient of determination, RMSE – root mean squared error, RPIQ – ratio of performance to inter-quartile distance.

method to combine spectral data information was also analysed. The model performance using the OPA was better than the two-channel of a two-dimensional model but similar to the two-channel of the one-dimensional model. Because of the computational cost involved in

creating the OPA matrix that did not contribute to the model performance, the use of raw spectral data in the CNN model was suggested (see Table 7).

Nevertheless, the fusion of vis-NIR and MIR did not result in a

Table 5

Relative improvement of the Convolutional Neural Network (CNN) model in comparison to Partial Least Squares Regression (PLSR) and Cubist model in terms of %RMSE.

Spectral data	Property	Relative improvement compared to PLSR (%)	Relative improvement compared to Cubist (%RMSE)
vis-NIR	TC (%)	39.64	25.56
	OC (%)	41.18	27.71
	CEC (cmol(+) / kg)	41.75	36.26
	Clay (%)	36.54	26.91
	Sand (%)	40.20	31.42
	pH	33.33	21.21
	TC (%)	43.24	25.00
	OC (%)	42.86	16.67
	CEC (cmol(+) / kg)	30.53	16.46
	Clay (%)	34.21	22.51
MIR	Sand (%)	30.75	26.70
	pH	38.78	23.08
	TC (%)	40.54	24.14
	OC (%)	41.67	16.00
	CEC (cmol(+) / kg)	30.53	12.39
Combined	Clay (%)	27.06	20.51
	Sand (%)	28.93	24.00
	pH	35.42	13.89

significant improvement over only using MIR spectra. As discussed previously, the performance of MIR in the prediction was already excellent ($R^2 > 0.95$) and thus no further gain could be obtained by including vis-NIR spectra.

3.4. Correlation between soil properties

We further assessed the models in terms of maintaining the correlation coefficient between soil properties using the Pearson correlation coefficient (r). Soil properties had inherent correlation, e.g., sand was inversely correlated with CEC ($r = -0.60$) and clay content ($r = -0.73$) and CEC was positively correlated with clay content ($r = 0.71$). The correlation matrices between soil properties using the observed values and predicted values were included in Supplementary

Materials Table S4–S6. Although multi-tasking CNN does not explicitly use the correlation structure of the outputs in its learning process, multi-tasking optimised all output variables together, thus indirectly maintain the correlation. Maintaining output correlation is one of the objectives in multitasking modelling (Zhang and Yang, 2017).

The performance of a single-task CNN model (a model is predicting a property at a time) is also tested and included in Supplementary Materials Fig. S4 and Table S3 for comparison. Single-task CNN perform worse than multi-task for most soil properties in all types of spectra input, except for pH. This could be due to the large correlation of soil C, clay, sand, and CEC which enable the CNN to learn inherent structure from such data to improve prediction, supporting the proposal of Xu et al. (2017). As soil pH has small correlation with other soil properties, including it in the multi-task model did not help improving the accuracy of the model.

Using the vis-NIR spectral data, the mean of the absolute relative difference of the correlation coefficients for all six soil properties were 0.97, 0.51 and 0.23 for PLSR, Cubist, and CNN respectively in comparison to the correlation of the original data. Using the MIR spectral data, the mean of absolute relative difference of the correlation coefficient of all soil properties using PLSR, Cubist, and CNN model were reduced to 0.09, 0.07 and 0.05 (see Supplementary Material Table S7 and S8 for the relative difference in the correlation matrices). Regardless of the type of input spectral data, multi-tasking CNN was found to work best in maintaining the correlation coefficient values of the dataset as shown by the lowest absolute differences. Nevertheless, the improved correlation structure of the predicted properties could also be due to improved accuracy of the model, e.g., in the case of vis-NIR vs. MIR.

CNN on MIR spectra produced the most accurate prediction and maintain almost the same correlation between soil properties. As PLSR and Cubist models predicted each soil property individually, the correlation between soil properties was not maintained. We tried to derive an indirect prediction of two properties from the predicted values that we already obtained.

3.4.1. Predictions of inorganic carbon

Our models only predicted TC and OC from the data. We could derive inorganic carbon prediction (IC) and compared the results with samples that have measured IC data ($n = 957$). The predicted IC was calculated as:

Table 6

Goodness of fit of various soil properties prediction on the test set using visible near-infrared (vis-NIR), mid-infrared (MIR) and combined vis-NIR-MIR spectral data as both one-dimensional and two-dimensional input with Convolutional Neural Network (CNN) model. R^2 – coefficient of determination, RMSE – root mean squared error, RPIQ – ratio of performance to inter-quartile distance, TC – Total Carbon, OC – Organic Carbon, CEC – Cation Exchange Capacity.

Model	Properties	One-dimensional input				Two-dimensional input			
		R^2	RMSE	bias	RPIQ	R^2	RMSE	bias	RPIQ
vis-NIR	TC (%)	0.83	0.67	-0.03	0.99	0.69	0.91	-0.03	0.74
	OC (%)	0.83	0.60	-0.02	0.81	0.71	0.77	-0.06	0.63
	CEC (cmol(+) / kg)	0.87	3.92	-0.20	1.81	0.71	5.79	-0.42	1.23
	Clay (%)	0.87	5.73	-0.44	1.90	0.70	8.69	-0.64	1.25
	Sand (%)	0.80	13.03	-0.36	1.58	0.60	18.45	-0.37	1.12
	pH	0.85	0.52	0.01	1.76	0.72	0.70	-0.01	1.31
MIR	TC (%)	0.98	0.21	0.01	3.12	0.97	0.31	-0.01	2.17
	OC (%)	0.98	0.20	0.01	2.37	0.96	0.29	-0.01	1.65
	CEC (cmol(+) / kg)	0.97	1.98	0.02	3.58	0.95	2.40	0.04	2.96
	Clay (%)	0.96	3.27	-0.22	3.33	0.93	4.36	-0.23	2.50
	Sand (%)	0.95	6.26	-0.35	3.30	0.92	8.04	-0.11	2.57
Combined	pH	0.95	0.30	-0.01	3.10	0.92	0.37	-0.02	2.48
	TC (%)	0.98	0.22	-0.01	3.01	0.95	0.37	-0.04	1.83
	OC (%)	0.98	0.21	-0.01	2.27	0.95	0.33	-0.03	1.44
	CEC (cmol(+) / kg)	0.97	1.98	-0.20	3.59	0.94	2.66	-0.25	2.67
	Clay (%)	0.95	3.45	-0.34	3.16	0.92	4.56	-0.40	2.39
	Sand (%)	0.95	6.46	-0.01	3.20	0.91	8.69	-0.22	2.38
	pH	0.95	0.31	-0.02	3.00	0.90	0.42	-0.03	2.19

Table 7

The goodness of fit of various soil properties prediction on the test set using Convolutional Neural Network (CNN) model with various methods to combine vis-NIR and MIR data: (i) two-channel one-dimensional, (ii) two-channel two-dimensional, and (iii) outer product analysis (OPA). R^2 – coefficient of determination, RMSE – root mean squared error, RPIQ – ratio of performance to inter-quartile distance, TC – Total Carbon, OC – Organic Carbon, CEC – Cation Exchange Capacity.

Spectral type	Properties	R^2	RMSE	bias	RPIQ
One-dimensional	TC (%)	0.98	0.22	-0.01	3.01
	OC (%)	0.98	0.21	-0.01	2.27
	CEC (cmol(+) / kg)	0.97	1.98	-0.20	3.59
	Clay (%)	0.95	3.45	-0.34	3.16
	Sand (%)	0.95	6.46	-0.01	3.20
	pH	0.95	0.31	-0.02	3.00
Two-dimensional	TC (%)	0.95	0.37	-0.04	1.83
	OC (%)	0.95	0.33	-0.03	1.44
	CEC (cmol(+) / kg)	0.94	2.66	-0.25	2.67
	Clay (%)	0.92	4.56	-0.40	2.39
	Sand (%)	0.91	8.69	-0.22	2.38
	pH	0.90	0.42	-0.03	2.19
OPA	TC (%)	0.98	0.25	0.01	2.68
	OC (%)	0.97	0.23	0.01	2.07
	CEC (cmol(+) / kg)	0.96	2.11	0.13	3.37
	Clay (%)	0.94	3.83	-0.02	2.85
	Sand (%)	0.95	6.69	-0.08	3.09
	pH	0.93	0.34	0.00	2.68

$$\text{Inorganic Carbon (IC)} = \text{Total Carbon (TC)} - \text{Organic Carbon (OC)} \quad (11)$$

Using just the vis-NIR spectral data predictions, the highest accuracy for the prediction of inorganic carbon was achieved using the CNN model followed by the PLSR and Cubist model with R^2 values of 0.84, 0.67 and 0.61 respectively (see Fig. 9).

For MIR spectral predictions, similar performances were achieved regardless of the model type. PLSR model was performing slightly better than other models with an R^2 value of 0.99 in comparison to 0.98 and 0.97 for CNN and Cubist model respectively. Using the combined vis-NIR and MIR spectral predictions, the performance of the PLSR was still better at R^2 of 0.99 in comparison to 0.98 and 0.95 for CNN and Cubist model respectively. Nonetheless, the performances of all models did not differ much. These results indicate that we do not explicitly need a separate model to predict IC.

3.4.2. Prediction of silt content

Furthermore, we predicted silt content based on the sand and clay content predictions. The silt content was calculated as:

$$\text{Silt content (\%)} = 100 - \text{Sand content (\%)} - \text{Clay content (\%)} \quad (12)$$

Using the vis-NIR data predictions, both the PLSR and Cubist generated poor predictions of silt content with R^2 values of 0.32 and 0.42 respectively in comparison to the CNN model with $R^2 = 0.75$ (see Fig. 10). When the MIR spectral predictions were used, CNN performed

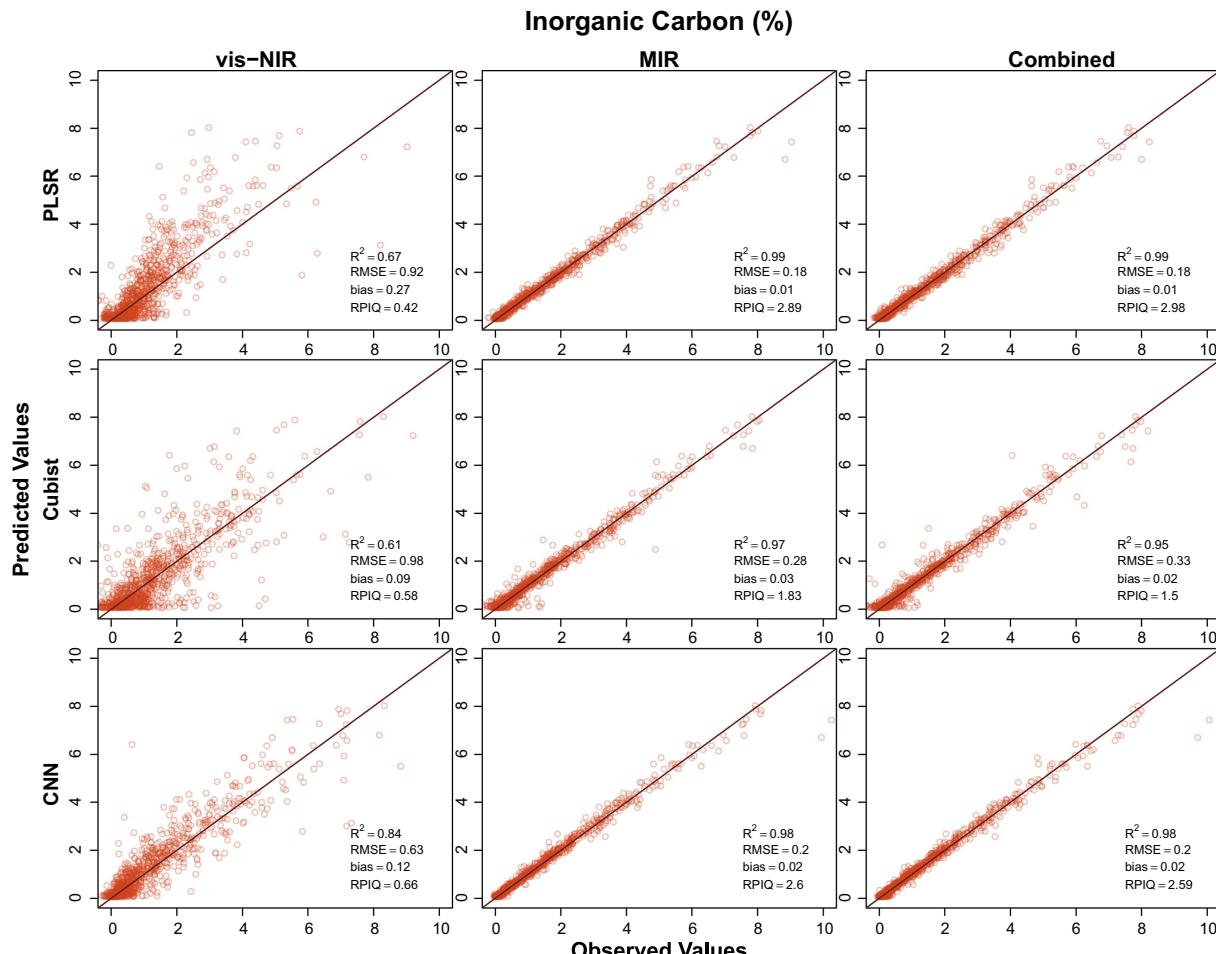


Fig. 9. Goodness of fit showing the relationship of observed and predicted inorganic carbon (IC) values calculated by subtracting the total carbon and organic carbon predictions ($\text{IC} = \text{TC}-\text{OC}$) with vis-NIR, MIR and combined spectral data input and Cubist, PLSR and CNN model. Vis-NIR – visible-near-infrared, MIR – mid-infrared, PLSR – Partial Least Squares Regression, CNN – Convolutional Neural Network, R^2 – coefficient of determination, RMSE – root mean squared error, RPIQ – ratio of performance to inter-quartile distance.

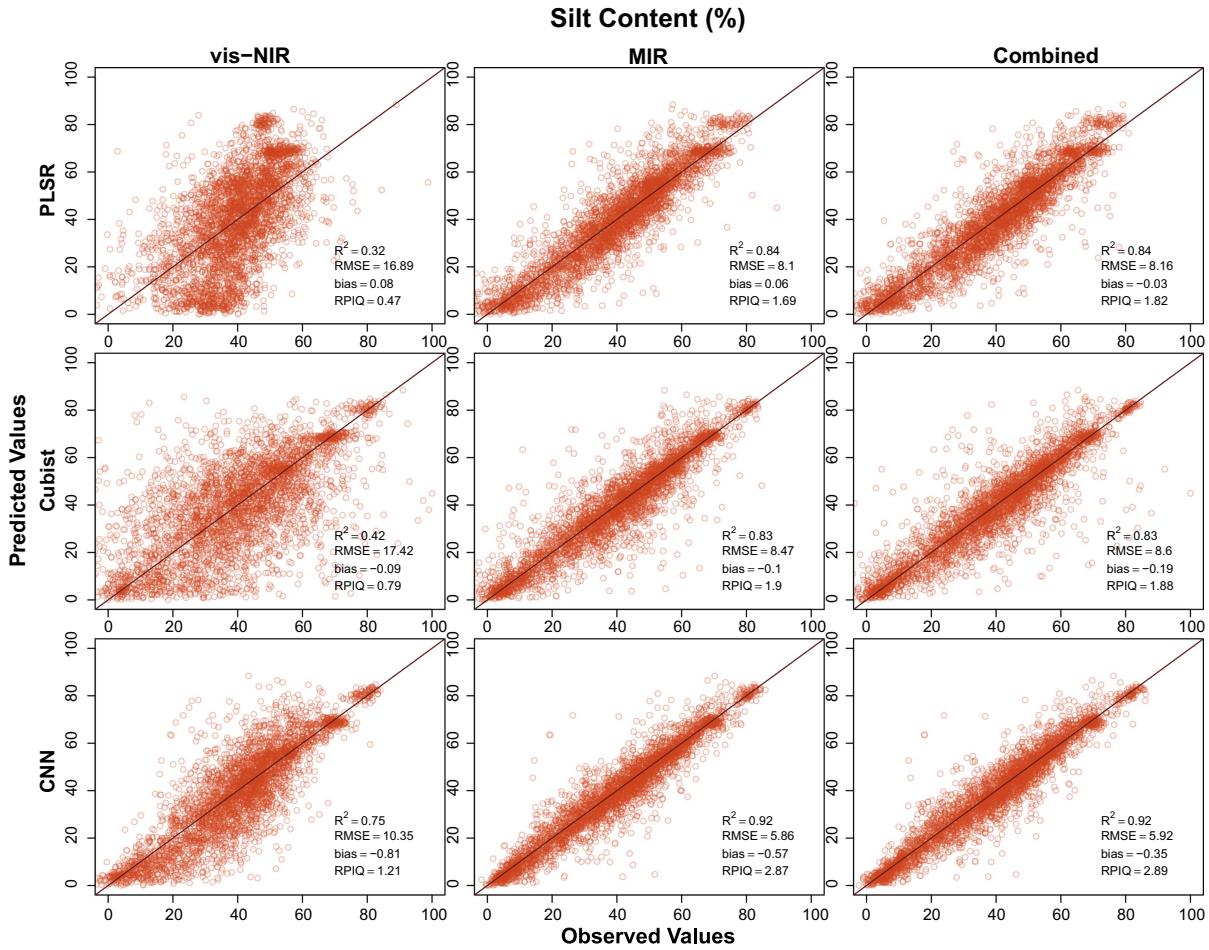


Fig. 10. Goodness of fit showing the relationship of observed and predicted silt content values. (silt = 00 - sand - clay) with vis-NIR, MIR and combined spectral data input and Cubist, PLSR and CNN model. Vis-NIR – visible-near-infrared, MIR – Mid-infrared, PLSR – Partial Least Squares Regression, CNN – Convolutional Neural Network, R^2 – coefficient of determination, RMSE – root mean squared error, RPIQ – ratio of performance to inter-quartile distance.

best with R^2 of 0.92 compared to PLSR and Cubist ($R^2 = 0.84$ and 0.83). The use of combined spectra yielded similar accuracies to those using just the MIR spectral predictions ($R^2 = 0.92$, 0.84 and 0.83 for CNN, PLSR and Cubist model respectively).

The results indicated that multi-task CNN maintained a correlation between soil properties and predicted more accurately compositional soil properties compared with PLSR and Cubist, which model each property independently. It could also be due to the larger R^2 values achieved by the CNN model.

3.5. Determining important wavelength variables

One of the shortcomings of CNN is that it is difficult to interpret the results from the neural network. We tried to assess how the CNN model used spectral variables for the prediction of soil properties via sensitivity analysis. Only analysis from the one-dimensional spectral model using MIR spectra was shown here.

First, a new data frame was created using 1000 randomly sampled test dataset and averaged. The first five wavelengths of these new averaged spectra were replaced with the actual value from the selected data instead of the average. This simulated spectral data frame was fed into the trained CNN model, and the 1000 realization of soil properties were given. The variance for the prediction of each of the six properties was then calculated as a measure of the importance of the spectral variables. This process was repeated until all the wavelengths had been evaluated.

If the wavelengths that were varied are important in predicting that

particular properties, we expected the variance to be higher. Conversely, if the wavelengths that were changed were not important, then there should be low variance because the rest of the wavelengths' variables are the averages of the test dataset. The variances for the prediction of various soil properties as a function of MIR wavelength were shown in Fig. 11. The sensitivity of spectral variables was unique for each soil property and could be related to the band assignments for the particular wavelengths. Because the CNN model contained several shared layers, it made sense that some of these selected spectral variables overlapped with different level of importance.

4. Discussion

In this study, we utilized a large spectral library from USDA-NRCS KSSL soil database to test different model performance. The CNN model outperformed both Cubist and PLSR models, which was in agreement to results found by Padarian et al. (2019). This study showed that the 1-D spectra input to the CNN model is more effective than converting the spectra to a 2-D spectrogram as suggested by Padarian et al. (2019). The CNN model is able to take multi-dimensional input; thus it can take Vis-NIR and MIR spectra or their combination together. However, this study found that the combination of Vis-NIR and MIR did not result in improved prediction as the prediction using MIR is already excellent.

The combination of high accuracy prediction and the KSSL database resulted in similar, if not higher performance in comparison to other studies. Compared to the study by Viscarra Rossel et al. (2016), our model had better performance (in terms of lower RMSE) for the same

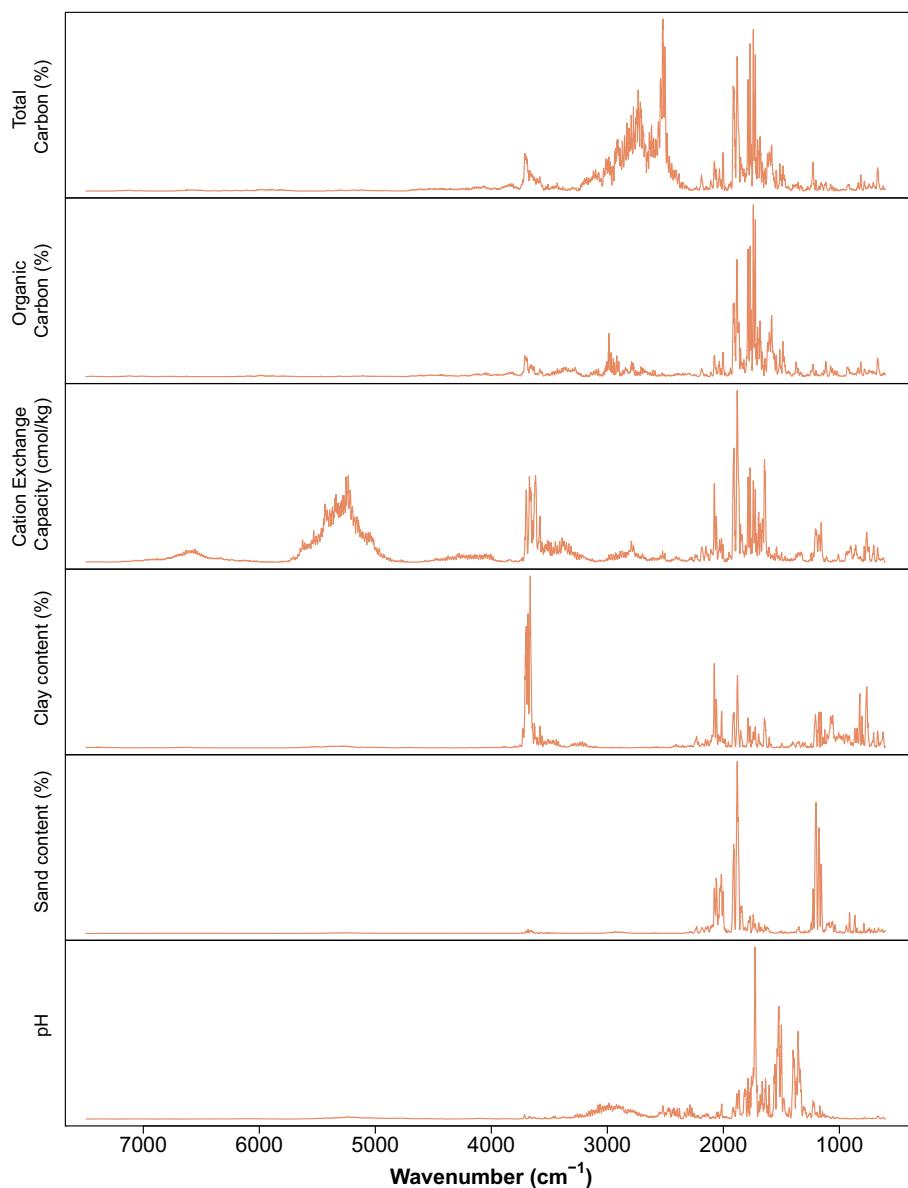


Fig. 11. Plot of average variances for the prediction of various soil properties (total carbon, organic carbon, clay and sand content, cation exchange capacity and pH) using mid-infrared (MIR) spectral and convolutional neural network (CNN) model.

properties predicted except for CEC using vis-NIR spectral data input. Our model achieved RMSE values of 0.60, 5.73, 13.03, 3.92 and 0.85 in comparison to 1.11, 10.26, 18.83, 6.86 and 0.82 respectively for the prediction of organic carbon, clay, sand, CEC and pH respectively. Regression models by [Padarian et al. \(2019\)](#) achieved RMSE values of 1.68, 7.29, 17.00, 6.51 and 0.53 for the same properties respectively. Note that the data we used is limited to SOC < 10% and thus the RMSE on SOC may not be directly comparable to other studies with larger range values.

Compared to another study that utilized a similar spectral library ([Dangal et al., 2019](#)) but a different regression model, we found that our model performed similarly if not better for most soil properties. We compared our CNN model to their machine-based learning (MBL) model using MIR spectral data input. We were able to achieve RMSE value of 0.20, 3.27, 1.98, and 0.30 in comparison to 0.64, 2.47, 2.3, and 0.34 for the prediction of OC, clay, CEC, and pH respectively. Note that our SOC training data has a median of 0.77% and with upper and lower quartile of 0.29–1.59%, while the study by [Dangal et al. \(2019\)](#) used a

wider range SOC data with a median of 1.33% (quartile: 0.42–4.95%).

We also unravel how CNN predicts different soil properties using a sensitivity analysis. The wavenumbers between 2900 and 2500 cm^{-1} and between 1900 and 1600 cm^{-1} were deemed to be important in predicting the total carbon content in the soil. These regions correspond to the alkyl –CH₂ and –CH₃ fundamental stretching frequencies, and aromatic CH-, aromatic C and aromatic C=O ([Forrester et al., 2013](#); [Hobley et al., 2014](#)). These bands were similar to findings by [Terhoeven-Urselmanns et al. \(2010\)](#) and [Tinti et al. \(2015\)](#). Interestingly, for the prediction of organic carbon, only regions between 1900 and 1600 cm^{-1} were used extensively by the model compared to the 2900–2500 cm^{-1} region. The total C prediction included wavenumbers around 2500–2600 cm^{-1} , which was the signature peak of carbonate, but that peak was absent for the prediction of organic C.

For predictions of CEC, the wavenumbers between 1900 and 1600 cm^{-1} were also utilized in addition to regions between 5500 and 5000 cm^{-1} , 3690 to 3300 cm^{-1} , 2150 to 2000 cm^{-1} and 1200 to 1100 cm^{-1} . Regions between 3690 and 3620 cm^{-1} had been reported

to be relevant in predicting clay minerals (such as kaolinite, smectite, and illite) by Hobley et al. (2014) but these regions overlapped with those of water peaks between 3600 and 3000 cm⁻¹ (Forrester et al., 2013). The 5300 to 5000 cm⁻¹ and band at 2127 cm⁻¹ also corresponded to the stretching vibrations and bending vibration of H—O—H bond and combination band of H₂O (Johnston, 2017). These water bonds were probably selected by the CNN model to account for the overlapping regions found between 3690 and 3300 cm⁻¹ regions. The bands near the 1100 cm⁻¹ region corresponded to the presence of quartz mineral (Sila et al., 2016). The clay content prediction utilized similar wavelengths to that of CEC prediction except for the 5300 to 5000 cm⁻¹ regions. Nonetheless, the importance of the wavelengths between 3690 and 3300 cm⁻¹ was more emphasized (higher variance) compared to that between 2150 and 2000 cm⁻¹ region. Important wavenumbers for the sand prediction were between 2100 and 1800 cm⁻¹ and 1250–1100 cm⁻¹. These bands were associated with quartz overtone as well as stretching vibrations of Si—O groups (Tinti et al., 2015). Wavenumbers of 1800–1250 cm⁻¹ was important in predicting soil pH. These bands were associated with mineral features and —OH stretching vibrations (Terhoeven-Urselmans et al., 2010; Sila et al., 2016).

Although CNN outperformed the PLSR and the Cubist model, we do not suggest that it would be suitable for everyone because of the high demand in the amount of data and computing power.

The benefits of the CNN model include:

1. Multi-task learning ability; the model can predict multiple outputs simultaneously while maintaining the correlation among all the outputs and yielding an accurate estimation for compositional properties.
2. Multi-dimensional input utilization; this model can incorporate multiple inputs channels of various dimensions (i.e., one or two-channel).
3. Higher performance; the CNN model can achieve higher accuracy performance in comparison to other regression models, such as Cubist and Partial Least Squares Regression.
4. Interpretability of the important wavelengths variables used in predicting soil properties through sensitivity analysis.

The limitations of the CNN model are:

1. Deep learning is data hungry; it requires a lot of data to be able to make a good prediction.
2. Many hyperparameters; the numbers of the hyperparameters utilized are relatively large, depending on the network architectures. This causes fine tuning process to be a bit tricky to prevent overfitting.
3. Requirement of better computing hardware; because many hyperparameters need to be trained, computing hardware that could handle large data are essential.

5. Conclusions

The availability of a large high-quality KSSL soil and spectral database enabled a good comparison of what could be achieved for prediction using vis-NIR, MIR, and a combination of both spectral data. We demonstrated that the CNN model could provide better performance in comparison to the commonly used regression models (PLSR and Cubist) regardless of the types of spectral data input. Although in this study, the use of combined vis-NIR and MIR with the CNN model did not improve the prediction accuracies much, the finding might be case dependent. As the CNN model is quite flexible, future work could test the combinations of other instruments (e.g., LIBS, XRF, Raman) to predict other soil properties of interest. The use of higher level data fusion (mid- to high-level) could also be explored.

Acknowledgments

This paper is in memory of Jon Hempel, who was the Director of the National Soil Survey Center (NSSC) 2009–2015, and initiated the MIR program at the NSSC in Lincoln. The authors acknowledge the staff at USDA-NRCS, Soil and Plant Science Division and the National Soil Survey Center Kellogg Soil Survey Laboratory (Lincoln, Nebraska, USA) who have collected and analysed the soil samples that constitute the KSSL Soil Characterization database. We acknowledge the Sydney Informatics Hub and the High Performance Computing service at the University of Sydney for providing resources that have contributed to the research results reported within this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.geoderma.2019.06.016>.

References

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D.G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: large-scale machine learning on heterogeneous systems. Software available from tensorflow.org. <https://www.tensorflow.org/>.
- Bellon-Maurel, V., Fernandez-Ahumada, E., Palagos, B., Roger, J.M., McBratney, A., 2010. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *Trac-Trends in Analytical Chemistry* 29 (9), 1073–1081.
- Borras, E., Ferre, J., Boque, R., Mestres, M., Acena, L., Bustos, O., 2015. Data fusion methodologies for food and beverage authentication and quality assessment - a review. *Anal. Chim. Acta* 891, 1–14.
- 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.
- Chollet, F., 2015. Keras. <https://github.com/fchollet/keras>.
- Clairotte, M., Grinand, C., Kouakoua, E., Thebaud, A., Saby, N.P.A., Bernoux, M., Barthes, B.G., 2016. National calibration of soil organic carbon concentration using diffuse infrared reflectance spectroscopy. *Geoderma* 276, 41–52.
- Dangal, S., Sanderman, J., Wills, S., Ramirez-Lopez, L., 2019. Accurate and precise prediction of soil properties from a large mid-infrared spectral library. *Soil Systems* 3, 11.
- Daniel, K.W., Tripathi, X.K., Honda, K., 2003. Artificial neural network analysis of laboratory and in situ spectra for the estimation of macronutrients in soils of Lop Buri (Thailand). *Aust. J. Soil Res.* 41 (1), 47–59.
- Devos, O., Ruckebusch, C., Durand, A., Duponchel, L., Huvenne, J.-P., 2009. Support vector machines (SVM) in near infrared (NIR) spectroscopy: focus on parameters optimization and model interpretation. *Chemom. Intell. Lab. Syst.* 96 (1), 27–33.
- Dollar, P., Tu, Z., Perona, P., Belongie, S., 2009. Integral Channel Features. In: British Machine Vision Conference, London, England.
- Forrester, S.T., Janik, L.J., McLaughlin, M.J., Soriano-Disla, J.M., Stewart, R., Dearman, B., 2013. Total petroleum hydrocarbon concentration prediction in soils using diffuse reflectance infrared spectroscopy. *Soil Sci. Soc. Am. J.* 77 (2), 450–460.
- Griffin, D., Lim, J., 1984. Signal estimation from modified short-time Fourier transform. *IEEE Trans. Acoust. Speech Signal Process.* 32 (2), 236–243.
- He, K.M., Zhang, X.Y., Ren, S.Q., Sun, J., 2015. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. 2015 Ieee International Conference on Computer Vision (ICCV) 1026–1034.
- Hobley, E., Willgoose, G.R., Frisia, S., Jacobsen, G., 2014. Vertical distribution of charcoal in a sandy soil: evidence from DRIFT spectra and field emission scanning electron microscopy. *Eur. J. Soil Sci.* 65 (5), 751–762.
- Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research.
- Islam, K., Singh, B., McBratney, A., 2003. Simultaneous estimation of several soil properties by ultra-violet, visible, and near-infrared reflectance spectroscopy. *Aust. J. Soil Res.* 41 (6), 1101–1114.
- Jaillais, B., Pinto, R., Barros, A.S., Rutledge, D.N., 2005. Outer-product analysis (OPA) using PCA to study the influence of temperature on NIR spectra of water. *Vib. Spectrosc.* 39 (1), 50–58.
- Janik, L.J., Merry, R.H., Skjemstad, J.O., 1998. Can mid infrared diffuse reflectance analysis replace soil extractions? *Aust. J. Exp. Agric.* 38 (7), 681–696.
- Johnston, C.T., 2017. Chapter 9 - infrared studies of clay mineral-water interactions. In: Gates, W.P., Kloprogge, J.T., Madejová, J., Bergaya, F. (Eds.), *Developments in Clay Science*, vol. 8. Elsevier, pp. 288–309.
- Kingma, D.P., Ba, J., 2014. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, Lake Tahoe, Nevada.

- Kuhn, M., Quinlan, R., 2018. Cubist: rule- and instance-based regression modeling. R package version 0.2.2.. <https://CRAN.R-project.org/package=Cubist>.
- Lee, S., Choi, H., Cha, K., Chung, H., 2013. Random forest as a potential multivariate method for near-infrared (NIR) spectroscopic analysis of complex mixture samples: gasoline and naphtha. *Microchem. J.* 110, 739–748.
- Matsugu, M., Mori, K., Mitari, Y., Kaneda, Y., 2003. Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Netw.* 16 (5–6), 555–559.
- McCarty, G.W., Reeves, J.B., Reeves, V.B., Follett, R.F., Kimble, J.M., 2002. Mid-infrared and near-infrared diffuse reflectance spectroscopy for soil carbon measurement. *Soil Sci. Soc. Am. J.* 66 (2), 640–646.
- Mevik, B.-H., Wehrens, R., Liland, K.H., 2018. Pls: partial least squares and principal component regression. R package version 2.7.0. <https://CRAN.R-project.org/package=pls>.
- Minasny, B., McBratney, A.B., 2008. Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemom. Intell. Lab. Syst.* 94 (1), 72–79.
- Niu, Z., Zhou, M., Wang, L., Gao, X., Hua, G., 2016. Ordinal regression with multiple output cnn for age estimation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Nocita, M., Stevens, A., van Wesemael, B., Aitkenhead, M., Bachmann, M., Barthes, B., Ben Dor, E., Brown, D.J., Clairotte, M., Csorba, A., Dardenne, P., Dematte, J.A.M., Genot, V., Guerrero, C., Knadel, M., Montanarella, L., Noon, C., Ramirez-Lopez, L., Robertson, J., Sakai, H., Soriano-Disla, J.M., Shepherd, K.D., Stenberg, B., Towett, E.K., Vargas, R., Wetterlind, J., 2015. Soil spectroscopy: an alternative to wet chemistry for soil monitoring. *Adv. Agron.* 132, 139–159.
- O'Rourke, S.M., Minasny, B., Holden, N.M., McBratney, A.B., 2016. Synergistic use of Vis-NIR, MIR, and XRF spectroscopy for the determination of soil geochemistry. *Soil Sci. Soc. Am. J.* 80 (4), 888–899.
- Padarian, J., Minasny, B., McBratney, A.B., 2019. Using deep learning to predict soil properties from regional spectral data. *Geoderma Regional* 16, e00198.
- Python Software Foundation, 2017. Python language reference. Python Software Foundation <https://www.python.org>.
- Quinlan, J.R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers Inc., San Mateo, California.
- R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Ramsundar, B., Kearnes, S., Riley, P., Webster, D., Konerding, D., Pande, V., 2015. Massively Multitask Networks for Drug Discovery. *arXiv preprint arXiv:1502.02072*.
- Ribeiro, D., Carneiro, G., Nascimento, J.C., Bernardino, A., 2017. Multi-channel convolutional neural network ensemble for pedestrian detection. *Pattern Recognition and Image Analysis (IbPRIA 2017)* 10255, 122–130.
- Rinnan, A., van den Berg, F., Engelsen, S.B., 2009. Review of the most common pre-processing techniques for near-infrared spectra. *Trac-Trends in Analytical Chemistry* 28 (10), 1201–1222.
- Ruder, S., 2017. An Overview of Multi-Task Learning in Deep Neural Networks. *CoRR (abs/1706.05098)*.
- Sarathjith, M.C., Das, B.S., Wani, S.P., Sahrawat, K.L., 2016. Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma* 267, 1–9.
- Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 36 (8), 1627–1639.
- Sermanet, P., Eigen, D., Zhang, X., Mathieu, M., Fergus, R., LeCun, Y., 2013. Overfeat: Integrated Recognition, Localization and Detection Using Convolutional Networks. *arXiv preprint arXiv:1312.6229*.
- Shepherd, K.D., Walsh, M.G., 2002. Development of reflectance spectral libraries for characterization of soil properties. *Soil Sci. Soc. Am. J.* 66 (3), 988–998.
- Shibusawa, S., Imade Anom, S.W., Sato, S., Sasao, A., Hirako, S., 2001. Soil mapping using the real-time soil spectrophotometer. In: Grenier, G., Blackmore, S. (Eds.), *ECPA. Third European Conference on Precision Agriculture*. Agro Montpellier, 1. Montpellier, France, pp. 497–508.
- Sila, A.M., Shepherd, K.D., Pokhriyal, G.P., 2016. Evaluating the utility of mid-infrared spectral subspaces for predicting soil properties. *Chemom. Intell. Lab. Syst.* 153, 92–105.
- Soil Survey Staff, 2014. Kellogg soil survey laboratory methods manual. In: Soil Survey Investigations Report No. 42. N. R. C. S. U.S. Department of Agriculture.
- Soriano-Disla, J.M., Janik, L.J., Rossel, R.A.V., Macdonald, L.M., McLaughlin, M.J., 2014. The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl. Spectrosc. Rev.* 49 (2), 139–186.
- Stenberg, B., Viscarra Rossel, R.A., Mouazen, A.M., Wetterlind, J., 2010. Chapter five - visible and near infrared spectroscopy in soil science. In: Sparks, D.L. (Ed.), *Advances in Agronomy*. vol. 107. Academic Press, pp. 163–215.
- Terhoeven-Urselmans, T., Vagen, T.G., Spaargaren, O., Shepherd, K.D., 2010. Prediction of soil fertility properties from a globally distributed soil mid-infrared spectral library. *Soil Sci. Soc. Am. J.* 74 (5), 1792–1799.
- Terra, F.S., Dematté, J.A.M., Rossel, R.A.V., 2015. Spectral libraries for quantitative analyses of tropical Brazilian soils: comparing Vis-NIR and mid-IR reflectance data. *Geoderma* 255, 81–93.
- Terra, F.S., Rossel, R.A.V., Dematté, J.A.M., 2019. Spectral fusion by outer product analysis (OPA) to improve predictions of soil organic C. *Geoderma* 335, 35–46.
- Tinti, A., Tugnoli, V., Bonora, S., Franciosi, O., 2015. Recent applications of vibrational mid-infrared (IR) spectroscopy for studying soil components: a review. *J. Cent. Eur. Agric.* 16, 1–22.
- Vesela, A., Barros, A.S., Syntysa, A., Delgadillo, I., Copikova, J., Coimbra, M.A., 2007. Infrared spectroscopy and outer product analysis for quantification of fat, nitrogen, and moisture of cocoa powder. *Anal. Chim. Acta* 601 (1), 77–86.
- Viscarra Rossel, R.A., Behrens, T., Ben-Dor, E., Brown, D.J., Dematte, J.A.M., Shepherd, K.D., Shi, Z., Stenberg, B., Stevens, A., Adamchuk, V., Aichi, H., Barthes, B.G., Bartholomeus, H.M., Bayer, A.D., Bernoux, M., Bottcher, 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., Morras, 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.
- Wang, D., Chakraborty, S., Weindorf, D.C., Li, B., Sharma, A., Paul, S., Ali, M.N., 2015. Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. *Geoderma* 243–244, 157–167.
- Xu, Y.T., Ma, J.S., Liaw, A., Sheridan, R.P., Svetnik, V., 2017. Demystifying Multitask Deep Neural Networks for Quantitative Structure-Activity Relationships. *Journal of Chemical Information and Modeling* 57 (10), 2490–2504.
- Xuemei, L., Hailiang, Z., Xudong, S., 2010. NIR sensitive wavelength selection based on different methods. In: 2010 International Conference on Mechanic Automation and Control Engineering, 26–28 June 2010.
- Zhang, Y., Yang, Q., 2017. A survey on multi-task learning. *arXiv e-prints*. Retrieved from: <https://ui.adsabs.harvard.edu/abs/2017arXiv170708114Z>.
- Zou, X.B., Zhao, J.W., Povey, M.J.W., Holmes, M., Mao, H.P., 2010. Variables selection methods in near-infrared spectroscopy. *Anal. Chim. Acta* 667 (1–2), 14–32.