Plant Soil (2025) 513:3157–3179
https://doi.org/10.1007/s11104-025-07306-9

METHODS PAPER

Spectral clustering analysis: discrimination of grass-herb roots and live-dead roots in VISNIR and SWIR regions

Pavel Baykalov · Gernot Bodner · Ivika Ostonen · Boris Rewald

Received: 3 April 2024 / Accepted: 13 February 2025 / Published online: 24 February 2025

Abstract

Background and aims Hyperspectral imaging is becoming a key, high-throughput technique in plant research. However, its application to roots has not yet received sufficient attention. The aims of this study are to identify spectral features that distinguish fine roots from soil, non-woody roots of different species, and dead from living roots, and to identify appropriate analytical techniques.

Methods Roots of *Alopecurus pratensis* (meadow foxtail) and *Urtica dioica* (nettle) and the rhizosphere were imaged in rhizoboxes in the wavelength range 400–1700 nm, covering both visible near- (VISNIR) and shortwave infrared (SWIR) regions. Principal Component Analysis, K-means clustering, and Generalised Linear Model, Partial Least Squares Discriminant Analysis, and Distributed Random Forest models were used to classify groups. Wavebands critical for classification were identified.

Results Our results demonstrate the intricate nature of spectra clustering, highlighting the challenges in the VISNIR range and the promise of SWIR data for enhanced separability. While species differentiation is challenging, the determination of the living conditions of the roots is possible within the SWIR range. The analysis reveals the significance of specific spectral regions, notably those associated with water content and senescence, in distinguishing between living and dead roots. Water content regions (mainly 1245 nm and 1450 nm) were most important in discriminating between roots and soil.

Responsible Editor: Doris Vetterlein.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1007/s11104-025-07306-9.

P. Baykalov · B. Rewald (✉)
Vienna Scientific Instruments GmbH, Alland, Austria
e-mail: boris.rewald@mendelu.cz

P. Baykalov
e-mail: p.baykalov@vienna-scientific.com

P. Baykalov
Institute of Forest Ecology, University of Natural Resources and Life Sciences, Vienna (BOKU), Vienna, Austria

G. Bodner
Institute of Agronomy, University of Natural Resources and Life Sciences, Vienna (BOKU), Tulln, Austria
e-mail: gernot.bodner@boku.ac.at

I. Ostonen
Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia
e-mail: ivika.ostonen@ut.ee

B. Rewald
Faculty of Forestry and Wood Technology, Department of Forest Protection and Wildlife Management, Mendel Universityin Brno, Brno, Czech Republic

Conclusions This study highlights the potential of spectral analysis, particularly in the SWIR region, for distinguishing roots by species and vitality. Further efforts are needed to develop robust methods for mixed data sets containing roots of different species and degrees of vitality.

Keywords Infrared spectrometry · Root taxa · Root vitality · Species discrimination · Spectral analysis · Mixed plant communities

Abbreviations

DRF	Distributed Random Forest
GLM	General Linear Model
K-means	"K" – number of clusters
MCC	Matthews Correlation Coefficient
PLS-DA	Partial Least Squares – Discriminant Analysis
SWIR	Short Wave Infrared
VISNIR	Visible Near Infrared

Introduction

Plant tissues are composed of various chemical compounds arranged in different structures, and these components, together with water, can interact with electromagnetic radiation / photons of different energy levels (da Luz and Crowley 2007; Knipling 1970; Turker-Kaya and Huck 2017). While light in the visual (VIS) spectrum (400–700 nm) is commonly used for 'normal' RGB imaging, near infrared (NIR; 700 ~ 1000 nm) and short-wave infrared (SWIR; ~ 1000–2500 nm) spectroscopy (Mishra et al. 2017) is increasingly used in plant research for purposes such as stress (Zahir et al. 2022) and disease/decay identification (Ali et al. 2019; Min et al. 2023), NDVI computation (Calera et al. 2001), canopy structure assessment (Serbin and Townsend 2020), and nutrient content estimation (Johnson et al. 2021). Beyond plant science, many other materials are cost-effectively analysed by infrared spectroscopy, leading, for example, to the increasing development of soil spectral libraries (Dangal et al. 2019; Rossel et al. 2008).

Plant communities are often composed of different species, and this organisation subsequently impacts ecosystem characteristics (Violle et al. 2007). Transforming individual species-level traits into community function involves relating trait values to abundance (Violle et al. 2007). While the value of diverse, mixed plant communities is increasingly recognised for sustaining functions and yields of (semi-) natural grasslands and forests (Bhattarai et al. 2023; Blondeel et al. 2024; Lindenmayer et al. 2000) but also in agriculture (Cappelli et al. 2022; Udawatta et al. 2019; Thrupp 2000) under progressive climate change, studies in mixed communities are considerably hampered by the lack of high-throughput methods for species identification and quantification. Aboveground, an increasing number of studies classify plant functional types or species based on the unique waveband patterns created by leaf reflectance or absorption (Buitrago et al. 2018; Paz Kagan et al. 2017; Zhao et al. 2016; Zhou et al. 2016). In contrast, spectroscopic methods on roots have been mainly used to determine chemical composition in dried and ground samples, such as starch (Nkouaya Mbanjo et al. 2022), lignin (Elle et al. 2019), and cellulose contents, as well as Carbon–Nitrogen ratios (White et al. 2011), often using Fourier-transform infrared (FTIR) spectroscopy (Jaggi and Vij 2006). To date, only a few studies have explored the potential to identify plant species using the spectral patterns obtained from their roots (Naumann et al. 2010; Rewald and Meinen 2013; Straková et al. 2020; Tong et al. 2016). Based on chemical differences, researchers have used dried and ground root material to disentangle root system biomass of different crop or weed species growing in mixtures, although labour-intensive calibration with artificial mixtures of two or three species was required (Meinen and Rauber 2012; Naumann et al. 2010). In certain instances this has been achieved even with species of the same genus, e.g. motivated by fraud detection; for example, Laasonen et al. (2002) demonstrated that IR spectroscopy (1640–1754 nm) is a reliable method for rapid identification of *E. purpurea* roots in the pharmaceutical industry. However, challenges remain when trying to distinguish intact roots by species in soil and/or rinsed from soil samples (Freschet et al. 2021; Rewald et al. 2012). In particular, minirhizotron or rhizobox studies, using repeated imaging to determine root growth and turnover at transparent soil-interfacing surfaces, not only benefit from automatic root detection/segmentation (Baykalov et al. 2023; Khoroshevsky et al. 2024),

but the methodology could be greatly enhanced by the ability to classify detected roots. However, direct interference from soil spectra and water droplets, particularly at the transparent interface of MR tubes and rhizoboxes/root windows, can be expected to affect spectral imaging of roots to an extent at least similar to that of dust particles deposited on leaves (Lin et al. 2021). In addition, many spectrometric cameras still have lower resolutions than RGB cameras (Zhao et al. 2022), which is particularly challenging for imaging fine roots, which can be as thin as 50 µm (Pierret et al. 2005).

Besides species identification, distinguishing between living and dead roots is another important methodological constraint in biomass sampling (Freschet et al. 2021; Rewald and Meinen 2013) and minirhizotron studies (Rewald and Ephrath 2013). Knowledge of the proportion of root systems actively contributing to water (Vetterlein and Doussan 2016) and nutrient (Ehrenfeld et al. 1997) uptake is key to determining the spatiotemporal patterns of root system function. Furthermore, living and dead roots are currently categorised by humans, often based on colour (e.g. white vs. black) or turgescence (e.g. turgescent vs. shriveled). A less subjective and automated distinction between living and dead roots would therefore help to determine more accurate root turnover rates (Freschet et al. 2021; Picon-Cochard et al. 2009). However, few studies have yet used IR spectroscopic methods to distinguish between living and dead plant tissue. For example, vegetation indices based on hyperspectral data (from 350 to 2500 nm) in mixed grasslands have achieved high correlations (> 0.9 R^2) in estimating the proportion of dead shoots (Yang and Guo 2014), suggesting strong and persistent spectral differences between living and dead leaves (i.e. via chlorophyll content). A recent study on grasslands showed that spectral complexity is sensitive to living and dead plant biomass (Rossi et al. 2022). For roots, Picon-Cochard et al. (2009) were able to predict the proportion of dead versus living roots of five grass species with an error of 15% using washed and dried samples that were manually sorted into vitality classes before drying. Similarly, hyperspectral imaging (in the VISNIR range) has been used to discriminate between dead and living tree roots by extracting pixel values from poplar (Nakaji et al. 2008). Hyperspectral Imaging (HSI), a cutting-edge technology that integrates imaging and spectroscopy, has the potential to surpass NIR spectroscopy. Unlike NIR spectroscopy, HSI scans entire samples, allowing the averaging of numerous narrow spectra to create representative samples for the analysis of highly heterogeneous objects of study, such as soils (Pudełko et al. 2020). Major applications of this technology include environmental monitoring (Stuart et al. 2019), precision agriculture (Sethy et al. 2022), food quality assessment (Min et al. 2023), medical diagnostics (Fei 2019), and mineral exploration (Bedini 2017). Hyperspectral imaging of rhizobox systems has previously been proposed to distinguish rhizosphere components such as soil, decomposed leaf litter, or living and dead roots (Nakaji et al. 2008; Pierret 2008), and has even been tested to segment roots from the surrounding soil (Bodner et al. 2018). The study of plant roots using hyperspectral imaging thus opens up new possibilities for the non-destructive study of soils, roots, and their symbionts. Machine learning (ML) algorithms applied to spectroscopic data provide a rapid and non-invasive method of analysing the composition of a target sample by mapping absorbance or reflectance values to the desired output, providing an efficient means of spectral data analysis (Zhang et al. 2022). The synergistic integration of hyperspectral imaging with ML methods has significantly advanced the classification and analysis of complex hyperspectral images (Ozdemir and Polat 2020). Two general ML approaches can be distinguished: unsupervised and supervised learning. Unsupervised methods are valuable for exploring inherent patterns in spectral data without prior labels, providing insight into the natural grouping of the features. In contrast, supervised methods use known groups to classify the data and identify the key features that drive group differentiation. Both supervised classification and unsupervised clustering are used for qualitative analysis of hyperspectral data (Li et al. 2022).

Root and rhizosphere research is hampered by the lack of non-destructive methods to distinguish visually similar roots of different species and vitality stages, up to and including senescent roots, in minirhizotrons and rhizoboxes. In this study, we first determine the spectral differences, in both the VISNIR and SWIR regions, that set apart fine roots of a common grass (*Alopecurus pratensis*) and a common herb (*Urtica dioica*) species from average soil spectra. Using unsupervised ML, we then investigate whether the spectral signatures are already

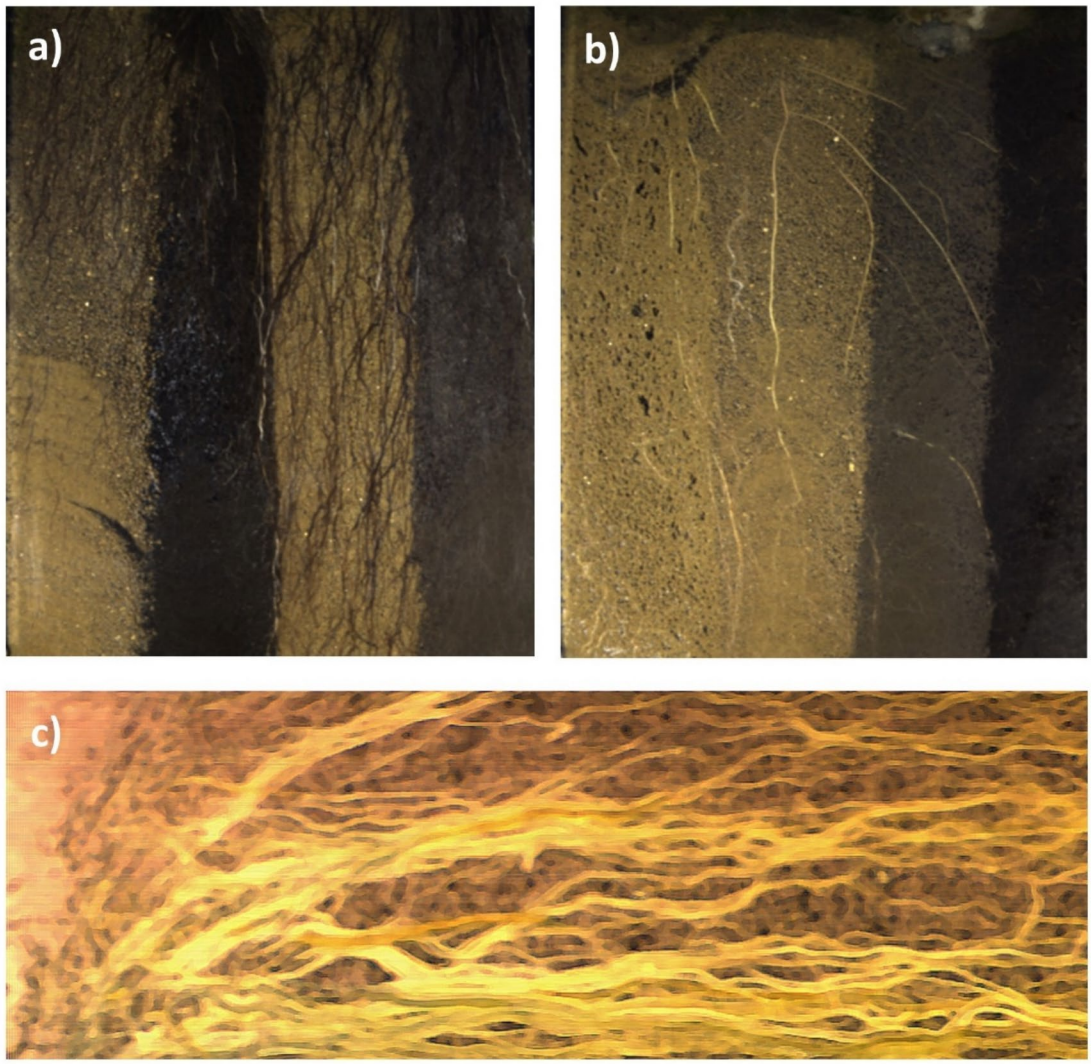

Fig. 1 RGB images of **a**) *Alopecurus pratensis* and **b**) *Urtica dioica* captured by the SPECIM IQ; different rhizoboxes present a different permutation of the four soil types, e.g. in b) from left to right: quartz sand, light subsoil, dark topsoil, peat. Subpanel **c**) depicts a synthetic image used for SWIR labelling purposes. The synthetic image was constructed by combining three different wavebands (ImSpector N17E), see text for details

present in the studied categories and could be used to classify roots of the two plant species, as well as to distinguish between living and dead roots. Three supervised machine learning methods were used to facilitate the discrimination of roots from the soil, roots of different species, or viability groups, with the aim of exploring the level of accuracy in determining groups and identifying the wavebands that contribute most to group classification.

Materials and methods

Experimental set-up

Six rhizoboxes (Fig. 1) were prepared, each measuring 30×20×3 cm (H×W×D), made of polymethylmethacrylate (PMMA; 3 mm) with a transparent front panel, and black rear, sides, and bottom (Vienna Scientific Instruments, Alland, AT). Four different soil types were

used, differing widely in organic content, colour, and texture, to create heterogeneity in soil spectral patterns. These included i) commercial garden peat substrate, ii) soil organic matter-rich, dark (A horizon of Chernozem), and iii) soil organic matter poor, light (subsoil of Chernozem) arable soils from eastern Austria (Tulln an der Donau, Lower Austria), and iv) pure quartz sand. The soils were sieved to a particle size of ≤ 2 mm. See Supplementary Table S1 for details. The rhizoboxes were filled with the soils in four vertical strips of ~ 5 cm width, using pieces of cardboard as dividers during rhizobox filling, to realistic densities of 0.5–2.6 g cm^{-3} (depending on soil type). Different boxes were filled with different permutations of the four soil types to ensure that some soils would not only hold lateral roots, while others would be consistently more densely rooted. Subsequently, two plant species were entered in three rhizoboxes each: *Alopecurus pratensis* L. (AP) and *Urtica dioica* L. (UD). AP, the meadow foxtail, is a perennial grass of the Poaceae family, known for its preference for moist, nutrient-rich soils (Kreyling et al. 2012). UD, stinging nettle, is a perennial herb of the Urticaceae family, commonly inhabits weakly acidic or weakly alkaline, richly fertile, and moist soils, occurring throughout Europe – particularly on disturbed sites (Taylor 2009). 16 seeds were planted per rhizobox; 2 weeks after germination seedlings were thinned to approximately 6 plants per rhizobox and grown for 5 months in a growth chamber (Fitotron HGC, Weiss–Gallenkamp, UK). The chamber was set to a temperature of 25 °C with a relative humidity of 60%. The day/night cycle in the chamber was set to 10/14 h, respectively; PAR was ~ 1450 µmol m^2 s^{-1} at mid-chamber height (Fitzky et al. 2023). Manual irrigation was performed every two to three days with tap water to field capacity; no fertiliser was added. The drainage holes at the bottom of the rhizoboxes were partially sealed to reduce moisture loss.

After taking rhizobox images (see below) of vigorous-looking plants with well-established root systems in mid-March, i.e. ~ 5 months after sowing (Fig. 1), the plants were killed by cutting the shoots. The rhizoboxes were then placed in a dark room at 25 °C; no algae growth was observed on the transparent rhizobox panel. The root systems were imaged again at the end of April, 42 days after the above-ground part of the plants had been cut. During this time, the rhizoboxes were watered once a week, to ensure constant soil moisture between imaging dates. Images taken in March and April are referred to in the results as 'living' and 'dead' root datasets, respectively.

Imaging and pre-processing

Two camera devices were used in this study to cover a wide wavelength range from the VIS (400 nm) to the SWIR (1700 nm) spectra. The cameras were: 1) SPECIM IQ (Specim, Spectral Imaging Ltd, Oulu, FI): This camera captures 204 wavebands from 400 to 1000 nm (VISNIR range), with a spectral resolution of 7 nm. The SPECIM IQ camera produces an image of the entire rhizobox (30 × 20 cm), albeit at a lower resolution than the ImSpector N17E (see below). Each image taken by the SPECIM IQ has a size of 512 × 512 pixels. The hyperspectral data was saved in .hdr and .raw formats. In addition, an RGB image (.png format), corresponding to the hyperspectral data, was taken as a reference. 2) ImSpector N17E (Specim, Spectral Imaging Ltd, Oulu, FI): This camera captures 256 wavebands from 900 to 1700 nm (SWIR range), with a spectral resolution of 5 nm. Instead of capturing a complete image of the rhizobox, the camera generates 7 vertical strips/sections of the rhizobox, each stored as a separate file. Each section is 320 × 3000 pixels (H × W), with 256 wavebands, and is saved in.sif format. In total, 7 hyperspectral images were taken from each of the 6 rhizoboxes. For greater versatility, efficiency, and compatibility between programming languages, the data was converted from.sif to HDF5 (.h5) format.

The hyperspectral data was normalized using white and dark references, according to the following equation (Bodner et al. 2018):

$$Normalized = \frac{R_{i,\lambda} - D_{i,\lambda}}{W_{i,\lambda} - D_{i,\lambda}} \quad (1)$$

where $R_{i,\lambda}$ is the raw data and $D_{i,\lambda}$ is the dark reference, $W_{i,\lambda}$ is the white reference, normalised at waveband λ and pixel i in the dataset. The data were also scaled from 0 to 1 (float) to facilitate further analysis using algorithms and statistical methods (Martinez-Vega et al. 2022):

$$Scaled(x_{\lambda,i}) = \frac{x_{i,\lambda} - \min(x_\lambda)}{\max(x_\lambda) - \min(x_\lambda)} \quad (2)$$

where $x_{i,\lambda}$ denotes a data point at waveband λ and pixel i in the dataset. Normalisation and scaling were applied to the entire rhizobox image or to the 7 strips/sections per rhizobox in the VISNIR and SWIR datasets, respectively.

The data were sparsely annotated by selecting regions of interest (ROIs) using a custom program written in Python (Baykalov 2024). Each sparse label or ROI (1–13 px) appeared on a mask that was saved for later processing. The pixels for the analysis were extracted from these labels. Due to the significant difference in resolution between the VISNIR and SIWR cameras, accurate detection of very fine roots in the low-resolution VISNIR data was limited and larger root diameters were labelled—creating a potential bias in the recorded spectra towards larger fine roots. However, it was beyond the scope of this study to evaluate the spectra of different root development/branching stages. To ensure consistency of the labels for both the March (living) and April (dead) datasets, the roots/soil from the March images were first labelled to generate the masks on living roots, and then the reference images were aligned using a custom algorithm consisting of the ORB feature detector (Rublee et al. 2011), the descriptor matcher, and the RANSAC homology finder (Shen et al. 2020; Vinay et al. 2015) from the 'cv2' library (Bradski 2000) on Python v. 3.7.13. The masks for the dead roots and soil in April were generated by matching the March masks to the same homology as the reference images.

VISNIR (visual and near infrared) data processing

A custom labelling tool (Baykalov 2024), VISNIR_labelling_tool.py, published on GitHub, was used to directly label the root pixels on an RGB image (.png format file) acquired directly with SPECIM IQ. This approach was chosen due to the direct correspondence between RGB and hyperspectral data (i.e. same orientation and image size). Labelling the roots on the RGB images proved to be much more straightforward than trying to identify specific wavebands where the roots are clearly visible. For each individual rhizobox, a subset of the root and soil pixels (ROIs: 15–260 px in size) were labelled and stored in different masks, one mask for roots and another for soil. All four soil types and the white stones present in the quartz sand were labelled as soil; soil types were not distinguished but labelled in approximately equal amounts. RGB images of the rhizoboxes were used as a reference for the date (March–April) alignment of the masks. For the analysis, the annotated pixels were extracted by filtering the hyperspectral data through the masks (Online Resource 3, Supplementary Fig. 1). These annotated pixels were then stored in a matrix with 204 columns, corresponding to the waveband number, and a number of rows corresponding to the number of pixels in the mask. To represent the inherent variability of soil and root (age), a minimum number of 500 pixels per category was selected.

SWIR (short wave infrared) data processing

The SWIR data exhibited a distinct landscape and conditions that required additional pre-processing steps. First, a 200-pixel wide area was cropped from the upper part of the data, primarily containing the shoots of the plants. After cropping, a waveband filtering process was applied, removing the first 37 bands and retaining 219 wavebands as the final result. Reference wavebands for date (March–April) alignment were generated from the 6th waveband (1034–1035 nm) after filtering, as this band provided the clearest image compared to the first five bands. This reference waveband was chosen by selecting the first clearest band starting from the beginning. The reference wavebands were saved in .npy format, as the alignment algorithm was implemented and executed in Python. To reduce stripe noise in each waveband, median filtering with a 5×5 diamond kernel was applied using the 'mmand' R library (Clayden 2020). Due to the presence of 7 stripes/sections, normalisation was performed within each rhizobox and day to standardise the data. The within-rhizobox normalisation facilitated the scaling of the data.

Similar to the approach used for VISNIR data, a custom tool (SWIR_labelling_tool.py) was used for labelling (Baykalov 2024). To label the SWIR data, a synthetic image was constructed by combining three different wavebands (60, 70, and 150, corresponding to 1135, 1189, and 1457 nm, respectively; Fig. 1c). These specific wavebands were chosen because they showed good quality with clear images. The main criterion for labelling was a clear distinction between roots and soil, validated by visual inspection of the RGB image of the rhizobox. Only a small random portion of the visible root length, ~5%, was labelled.

The four soil types present in the study were considered in the labelling process. The soil type was, however, not considered in the spectral analysis, all soil types were combined into one cluster corresponding to an average soil spectrum. The small bright stones found predominantly in the quartz sand were selected as soil to augment data diversity. Similar to the VISNIR data processing, the ROI pixels were extracted by filtering the data through the mask and stored in a two-dimensional matrix (Online Resource 3, Supplementary Fig. 1). The matrix consisted of 219 columns corresponding to the wavebands remaining after the previous filtering step. Due to the higher resolution of the SWIR images, the number of labelled pixels was higher than for the VISNIR data, with a minimum of 20,000 pixels per category.

Additionally, based on the observations from the first analysis (of March vs April data within the same species), which included the entire spectra ~ 1000–1700 nm, a region of high differences was detected. This region is located between 1450 and 1700 nm, consequently a subset of data within the same range was made and named as '1450–1700' ($AP_{1450–1700}$ and $UD_{1450–1700}$); these data were separately clustered and classified only for the living/dead distinction.

Spectral data analysis

Data analysis was performed separately for the VISNIR and SWIR data sets. Image alignment between the VISNIR data and SWIR data was not achieved due to the large difference in resolution (~ 22 pixel cm^{-1} of VISNIR versus ~ 100 pixel cm^{-1} of SWIR), resulting in mismatched labels.

Visualization of reflectance patterns

Six types of data points were analysed, grouped in three pairs: Roots:Soil, AP:UD roots (species), and Living:Dead roots. For each group, the mean pixel values for each waveband were computed, resulting in a mean spectrum for that particular group. The standard error was then added or subtracted from the mean to produce a range of values reflecting the variability of the pixel values. Principal component analysis (PCA) was used as a dimensionality reduction technique to visualise the ROI pixels for both VISNIR and SWIR data. PCA was applied separately for roots-soil, species, and live-dead (by date) visualisation. The R libraries 'ggplot' (Wickham 2016) and 'ggfortify' (Li 2016; Tang et al. 2016) were used to generate plots.

Clustering and classification algorithms

An unsupervised clustering method was used to identify potential inherent groups within the data. These clusters were then compared with the actual groups, which were either roots or soil, species, or date (live-dead), to assess clustering performance. K-means (Forgy 1965), a widely used unsupervised machine learning algorithm, was used to partition the dataset into 'k' clusters, where 'k' is the number of actual groups, by iteratively assigning data points to the nearest centroid and recalculating centroids based on the mean of the points within each cluster. All hyperparameters were set as default, except for the random seed, which was varied to determine the maximum Matthews Correlation Coefficient (MCC; see below). To ensure robustness, K-means was executed 500 times to identify the best MCC score and to generate the distribution of MCC values (see Online Resource 3).

Three supervised machine learning methods were used to facilitate the detection of groups, with the aim of exploring two key aspects: i) the level of accuracy in determining the groups and ii) the identification of wavebands providing the most valuable information in group classification. 1) Generalised Linear Model (GLM), a statistical framework extending the traditional linear regression model to accommodate various types of data distributions. By specifying an appropriate link function (in this case multinomial) and error distribution, GLM enables the analysis of non-normally distributed dependent variables (Nelder and Wedderburn 1972). 2) The Partial Least Squares Discriminant Analysis (PLS-DA) model, popular for modelling spectral data (Ali et al. 2019; Lin et al. 2012; Yu et al. 2018), is a multivariate statistical technique used for supervised classification, specifically designed to distinguish between predefined groups or classes within a dataset by maximising the separation between them, considering both predictor variables and class information (Lee and Liong 2018). A similar method but for regression instead of classification, is widely used and known in the field of spectral analysis (Bodner et al. 2018; Straková et al. 2020; Tong et al. 2016; Zahir et al. 2022). 3) Distributed Random Forest (DRF) (LeDell and Poirier 2020), an ensemble

learning method designed for distributed computing environments. In DRF, multiple decision trees are constructed independently on different subsets of the data, and their predictions are combined to generate a more accurate and robust model (Breiman 2001).

GLM, PLS-DA and DRF models were run with default hyperparameter values (H2O.ai 2023a; Kuhn 2008; LeDell and Poirier 2020). After training, each model allows retrieving the sorted priority list of the variables used for the classification (H2O.ai 2023b). The first 20 important variables, corresponding to the wavebands, were extracted from each cross-validation training and then retained. The important wavebands identified by each algorithm over five cross-validations were combined, resulting in a total amount of > 300 accumulated wavebands. Only wavebands that appeared at least five times in this combined set were selected for further analysis, reducing the number of wavebands to 52–72; identified, combined wavebands are reported per discrimination task (Online Resource 1). Cross-validation is an essential technique that facilitates the testing of algorithms using the same dataset while providing an assessment of the average performance on unseen data. For this study, a five-fold cross-validation approach was employed. This involved randomly dividing the data into five subsets, training the algorithms on four of these subsets, and evaluating them on the fifth subset. By repeating this process for all possible combinations, the validation results were averaged and reported along with the corresponding standard error (SE). Additional testing was conducted on the data per month. Each model underwent training using the March training set data and was subsequently evaluated on the entire dataset from April, rather than the validation set.

The 'caret' (Kuhn 2008) and 'h2o' (LeDell and Poirier 2020) packages for R (R Core Team 2022) were used for both unsupervised and supervised algorithms. A computer equipped with an NVIDIA GeForce RTX 2080 with Max-Q Design GPU and 24 GB of memory was used for training. The computer had an Intel i7 Intel processor and 32 GB of RAM.

Evaluation of the algorithms

Two metrics were used to evaluate the algorithms: regular accuracy and the Matthews Correlation Coefficient (MCC). Accuracy provides a broad overview of the classification performance (Grandini et al. 2020), representing the ratio of correctly labelled instances to the total number of instances.

$$Accuracy = \frac{Number\ of\ correctly\ labelled\ instances}{Total\ number\ of\ instances} \quad (3)$$

In contrast, MCC provides a superior measure for evaluating multi-class classification that is robust to unbalanced data (Boughorbel et al. 2017; Chicco et al. 2021). It expresses the degree of correlation between two categorical random variables (Grandini et al. 2020). The MCC used in the analysis is from the R package 'mltools' (Gorman 2022).

Results

Differences between mean root and soil spectra

In the VISNIR analysis, there was a high degree of overlap between root and soil spectral features (Figs. 2, 3). The PCA plots and spectral pattern illustrate the intricate intertwining of soil and root spectra; the second component of the PCA falls significantly short of clearly categorising the two components, resulting in a modest explanatory power of 9–13%. The mean spectra of the two groups are very close, with the largest differences at the end of the range for living roots and in the middle (around 600 nm) for dead roots (Fig. 3; Online Resource 5, Supplementary Fig. 5). K-means clustering yielded only moderate accuracies and very low MCCs (Table 1). In terms of supervised classification, the tree-based algorithm DRF yielded the best results for the separation of Roots:Soil March pixels based on VISNIR, with accuracies between 0.9–0.94 and MMC of 0.79–0.87 (Table 2). In contrast, GLM and PLS-DA achieved results of 0.71–0.81 and 0.38–0.64 for accuracy and MMC, respectively (Table 2), with GLM performing better than PLS-DA. A subsequent test on dead root data (Roots:Soil) revealed a risk of overfitting within the DRF model, while GLM and PLS-DA maintained a consistent range of accuracy (Table 3). The key to VISNIR-based classification appears to be centred on wavebands located in the spectral regions around ~ 600 nm, as well as with ~ 810–850 nm (weaker), and particularly from ~ 950 to 1000 nm (Fig. 3, grey lines; Online Resource 2, Supplementary Table 2).

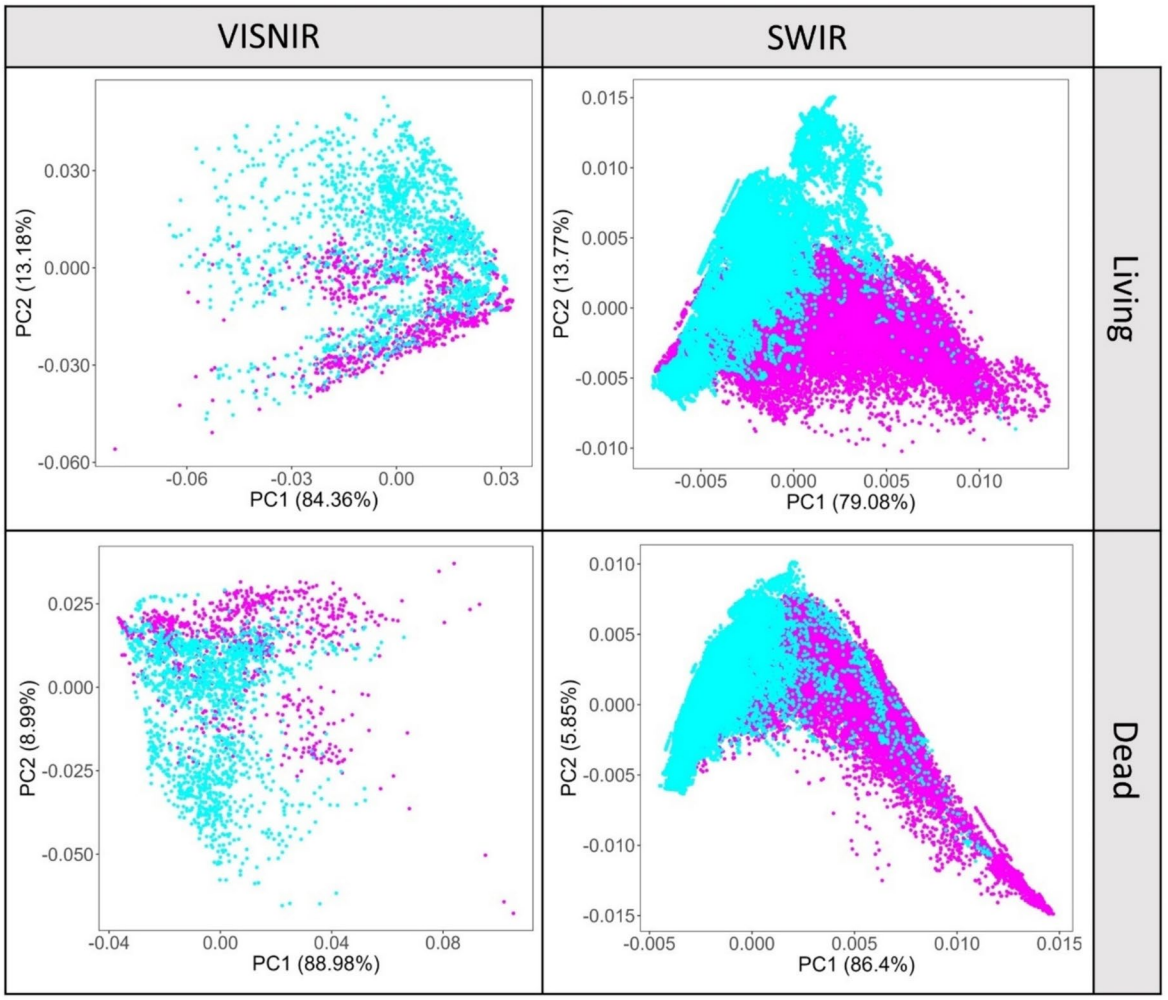

Fig. 2 PCA plots for soil (magenta) and root (cyan) spectral visualisation. Soil spectra are derived as the mean of four soil types (see text for details), root spectra are the mean of living *Alopecurus pratensis* and *Urtica dioica* roots (March) and dead roots of both species (April). Spectral range is visual-near infrared (VISNIR) or short-wave infrared (SWIR). VISNIR data are less dense due to the lower camera resolution

Similar to the VISNIR data, the PCAs for the SWIR data (Fig. 2) do not conclusively show a clear separation between roots and soil. In contrast, however, k-means clustering of SWIR data shows significantly higher MCC (0.51–0.59) (Table 1; Online Resource 4, Supplementary Fig. 2), accompanied by accuracies of around > 0.7 (Table 1). Exploring the spectral domain for wavebands that best discriminate between root and soil instances, average soil pixels have a reduced reflectance of around 1000–1250 and 1400–1600 nm (Fig. 3), compared to root pixels. Interestingly, the same region of 1450–1600 nm, which had more overlapping (between living root and soil spectra) in March, allowed for a higher means of separation of dead roots from the soil in April (Online Resource 5, Supplementary Fig. 5). Regarding the supervised classification, we observe a recurrent pattern of very good accuracies/MMC for the DRF algorithm, and a slightly less good performance of GLM (Table 2). Notably, the overall MCC values in SWIR exceed those in VISNIR for all classification algorithms. The key wavebands identified by the different algorithms for classification (approx. 1200–1260 nm, 1415–1440 nm; Fig. 3; Online Resource 2, Supplementary Table 3) show minor discrepancies between imaging dates.

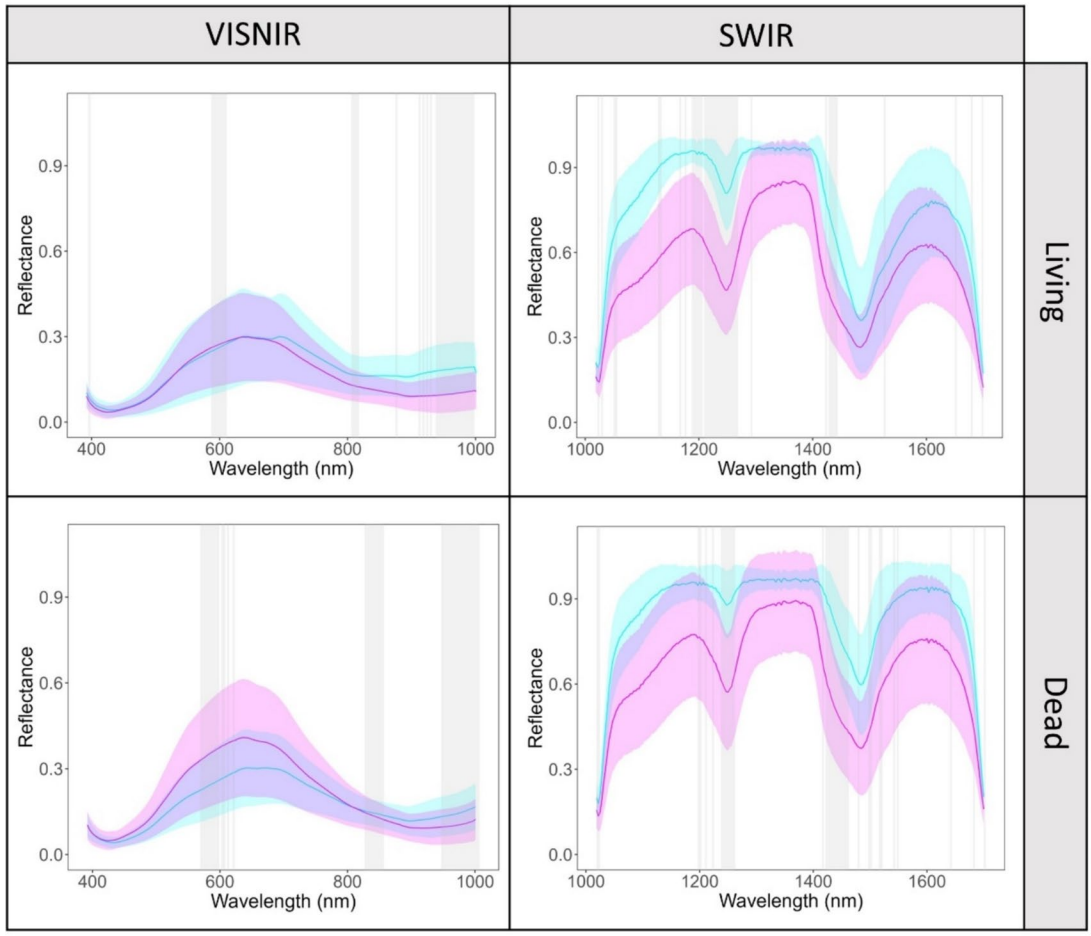

Fig. 3 Spectral plots for soil (magenta), as average of four soil types, and fine roots (cyan) as average of both species in March (Living) and April (Dead). Imaging campaign for visual-near infrared (VISNIR) and short-wave infrared (SWIR) regions. Means (solid lines) and standard errors (shaded area) of reflectance are given per wavelength (mean ± SE, $n > 500$). The grey vertical lines represent the important wavebands separating the two groups; see Online Resource 2, Supplementary Tables 1, 2 for details. Note: Difference in x-axis scale between VISNIR and SWIR spectra

Spectral differences between grass and herb fine roots

The PCAs on the species VISNIR spectral pattern (Fig. 4) illustrate that the spectra of *Alopecurus pratensis* and *Urtica dioica* species are tightly intertwined and have a significant overlap. Similarly, the K-means clustering of both living and dead root spectra is characterised by low MCC scores (0.27–0.36; Table 1). The supervised models DRF and GLM performed well (MCC 0.6–0.9), while PLS-DA scored < 0.5 MCC (Table 2). All three methods provided information on key VISNIR wavebands, certain sections in the middle of the spectrum become prominent, but these sections vary between the living (March; 710–770 nm) and dead (April; 560–580 and several regions from 830 to 1000 nm) roots (Fig. 5; Online Resource 2, Supplementary Table 4). A test of the models trained in March against the April data (Table 3) revealed differences that question the apparent spectral similarity of species at the two dates, but also provides potential signs of overfitting of the models on the March dataset (more pronounced in the DRF), echoing the analogous observations comparing root and soil spectra.

The SWIR data showed improved differentiation between the two species although a degree of overlap

Table 1 Accuracy and MMC of average K-means clustering results on validation data (VISNIR-SWIR); mean ± SE from fivefold cross-validation

Comparison sets	VISNIR		SWIR	
	Accuracy	MCC	Accuracy	MCC
Roots:Soil—Living	0.522 ± 0.007	0.100 ± 0.010	0.776 ± 0.001	0.595 ± 0.002
Roots:Soil—Dead	0.640 ± 0.010	0.180 ± 0.020	0.712 ± 0.002	0.507 ± 0.003
Species—Living	0.659 ± 0.004	0.274 ± 0.007	0.701 ± 0.004	0.467 ± 0.009
Species—Dead	0.670 ± 0.010	0.360 ± 0.020	0.734 ± 0.002	0.370 ± 0.009
Living:Dead—AP	0.477 ± 0.008	−0.040 ± 0.020	0.753 ± 0.001	0.493 ± 0.002
L:D—$AP_{1450–1700}$	-	-	0.782 ± 0.001	0.568 ± 0.002
Living:Dead—UD	0.530 ± 0.020	0.080 ± 0.030	0.688 ± 0.001	0.341 ± 0.003
L:D—$UD_{1450–1700}$	-	-	0.866 ± 0.004	0.745 ± 0.008

Roots:Soil comparison; Species: Species comparison; Living:Dead (L:D): March to April comparison for AP, *Alopecurus pratensis*, or UD, *Urtica dioica*; L:D $AP_{1450–1700}$ used only the wavelenght 1450 to1700 nm for clustering

Table 2 Accuracy and MMC of average GLM, PLS-DA and DRF models on validation data per VISNIR or SWIR data for different compared sets; mean ± SE from fivefold cross-validation

Algorithm	Comparison sets	VISNIR		SWIR	
		Accuracy	MCC	Accuracy	MCC
GLM	Roots:Soil—Living	0.747 ± 0.006	0.540 ± 0.009	0.984 ± 0.001	0.967 ± 0.001
	Roots:Soil—Dead	0.817 ± 0.005	0.640 ± 0.010	0.877 ± 0.001	0.754 ± 0.002
	Species—Living	0.834 ± 0.008	0.660 ± 0.010	0.999 ± 0.001	0.999 ± 0.001
	Species—Dead	0.813 ± 0.008	0.610 ± 0.010	0.992 ± 0.001	0.982 ± 0.001
	Living:Dead—AP	0.956 ± 0.005	0.910 ± 0.010	0.985 ± 0.001	0.970 ± 0.001
	Living:Dead—UD	0.898 ± 0.008	0.790 ± 0.020	0.983 ± 0.001	0.965 ± 0.002
PLS-DA	Roots:Soil—Living	0.714 ± 0.006	0.380 ± 0.010	0.913 ± 0.001	0.832 ± 0.002
	Roots:Soil—Dead	0.745 ± 0.008	0.410 ± 0.020	0.821 ± 0.001	0.642 ± 0.002
	Species—Living	0.724 ± 0.005	0.420 ± 0.010	0.956 ± 0.002	0.902 ± 0.004
	Species—Dead	0.750 ± 0.010	0.470 ± 0.030	0.876 ± 0.001	0.721 ± 0.002
	Living: Dead—AP	0.699 ± 0.005	0.400 ± 0.010	0.802 ± 0.002	0.595 ± 0.003
	Living: Dead—UD	0.652 ± 0.03	0.310 ± 0.060	0.912 ± 0.002	0.820 ± 0.006
DRF	Roots:Soil—Living	0.943 ± 0.003	0.876 ± 0.008	0.997 ± 0.001	0.993 ± 0.001
	Roots:Soil—Dead	0.906 ± 0.004	0.795 ± 0.007	0.970 ± 0.001	0.941 ± 0.001
	Species—Living	0.980 ± 0.005	0.960 ± 0.010	0.999 ± 0.001	0.998 ± 0.001
	Species—Dead	0.977 ± 0.002	0.952 ± 0.005	0.996 ± 0.001	0.992 ± 0.001
	Living:Dead—AP	0.991 ± 0.003	0.983 ± 0.006	0.998 ± 0.001	0.997 ± 0.001
	Living:Dead—UD	0.962 ± 0.005	0.920 ± 0.010	0.995 ± 0.001	0.991 ± 0.001

Species, species comparison: AP, *Alopecurus pratensis*; UD, *Urtica dioica*; Living:Dead, March to April comparison; MCC, Matthews Correlation Coefficient; GLM, General Linear Model; PLS-DA, Partial Least Squares Discriminant Analysis; DRF, Distributed Random Forest

remained (PCA; Fig. 4). The means SWIR spectra of *A. pratensis* and *U. dioica* roots showing considerable proximity in spectral pattern, with UD showing higher reflectance values in the range 1000 to 1100 nm, and AP showing higher values between c. 1400–1700 nm (Fig. 5). However, the standard errors of the two living species' root spectra overlap particularly strongly until 1400 nm, reflecting the variability in species' root spectra. The unsupervised k-means clustering holds accuracies of 0.70 and 0.73, and MCCs of 0.47 and 0.37 for the living (March) and dead (April) data, respectively (Table 1). The supervised classifications had high accuracies and MCCs, especially the DRF and GLM models (MCC 0.98–0.99), with PLS-DA holding slightly lower values for April (MCC 0.72), and revealing different SWIR wavebands contributing to the classification (Fig. 5; Online Resource 2, Supplementary Table 5). In contrast to VISNIR, the

Table 3 Accuracy and MCC of algorithms GLM, PLA-DA and DRF trained on living (March) and tested on dead root data (April) per VISNIR or SWIR data for different compared sets; mean ± SE from fivefold cross-validation

Algorithm	Comparison sets	VISNIR		SWIR	
		Accuracy	MCC	Accuracy	MCC
GLM	Roots:Soil	0.630 ± 0.003	0.432 ± 0.004	0.845 ± 0.001	0.690 ± 0.001
	Species	0.608 ± 0.001	0.171 ± 0.004	0.841 ± 0.001	0.653 ± 0.002
PLS-DA	Roots:Soil	0.680 ± 0.002	0.394 ± 0.002	0.801 ± 0.001	0.618 ± 0.001
	Species	0.650 ± 0.004	0.240 ± 0.010	0.698 ± 0.001	0.286 ± 0.001
DRF	Roots:Soil	0.760 ± 0.004	0.470 ± 0.008	0.801 ± 0.002	0.611 ± 0.004
	Species	0.621 ± 0.007	0.247 ± 0.008	0.782 ± 0.005	0.520 ± 0.010

Species contrast, *Alopecurus pratensis* and *Urtica dioica*; MCC, average Matthews Correlation Coefficient; GLM, General Linear Model; PLS-DA, Partial Least Squares Discriminant Analysis; DRF, Distributed Random Forest

methods point to the extremes of the spectrum (start and end wavebands) as crucial features, and additionally the spectra 1450 ~ 1500 nm for dead roots (April). When evaluating the models trained on living roots (March) and tested on dead roots (April), the SWIR data show a reduced level of species differentiation compared to the VISNIR spectral data (Table 3). Similarly to VISNIR, the testing on April's data is revealing lower MCC and accuracy scores (Table 3) compared to the within-month evaluation (Table 2), for all classification models tested.

Living-dead fine root differentiation by spectral patterns

The task of clustering roots according to vitality (Living:Dead) within a species is less challenging than distinguishing between species, as evidenced by improved accuracies and MCCs of the latter (Table 1). Neither the PCAs (Fig. 6) nor the visual spectral comparison (Fig. 7) show any discernible differences in the VISNIR region. The means start to diverge only from around ~ 820 nm, in both species (Fig. 7; Online Resource 5, Supplementary Fig. 7). The supervised models, however, point to some differentiating wavebands in the NIR region from ~ 900 to 940 nm (Fig. 7; Online Resource 2, Supplementary Table 6). At the same time the MCC values for Living:Dead comparisons are the highest for GLM modelling compared to the previous analyses (Roots:Soil, species differentiation), but no changes in MCC were observed for other models (Table 2).

In contrast, the SWIR region allowed for clearer discrimination of living and dead fine roots of either species by PCA (Fig. 6). Although the two clusters remain in proximity, they show a lower degree of overlap than observed with the species-related PCAs (Fig. 4). The means spectra of living and dead roots show, in particular for *U. dioica*, a marked difference in the region from 1410 to 1700 nm (Fig. 7; Online Resource 5, Supplementary Fig. 7). This agrees to some extent with the supervised methods in that the important wavebands for differentiation are concentrated from ~ 1460 to ~ 1570 nm (Fig. 7; Online Resource 2, Supplementary Table 7). Regarding the MCC score, the PLS-DA model shows opposite results, with UD having a higher score (0.82) than AP (0.59), compared to its performance in the VISNIR region where MCC of Living:Dead—AP is the highest, albeit with a modest 0.4 (Table 2). Moreover, the clustering results vary between species and are comparable to previous clustering analyses for species and roots-soil differentiation (Table 1; Online Resource 4, Supplementary Figs. 2, 3 and 4). However, the visual difference shown in Fig. 7, with the separation of the means (Online Resource 5, Supplementary Fig. 7) and the high concentration of important wavebands in the same location (Online Resource 2, Supplementary Table 7) allows improved accuracies/MCCs (Table 2) by focusing the clustering exceptionally on the 1450–1700 nm region of the SWIR spectra (AP/$UD_{1450–1700}$).

Discussion

In recent decades, the development of non-destructive spectral techniques has improved significantly,

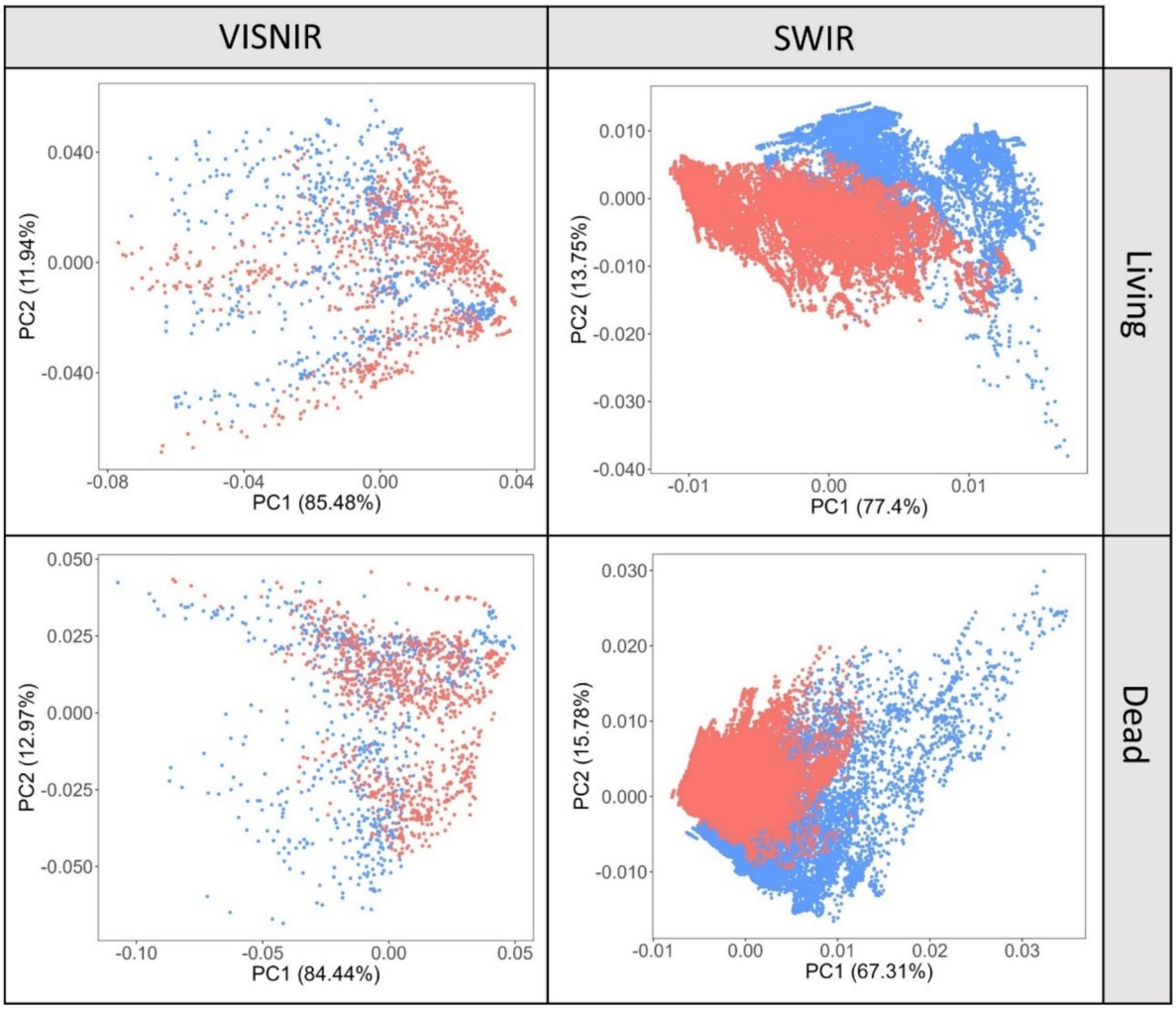

Fig. 4 PCA plots for the species *Alopecurus pratensis* (AP, red), and *Urtica dioica* (UD, blue) in the spectral ranges visual-near infrared (VISNIR) and short-wave infrared (SWIR), and for living (March) and dead roots (April). The VISNIR data is less dense due to a lower camera resolution

increasing the speed and reducing the cost of analysis. Near-infrared reflectance spectroscopy has emerged as an interesting alternative to conventional laboratory analysis in many fields, including soil science (Dangal et al. 2019; Serbin and Townsend 2020; Zahir et al. 2022) and root chemical analyses (Elle et al. 2019; White et al. 2011). While these analyses are almost exclusively performed on dried and homogenised samples, the aim of the present analysis was to determine the spectral differences between fresh soil and intact roots at transparent interfaces as common in rhizobox and minirhizotron studies (Neumann et al. 2009; Vamerali et al. 2012). While methods for automated detection of roots growing in soil and imaged in an RGB format are progressing rapidly (Baykalov et al. 2023; Khoroshevsky et al. 2024), it was hypothesized that extended spectral information of fine roots contains additional information that can be used for species and vitality discrimination. Both, species identification and vitality discrimination of roots in situ are key bottlenecks in root ecological studies aiming at e.g. unravelling species interactions below ground (Violle et al. 2007) and/or determining realistic soil C input by root turnover (Gill and Jackson 2000). While there have been attempts to define plant 'optical types' for vegetation above ground, relating spectral

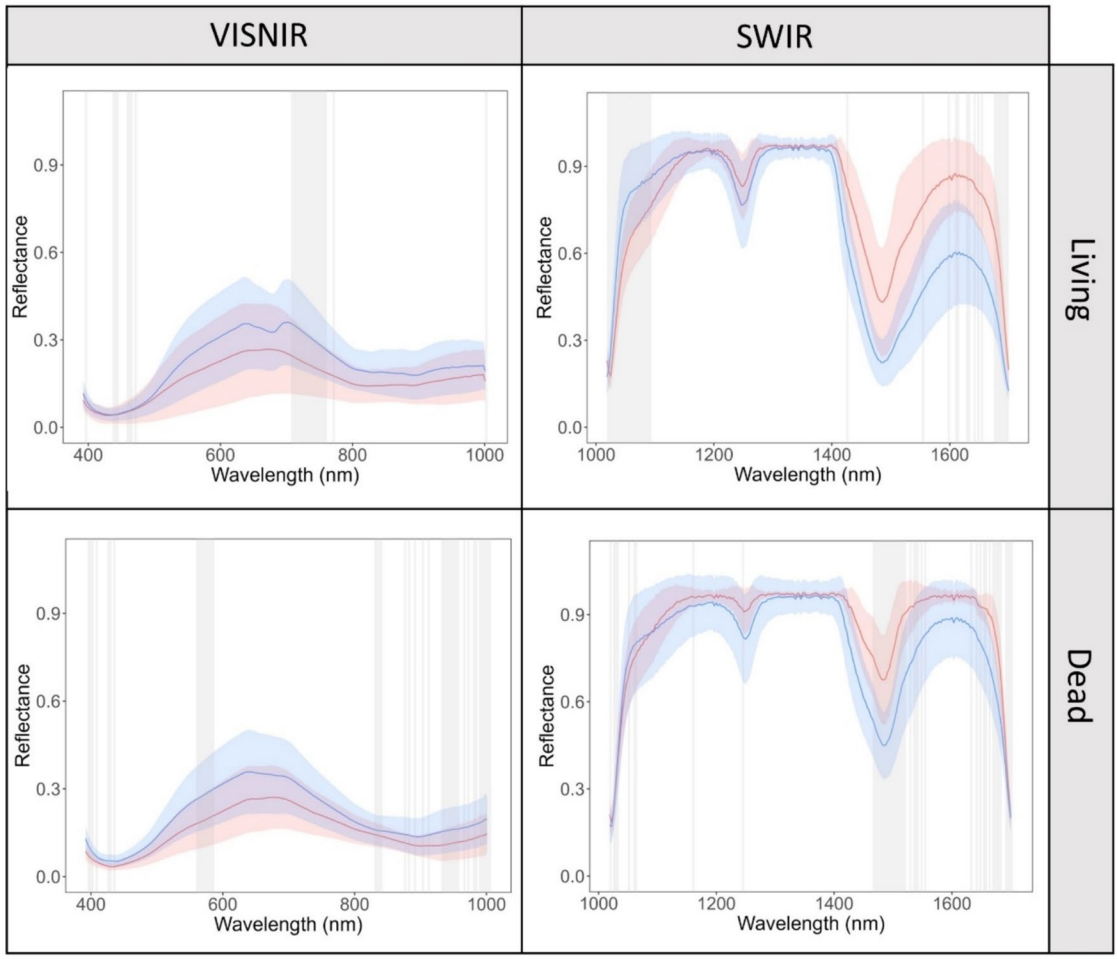

Fig. 5 Spectral reflectance plots for *Alopecurus pratensis* (AP, red) and *Urtica dioica* (UD, blue) in March (Living) and April (Dead) and for visual-near infrared (VISNIR) and shortwave infrared (SWIR) ranges. Means (solid lines) and standard errors (shaded area) of reflectance are given per wavelength (mean ± SE, $n > 500$). The grey vertical shadows represent the important wavebands separating the two groups; see Online Resource 2, Supplementary Tables 3, 4 for details. Note: Difference in x-axis scale between VISNIR and SWIR spectra

properties of leaves/canopies to functional plant traits (Ustin and Gamon 2010), information on root spectral patterns in situ remains scarce.

Water is the easiest component to identify in NIR spectra, with dedicated bands around 970 nm, 1245 nm, and 1450 nm (Hank et al. 2019; Thenkabail et al. 2013; Zahir et al. 2022). Roots often have high gravimetric water contents that are independent of soil water contents (Guo et al. 2013). Indeed, the water content valleys in the SWIR spectra emerge as a discernible cue that augments the probability of a pixel being attributed to a root rather than soil. This is shown in Fig. 3, where living fine roots had a higher reflectance than the averaged soil spectra—particularly at the upper end of the VISNIR / lower end of the SWIR spectra (around 1000–1250 nm), and the wavebands > 1400 nm, for dead roots (Fig. 3; Online Resource 5, Supplementary Fig. 5). While this alignment is generally consistent with previous research, highlighting the importance of wavebands related to water content in effectively discriminating between roots and soil (Bodner et al. 2018), our results show that the soil-root separation is less distinct at the 1450 nm water band (in living roots), with a large overlap in standard errors, compared to the ~ 1245 nm water band (Online Resource 5, Supplementary Fig. 5).

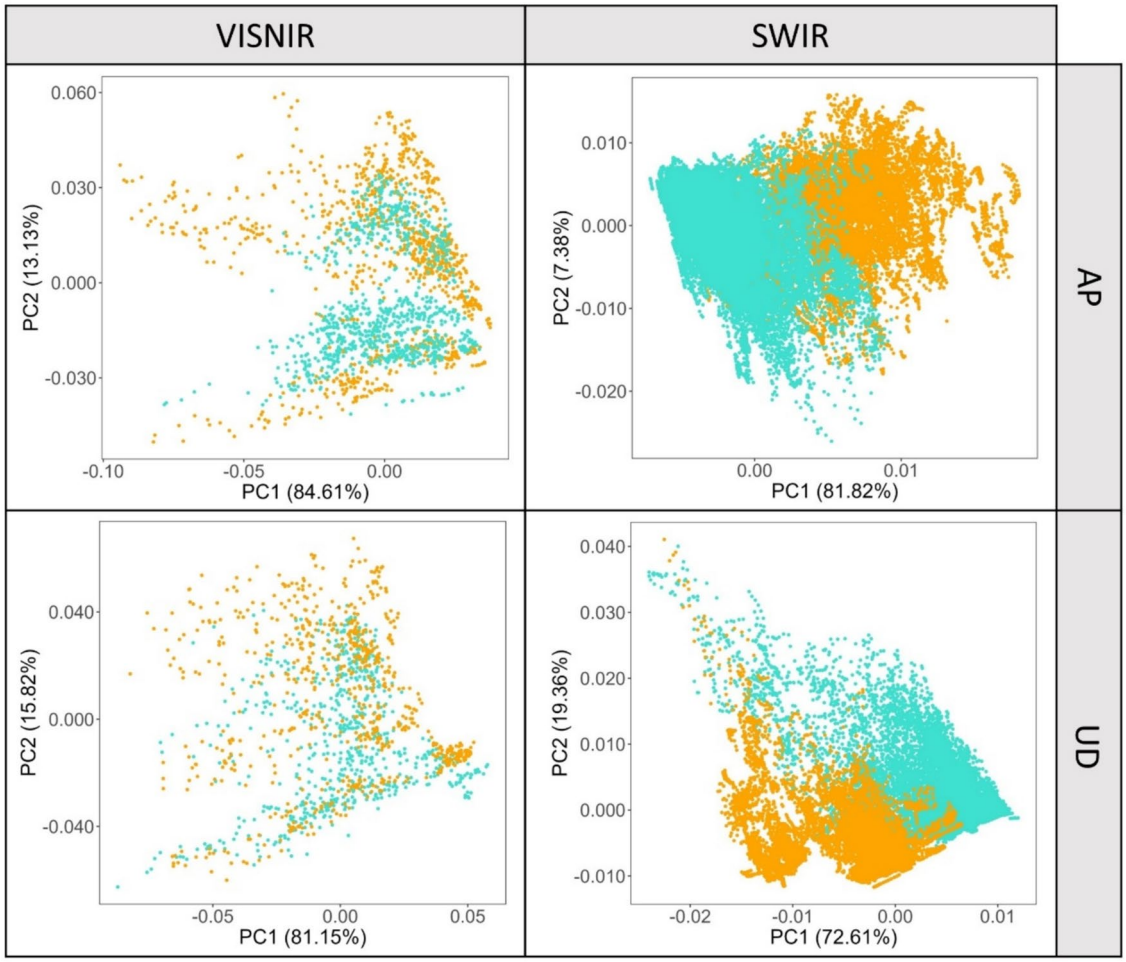

Fig. 6 PCA plots for spectra of living (March, orange) and dead (April, turquoise) roots of the species *Alopecurus pratensis* (AP), and *Urtica dioica* (UD) in the spectral ranges visual-near infrared (VISNIR) and short-wave infra-red (SWIR). The VISNIR data is less dense due to a lower camera resolution

Both the March and April datasets used combined soil spectra comprising four different soil types including sand, which probably contributes to expanded standard errors and increased mean reflectance values (Norouzi et al. 2021). This effect is attributed to the similarity in colour between sand and the white-yellowish hue of roots, as illustrated in Fig. 1b, and the close correspondence between root and soil spectra in VISNIR (Fig. 3). However, the SWIR range appears to be less affected by these colour-related issues, but is rather depended on particle size and mineral composition (Norouzi et al. 2021), with the mean soil reflectance consistently below the root spectrum in both March and April datasets (Fig. 3); and negligible differences observed between soils from both months (Online Resource 6, Supplementary Fig. 8). It is hypothesized that water contents significantly influence the proximity of the soil mean reflectance to the root mean reflectance within the 1450–1600 nm range in March (Fig. 3; Online Resource 5, Supplementary Fig. 5). Although we continued irrigation, it is speculated that April soils were slightly drier than in March, which is reflected in a slight increase in the average soil spectrum in April (Online Resource 6, Supplementary Fig. 8), as reflectance often increases with decreasing soil moisture (Yang et al. 2019). Nevertheless, the classification algorithms showed good results when discriminating average soil spectra from

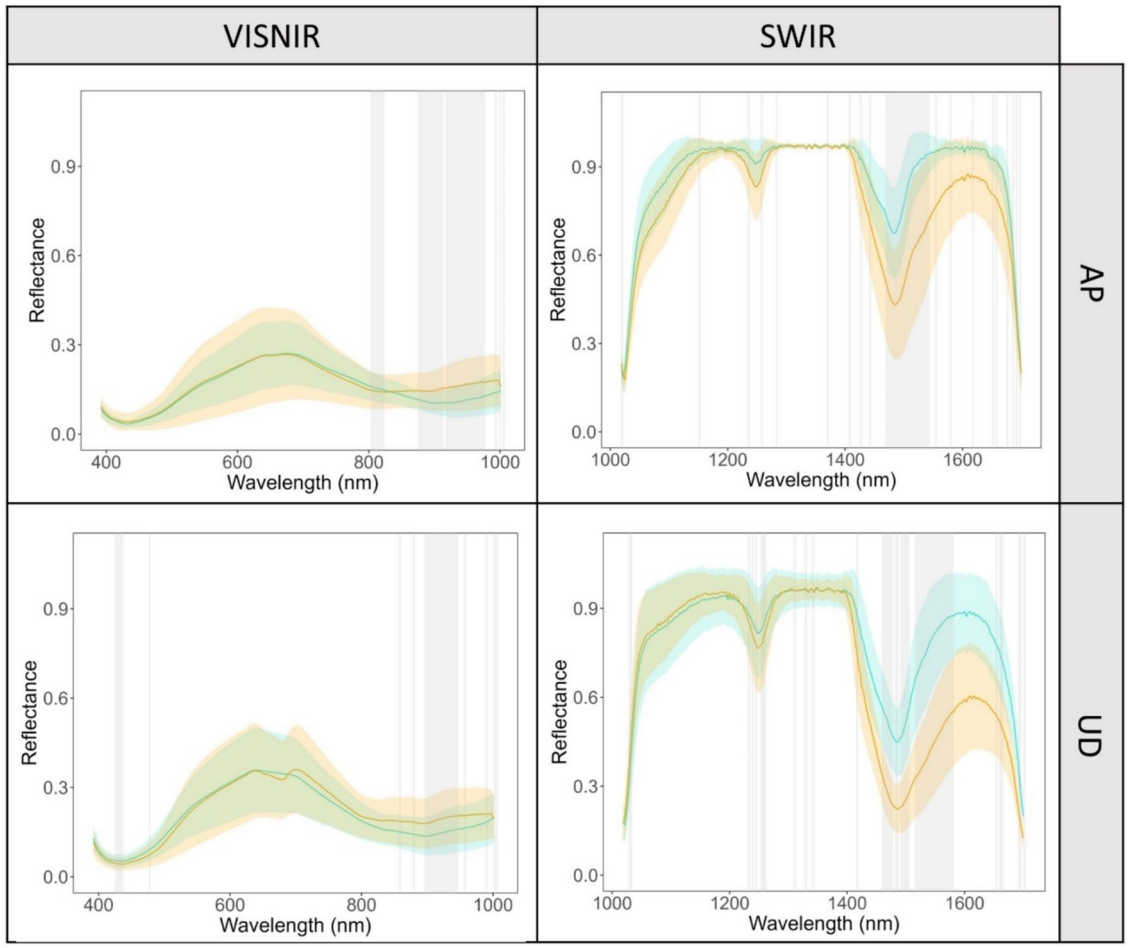

Fig. 7 Spectral plots for living (orange) and dead (turquoise) roots of the species *Alopecurus pratensis* (AP), and *Urtica dioica* (UD) in the spectral ranges visual-near infrared (VISNIR) and short-wave infrared (SWIR). Means (solid lines) and standard errors (shaded area) of reflectance are given per wavelength (mean ± SE, $n > 500$). The grey vertical shadows represent the important wavebands for separating the two groups; see Online Resource 2, Supplementary Tables 5, 6 for details. Note: Difference in x-axis scale between VISNIR and SWIR spectra

average living root spectra in SWIR (Table 2). Clustering, on the other hand, maintained an MCC score of 0.5 for both the March and April data (Table 1; Online Resource 4, Supplementary Fig. 2). This score is, however, often considered sufficient as MCC values range from −1 to 1 (Chicco et al. 2021).

When analysing NIR spectra, it's crucial to recognise that the wavelengths do not correspond to specific molecules or chemical species, but represent molecular bonds such as C-H, O–H, and N–H (Wu et al. 2014). These bonds absorb at multiple locations in the NIR spectrum, forming overtones and band combinations due to their proximity (Workman Jr 1996). While chemometric data to build a predictive model for in situ spectra of fine roots are still missing, fine roots of grassland species, for example, contain approx. 45% C, and 1.4% N (Ning et al. 2017). However, significant differences in chemical composition among the roots of grass and herb species have been reported (Bhattarai et al. 2024; Dawson et al. 2000; Furey and Tilman 2021). In addition, the moisture content in fresh roots of various herb species has been documented to range from 39% for *Poa sphondylodes* to 100% for *Heteropappus altaicus* (Zhang et al. 2019). Given these expected differences in root / rhizodermis stoichiometry and water

content (in herbs and grasses), we hypothesised that spectral patterns could be used to distinguish fresh roots of different species. Indeed, the reflectance spectra of living roots of both species showed less overlap in the PCA (Fig. 4) compared to the soil-root (Living/March) comparison (Fig. 2). Regarding the spectral plots in the SWIR region, however, more separated spectra (means and its corresponding standard errors) appeared in the 1400–1700 nm range than when comparing average soil and root spectra, while the opposite was true for the 1000–1250 nm range (Figs. 3, 5; Online Resource 3, Supplementary Figs. 4, 5). Excellent MCCs of ~ 0.9 were achieved with all classification models, which is even slightly better than the MCC scores achieved for roots-soil discrimination in the SWIR range. Nevertheless, testing the models trained on March data on April data showed that the supervised models performed worse compared to their evaluation on the same month they were trained on (similar to the roots-soil differentiation). These results could reveal the underlying differences between March and April data as well as the slight tendency of the models to overfit the March data, particularly evident in VISNIR. For instance, the DRF model significantly reduced its MCC score from 0.9 to 0.2 when tested in April within VISNIR (Tables 2, 3). Nevertheless, the primary distinction arises from the divergence between the living and dead conditions of the roots (March–April). This suggests that constructing a robust and accurate model for species or roots-soil differentiation requires the inclusion of mixed data from both living and dead roots.

While the water-related bands were key for the Roots:Soil discrimination (as discussed above), the wavebands at the extremes of the SWIR spectrum, specifically at the beginning (< 1100 nm) and end (> 1600 nm), were key features for species discrimination by the models. We speculate that this importance is based on the fact that *U. diocia* has a higher reflectance than *A. pratensis* roots in the 1000 ~ 1100 nm region and *A. pratensis* has a higher reflectance ≥ 1420 nm, while both regions simultaneously hold less variation in reflectance than the part of the spectra with the largest difference in means (i.e. ~ 1450–1600 nm). Previous studies suggested that key wavebands for species identification are situated at higher NIR wavelength > 1450 nm (Tong et al. 2016) – or even further in the infrared range, around 15,000 nm in the MIR range (Naumann et al. 2010; Straková et al. 2020).

The biochemical reason for the spectral patterns of intact roots remains unclear. Although only turgescent and healthy-looking roots were selected in March, we cannot completely exclude that some roots have senesced in the five months since germination. Furthermore, the spectral reflectance is likely to be constantly changing as roots age—develop root hairs, differentiate boundary tissue, etc. In the VIS spectrum, this may also be related to (gradual) changes in rhizodermis colour due to senescence (Niinemets and Ostonen 2020) or the physiological differentiation of living roots (Rewald et al. 2014). Although there may be several unexplored stages of spectral changes during root development, our dataset was averaged by randomly selecting a large number of pixels at various locations per root type. The cutting of shoots and the dark incubation over a period of 42 days resulted in a noticeable change in the spectral reflectance patterns (Fig. 7). Not surprisingly, the ongoing decomposition of the roots led to SWIR spectra that were more similar to the soil (Fig. 3) – resulting in relatively lower MCCs of models compared to living conditions (Tables 1, 2, Roots:Soil). Interestingly, the region of 1450–1600 nm, which was more intertwined in March, allowed for a higher means of separation of dead roots from soil in April (Online Resource 5, Supplementary Fig. 5). This could suggest a potential shift in the chemical composition or structural properties of the roots post-decomposition. Senescent herb/grass roots have been reported to reduce root N concentration (although less than in leaves), while soluble carbohydrate and starch concentrations increased in senescing absorptive roots (in contrast to leaves) (Wojciechowska et al. 2020) – leading to marked differences in chemical composition. Larsson and Steen (1988), for example, reported that roots of grasses decompose fast, losing 35–65% of their dry mass within 94 days. However, chemical changes during decomposition vary among grass species, with roots of e.g. *Stipa* spp. and *Festuca scabrella* exhibiting differing percent loss of carbohydrates, indicating distinct decomposition dynamics influenced by initial root chemical composition (Herman et al. 1977). Similar to our results, a previous study on tree fine roots reported that dead roots reflected less light on average in the VISNIR region 800 ~ 1000 nm than living roots (Nakaji et al. 2008). However, the data shows that the most significant differences between living and dead

roots occur in the SWIR spectral region > 1400 nm—with a higher reflectance of dead roots, especially in *U. dioica*. The supervised methods emphasize the water content regions ~ 970 nm and ~ 1450 nm as primary features for vitality classification (Online Resource 2, Supplementary Tables 5, 6), with dead roots having a lower water content than living ones. Upon closer analysis of the SWIR spectra, the latter portion of the range, beginning at approximately 1400 nm, exhibits additional discernible differences between dead and living roots (Fig. 7; Online Resource 5, Supplementary Fig. 7). The average spectra (and their standard errors) of each group are more widely separated. Besides water content, the spectral region from 1450 to 1700 nm is known to be related to lignin, and cellulose contents (Hank et al. 2019; Thenkabail et al. 2013).

In addition to chemical parameters, the physical state of a sample can also affect the spectra. NIR spectra often show baseline shifts, usually indicating variations due to particle size, bulk density or porosity (Askari et al. 2015; Norouzi et al. 2021). This appears to be particularly relevant at soil/root interfaces towards the transparent surfaces of root observation systems (rhizoboxes or minirhizotron tubes), where different particle sizes (sand grains or different soil aggregate sizes) and cylindrical roots create a mosaic of shadows and different reflection angles. In addition, water droplets can easily form on the transparent panels and accumulate in small cavities at the soil/panel interface (Baykalov et al. 2023). This poses a particular challenge in discriminating soil adjacent to root edges (Moradi et al. 2011), where higher water contents may prevent a clear transition (Fig. 3, 1450–1550 nm region in living roots). In addition, and although the labelling was concentrated along the central axis of the root, the lower resolution of the SPECIM IQ VISNIR camera may have (partially) blended soil and root pixels at their edges, occasionally preventing precise delineation. Besides, the low resolution that VISNIR dataset presented, did not enable us to properly detect the finest roots. This may partly explain the high overlap of VISNIR spectra in all clusters examined. However, it is likely that in situ root spectra are always influenced to some extent by the chemical composition of the soil, present as a 'smear' on the root surface, or soil water, 'washing over' the root surfaces during rain/irrigation events and depositing fine particles. Similarly, spectrometric measurements on leaves are known to be influenced by water droplets and dust particles on the leaf surface (Lin et al. 2021) or embedded in the wax layer on the cuticula (Liu et al. 2020). In addition, the microbial population on the rhizoplane, with an increasing density of bacterial cells from the root tip to the region of root cell maturation, can also significantly influence spectral patterns (Schmidt et al. 2018).

Through this study, it became clear that the VISNIR spectrum provides significantly fewer spectral differences in comparison to the SWIR range, resulting in higher MCCs for the latter in every analysis performed. This may be due to the higher resolution of the SWIR camera as well as the spectral information itself. Importantly, the visible differences observed between roots and soil in standard RGB images did not directly correspond to the spectral properties in VISNIR data. We suggest that these disparities might be more complex, potentially arising from subtle texture differences that affect closely related colours and spectral characteristics. Consequently, the connection between neighbouring pixels, i.e. spatial information, could add to interpret the spectral values of individual pixels. We, thus suggest that combining both spectral and spatial information may facilitate future spectral analyses of roots in situ (Fauvel et al. 2013; Qian et al. 2013; Ran et al. 2017; Xu and Gowen 2020). Nevertheless, the GLM model, on SWIR data, emerged as a better choice for spectrum-based classifications compared to the DRF and PLS-DA models. Its consistent accuracy, even when applied to April data but trained on March spectra, suggests a resilience to the overfitting that DRF has shown (Tables 2, 3), although DRF had the highest accuracies overall. Despite its popularity in many FTIR studies (Ali et al. 2019; Lin et al. 2012; Yu et al. 2018), PLS-DA did not perform well during classifications, not only scoring lower than the other methods but also being less robust when tested on April data. Further studies targeting in situ spectra of soil and roots should thus test different models to determine their suitability and cannot rely on the use of established models applied to spectra of dried and homogenized root and soil samples to yield the best in situ classifications.

Conclusions

This study illustrates the opportunities and challenges of using VISNIR and SWIR spectral imaging to differentiate non-woody roots of visually similar

species and to distinguish living from dead roots in rhizoboxes, rhizotrons or minirhizotron installations. The observed spectral patterns, particularly in SWIR, provide valuable insights into root species and vitality differentiation, offering a non-invasive means of assessing belowground processes. Species differentiation, while posing a challenge, shows promising a potential as significant deviations in reflectance were observed primarily in spectral regions 1000 ~ 1100 nm and > 1400 nm. The spectral region beyond 1400 nm has significant relevance for differentiating between living and dead roots at early stages of decomposition, indicating potential for improving non-destructive root turnover rate calculations. Additionally, analysis of averaged soil and root spectra revealed their spectral proximity, although roots presented a higher reflectance than the soil with notable variations in the water content regions specifically at 970 nm, 1245 nm, and 1450 nm. We suggest that future spectrometry studies should also consider spatial information to develop robust methods for the classification of rhizosphere elements and roots in particular. It was not considered during the time of the experiment, to extract and wash the roots and then get the spectra in these conditions. These additional spectra would have offered important information for the validation of our data. Moreover, our findings underscore the importance of selecting appropriate classification models tailored to in situ spectral data, rather than relying solely on established models applied to dried and homogenised samples. In conclusion, this study demonstrates the potential of spectral analysis for in situ root species and vitality identification.

Acknowledgements We thank Cosmo Lang, who helped to carry out the root labelling as part of his BSc studies. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 813114 (FutureArctic-ITN). BR received funding from the European Union's Horizon research and innovation programme under grant agreement No 101087262 (EXCELLENTIA) during the manuscript preparation stage. IO was supported by the Estonian Research Council (PRG916). We thank two anonymous reviewers for their suggestions on an earlier version of the manuscript, which greatly improved its accessibility to readers.

Author contributions BR and GB conceived the study conception and design. Material preparation, data collection and analysis were performed by PB. The first draft of the manuscript was written by PB and BR and all authors commented on previous versions of the manuscript. PB and BR revised the manuscript. All authors read and approved the final manuscript.

Funding Open access publishing supported by the institutions participating in the CzechELib Transformative Agreement. This work received funding from the European Union's Horizon 2020 research and innovation programme (Grant numbers 813114 (FutureArctic-ITN) and 101087262 (EXCELLENTIA)). IO was supported by the Estonian Research Council (PRG916).

Data availability The datasets generated during and/or analysed during the current study are available from the first or corresponding author upon reasonable request.

Declarations

Competing interests Financial interests: Authors GB and IO declare they have no financial interests. IO is an Editorial Board Member of Plant and Soil. Authors PB and BR were partially employed by the Vienna Scientific Instruments GmbH, a manufacturer of minirhizotron hardware and rhizoboxes.

References

Ali MM, Bachik NA, Muhadi NA, Yusof TNT, Gomes C (2019) Non-destructive techniques of detecting plant diseases: A review. Physiol Mol Plant P 108. https://doi.org/10.1016/j.pmpp.2019.101426

Askari MS, Cui J, O'Rourke SM, Holden NM (2015) Evaluation of soil structural quality using VIS–NIR spectra. Soil Tillage Research 146:108–117. https://doi.org/10.1016/j.still.2014.03.006

Baykalov P (2024) Labelling-tools-for-IR-data, GitHub. https://github.com/dIcarusb/Labelling-tools-for-IR-data Accessed 12 Feb 2024.

Baykalov P, Bussmann B, Nair R, Smith AG, Bodner G, Hadar O, Lazarovitch N, Rewald B (2023) Semantic segmentation of plant roots from RGB (mini-) rhizotron images—generalisation potential and false positives of established methods and advanced deep-learning models. Plant Methods 19:122. https://doi.org/10.1186/s13007-023-01101-2

Bedini E (2017) The use of hyperspectral remote sensing for mineral exploration: a review. Journal of Hyperspectral

Remote Sensing 7:189–211. https://doi.org/10.29150/jhrs.v7.4.p189-211
Bhattarai B, Sigurdsson BD, Sigurdsson P, Leblans N, Janssens I, Meynzer W, Devarajan AK, Truu J, Truu M, Ostonen I (2023) Soil warming duration and magnitude affect the dynamics of fine roots and rhizomes and associated C and N pools in subarctic grasslands. Ann Bot 132:269–279. https://doi.org/10.1093/aob/mcad102
Bhattarai B, Richter A, Metze D, Sigurdsson B, Sigurðsson P, Leblans N, Janssens I, Ostonen I (2024) Influence of soil warming magnitude and duration on soluble sugar pool in fine roots and rhizomes of subarctic grasslands: differences at species and plant community level adaptation. Plant Stress 11. https://doi.org/10.1016/j.stress.2024.100406
Blondeel H, Guillemot J, Martin-StPaul N, Druel A, Bilodeau-Gauthier S, Bauhus J, Grossiord C, Hector A, Jactel H, Jensen J, Messier C, Muys B, Serrano-León H, Auge H, Barsoum N, Birhane E, Bruelheide H, Cavender-Bares J, Chu C, Cumming JR, Damtew A, Eisenhauer N, Ferlian O, Fiedler S, Ganade G, Godbold DL, Gravel D, Hall JS, Hölscher D, Hulvey KB, Koricheva J, Kreft H, Lapadat C, Liang J, Liu X, Meredieu C, Mereu S, Montgomery R, Morillas L, Nock C, Paquette A, Parker JD, Parker WC, Paterno GB, Perring MP, Ponette Q, Potvin C, Reich PB, Rentch J, Rewald B, Sandén H, Sinacore K, Standish RJ, Stefanski A, Tobin PC, van Breugel M, Fagundes MV, Weih M, Williams LJ, Zhou M, Scherer-Lorenzen M, Verheyen K, Baeten L (2024) Tree diversity reduces variability in sapling survival under drought. J Ecol 112:1164–1180. https://doi.org/10.1111/1365-2745.14294
Bodner G, Nakhforoosh A, Arnold T, Leitner D (2018) Hyperspectral imaging: a novel approach for plant root phenotyping. Plant Methods 14. https://doi.org/10.1186/s13007-018-0352-1
Boughorbel S, Jarray F, El-Anbari M (2017) Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. PLoS ONE 12. https://doi.org/10.1371/journal.pone.0177678
Bradski G (2000) The OpenCV Library. Dr Dobb's Journal of Software Tools 25(120):122–125
Breiman L (2001) Random Forests. Mach Learn 45:5–32. https://doi.org/10.1023/A:1010933404324
Buitrago MF, Skidmore AK, Groen TA, Hecker CA (2018) Connecting infrared spectra with plant traits to identify species. ISPRS J Photogramm Remote Sens 139:183–200. https://doi.org/10.1016/j.isprsjprs.2018.03.013
Calera A, Martínez C, Melia J (2001) A procedure for obtaining green plant cover: Relation to NDVI in a case study for barley. Int J Remote Sens 22:3357–3362. https://doi.org/10.1080/01431160010020100
Cappelli SL, Domeignoz-Horta LA, Loaiza V, Laine A-L (2022) Plant biodiversity promotes sustainable agriculture directly and via belowground effects. Trends Plant Sci 27:674–687. https://doi.org/10.1016/j.tplants.2022.02.003
Chicco D, Tötsch N, Jurman G (2021) The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Mining 14:13. https://doi.org/10.1186/s13040-021-00244-z
Clayden J (2020) Mathematical morphology in any number of dimensions, GitHub. https://github.com/jonclayden/mmand. Accessed 18 Oct 2023
da Luz BR, Crowley JK (2007) Spectral reflectance and emissivity features of broad leaf plants: prospects for remote sensing in the thermal infrared (8.0-14.0 mu m). Remote Sens Environ 109:393–405. https://doi.org/10.1016/j.rse.2007.01.008
Dangal SRS, 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. https://doi.org/10.3390/soilsystems3010011
Dawson LA, Mayes RW, Elston DA, Smart TS (2000) Root hydrocarbons as potential markers for determining species composition. Plant, Cell Environ 23:743–750. https://doi.org/10.1046/j.1365-3040.2000.00592.x
Ehrenfeld JG, Parsons WFJ, Han X, Parmelee RW, Zhu W (1997) Live and dead roots in forest soil horizons: contrasting effects on nitrogen dynamics. Ecology 78:348–362. https://doi.org/10.1890/0012-9658(1997)078[0348:LADRIF]2.0.CO;2
Elle O, Richter R, Vohland M, Weigelt A (2019) Fine root lignin content is well predictable with near-infrared spectroscopy. Sci Rep 9:6396. https://doi.org/10.1038/s41598-019-42837-z
Fauvel M, Tarabalka Y, Benediktsson JA, Chanussot J, Tilton JC (2013) Advances in Spectral-Spatial Classification of Hyperspectral Images. Proc IEEE 101:652–675. https://doi.org/10.1109/JPROC.2012.2197589
Fei B (2019) Chapter 3.6 - Hyperspectral imaging in medical applications. In: Amigo JM (ed) Data handling in science and technology, Elsevier, Vol 32. pp 523–565. https://doi.org/10.1016/B978-0-444-63977-6.00021-3
Fitzky AC, Kaser L, Peron A, Karl T, Graus M, Tholen D, Halbwirth H, Trimmel H, Pesendorfer M, Rewald B, Sandén H (2023) Same, same, but different: Drought and salinity affect BVOC emission rate and alter blend composition of urban trees. Urban Forestry & Urban Greening 80. https://doi.org/10.1016/j.ufug.2023.127842
Forgy EW (1965) Cluster analysis of multivariate data: efficiency versus interpretability of classifications. Biometrics 21:768–769
Freschet GT, Pagès L, Iversen CM, Comas LH, Rewald B, Roumet C, Klimešová J, Zadworny M, Poorter H, Postma JA et al (2021) A starting guide to root ecology: strengthening ecological concepts and standardising root classification, sampling, processing and trait measurements. New Phytol 232:973–1122. https://doi.org/10.1111/nph.17572
Furey GN, Tilman D (2021) Plant biodiversity and the regeneration of soil fertility. Proc Natl Acad Sci 118. https://doi.org/10.1073/pnas.2111321118
Gill RA, Jackson RB (2000) Global patterns of root turnover for terrestrial ecosystems. New Phytol 147:13–31. https://doi.org/10.1046/j.1469-8137.2000.00681.x
Gorman B (2022) mltools, GitHub. https://github.com/ben519/mltools. Accessed 8 Sep 2023.
Grandini M, Bagli E, Visani G (2020) Metrics for multi-class classification: an Overview. arXiv [statML]. https://doi.org/10.48550/arXiv.2008.05756

Guo L, Lin H, Fan B, Cui X, Chen J (2013) Impact of root water content on root biomass estimation using ground penetrating radar: evidence from forward simulations and field controlled experiments. Plant Soil 371:503–520. https://doi.org/10.1007/s11104-013-1710-4

H2O.ai (2023b) Variable Importance in H2O.ai. H2O.ai. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/variable-importance.html. Accessed 26 Jan 2024

Hank T, Berger K, Bach H, Clevers JGPW, Gitelson A, Zarco-Tejada P, Mauser W (2019) Spaceborne Imaging Spectroscopy for Sustainable Agriculture: Contributions and Challenges. Surv Geophys 40:515–551. https://doi.org/10.1007/s10712-018-9492-0

Herman WA, McGill WB, Dormaar JF (1977) Effects of initial chemical composition on decomposition of roots of three grass species. Can J Soil Sci 57:205–215. https://doi.org/10.4141/cjss77-025

Jaggi N, Vij DR (2006) Fourier transform infrared spectroscopy. In: Vij DR (ed) Handbook of Applied Solid State Spectroscopy. Springer, US, Boston, MA, pp 411–450

Johnson J-M, Sila A, Senthilkumar K, Shepherd KD, Saito K (2021) Application of infrared spectroscopy for estimation of concentrations of macro- and micronutrients in rice in sub-Saharan Africa. Field Crop Res 270. https://doi.org/10.1016/j.fcr.2021.108222

Khoroshevsky F, Zhou K, Chemweno S, Edan Y, Bar-Hillel A, Hadar O, Rewald B, Baykalov P, Ephrath JE, Lazarovitch N (2024) Automatic Root Length Estimation from Images Acquired In Situ without Segmentation. Plant Phenomics 6:0132. https://doi.org/10.34133/plantphenomics.0132

Knipling EB (1970) Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. Remote Sens Environ 1:155–159. https://doi.org/10.1016/S0034-4257(70)80021-9

Kreyling J, Thiel D, Simmnacher K, Willner E, Jentsch A, Beierkuhnlein C (2012) Geographic origin and past climatic experience influence the response to late spring frost in four common grass species in central Europe. Ecography 35:268–275. https://doi.org/10.1111/j.1600-0587.2011.07173.x

Kuhn M (2008) Building predictive models in r using the caret package. J Stat Soft 28:1–26. https://doi.org/10.18637/jss.v028.i05

Laasonen M, Harmia-Pulkkinen T, Simard CL, Michiels E, Räsänen M, Vuorela H (2002) Fast Identification of Echinacea purpurea Dried Roots Using Near-Infrared Spectroscopy. Anal Chem 74:2493–2499. https://doi.org/10.1021/ac011108f

Larsson K, Steen E (1988) Changes in mass and chemical composition of grass roots during decomposition. Grass Forage Sci 43:173–177. https://doi.org/10.1111/j.1365-2494.1988.tb01885.x

LeDell E, Poirier S (2020) H2O AutoML: Scalable automatic machine learning. 7th ICML Workshop on Automated Machine Learning (AutoML). https://www.automl.org/wp-content/uploads/2020/07/AutoML_2020_paper_61.pdf

Lee L, Liong CY (2018) Partial Least Squares-Discriminant Analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. The Analyst 143. https://doi.org/10.1039/C8AN00599K

Li G, Ma S, Li K, Zhou M, Lin L (2022) Heterogeneity classification based on hyperspectral transmission imaging and multivariate data analysis. Infrared Phys Technol 123. https://doi.org/10.1016/j.infrared.2022.104180

Li Y, Ta MH aW (2016) c. The R Journal 8:478–489. https://doi.org/10.32614/RJ-2016-060

Lin W-S, Yang C-M, Kuo B-J (2012) Classifying cultivars of rice (Oryza sativa L.) based on corrected canopy reflectance spectra data using the orthogonal projections to latent structures (O-PLS) method. Chemom Intell Lab Syst 115:25–36. https://doi.org/10.1016/j.chemolab.2012.04.005

Lin W, Sun Y, Wang D, Li Y, Yu X (2021) Estimation model of dust deposition capacity of common vegetation based on spectral characteristics in Shanghai, China. Sustainable Cities and Society 70:102915. https://doi.org/10.1016/j.scs.2021.102915

Lindenmayer DB, Margules CR, Botkin DB (2000) Indicators of Biodiversity for Ecologically Sustainable Forest Management. Conserv Biol 14:941–950. https://doi.org/10.1046/j.1523-1739.2000.98533.x

Liu N, Zhao L, Tang L, Stobbs J, Parkin I, Kunst L, Karunakaran C (2020) Mid-infrared spectroscopy is a fast screening method for selecting Arabidopsis genotypes with altered leaf cuticular wax. Plant, Cell Environ 43:662–674. https://doi.org/10.1111/pce.13691

Martinez-Vega B, Tkachenko M, Matkabi M, Ortega S, Fabelo H, Balea-Fernandez F, La Salvia M, Torti E, Leporati F, Callico GM, Chalopin C (2022) Evaluation of Preprocessing Methods on Independent Medical Hyperspectral Databases to Improve Analysis. Sensors 22. https://doi.org/10.3390/s22228917

Meinen C, Rauber R (2012) FTIR-ATR spectroscopy–a new approach in root discrimination of crop and weed species. Julius-Kühn-Archiv 702. https://doi.org/10.5073/jka.2012.434.091

Min D, Zhao J, Bodner G, Ali M, Li F, Zhang X, Rewald B (2023) Early decay detection in fruit by hyperspectral imaging–Principles and application potential. Food Control 152. https://doi.org/10.1016/j.foodcont.2023.109830

Mishra P, Asaari MSM, Herrero-Langreo A, Lohumi S, Diezma B, Scheunders P (2017) Close range hyperspectral imaging of plants: A review. Biosys Eng 164:49–67. https://doi.org/10.1016/j.biosystemseng.2017.09.009

H2O.ai (2023a) Generalized Linear Modeling in H2O.ai. H2O.ai. https://docs.h2o.ai/h2o/latest-stable/h2o-docs/data-science/glm.html. Accessed 26 Jan 2024

Moradi AB, Carminati A, Vetterlein D, Vontobel P, Lehmann E, Weller U, Hopmans JW, Vogel H-J, Oswald SE (2011) Three-dimensional visualization and quantification of water content in the rhizosphere. New Phytol 192:653–663. https://doi.org/10.1111/j.1469-8137.2011.03826.x

Nakaji T, Noguchi K, Oguma H (2008) Classification of rhizosphere components using visible–near infrared spectral images. Plant Soil 310:245–261. https://doi.org/10.1007/s11104-007-9478-z

Naumann A, Heine G, Rauber R (2010) Efficient discrimination of oat and pea roots by cluster analysis of Fourier transform infrared (FTIR) spectra. Field Crop Res 119:78–84. https://doi.org/10.1016/j.fcr.2010.06.017

Nelder JA, Wedderburn RWM (1972) Generalized Linear Models. Royal Statistical Society Journal Series a: General 135:370–384. https://doi.org/10.2307/2344614

Neumann G, George TS, Plassard C (2009) Strategies and methods for studying the rhizosphere—the plant science toolbox. Plant Soil 321:431–456. https://doi.org/10.1007/s11104-009-9953-9

Niinemets Ü, Ostonen I (2020) Plant organ senescence above- and belowground in trees: how to best salvage resources for new growth? Tree Physiol 40:981–986. https://doi.org/10.1093/treephys/tpaa060

Ning Z-Y, Li Y-L, Yang H-L, Sun D-C, Bi J-D (2017) Carbon, nitrogen and phosphorus stoichiometry in leaves and fine roots of dominant plants in Horqin Sandy Land. Chinese Journal of Plant Ecology 41:1069

Nkouaya Mbanjo EG, Hershberger J, Peteti P, Agbona A, Ikpan A, Ogunpaimo K, Kayondo SI, Abioye RS, Nafiu K, Alamu EO, Adesokan M, Maziya-Dixon B, Parkes E, Kulakow P, Gore MA, Egesi C, Rabbi IY (2022) Predicting starch content in cassava fresh roots using near-infrared spectroscopy. Front Plant Sci 13. https://doi.org/10.3389/fpls.2022.990250

Norouzi S, Sadeghi M, Liaghat A, Tuller M, Jones SB, Ebrahimian H (2021) Information depth of NIR/SWIR soil reflectance spectroscopy. Remote Sens Environ 256. https://doi.org/10.1016/j.rse.2021.112315

Ozdemir A, Polat K (2020) Deep learning applications for hyperspectral imaging: a systematic review. Journal of the Institute of Electronics and Computer 2:39–56. https://doi.org/10.33969/JIEC.2020.21004

Udawatta RP, Rankoth LM, Jose S (2019) Agroforestry and biodiversity. Sustainability 11. https://doi.org/10.3390/su11102879

Paz Kagan T, Caras T, Herrmann I, Shachak M, Karnieli A (2017) Multiscale Mapping of Species Diversity under Changed Land-Use Using Imaging Spectroscopy. Ecol Appl: Publ Ecol Soc Am 27. https://doi.org/10.1002/eap.1540

Picon-Cochard C, Pilon R, Revaillot S, Jestin M, Dawson L (2009) Use of near-infrared reflectance spectroscopy to predict the percentage of dead versus living grass roots. Plant Soil 317:309–320. https://doi.org/10.1007/s11104-008-9810-2

Pierret A (2008) Multi-spectral imaging of rhizobox systems: New perspectives for the observation and discrimination of rhizosphere components. Plant Soil 310:263–268. https://doi.org/10.1007/s11104-008-9651-z

Pierret A, Moran CJ, Doussan C (2005) Conventional detection methodology is limiting our ability to understand the roles and functions of fine roots. New Phytol 166:967–980. https://doi.org/10.1111/j.1469-8137.2005.01389.x

Pudełko A, Chodak M, Roemer J, Uhl T (2020) Application of FT-NIR spectroscopy and NIR hyperspectral imaging to predict nitrogen and organic carbon contents in mine soils. Measurement 164. https://doi.org/10.1016/j.measurement.2020.108117

Qian Y, Ye M, Zhou J (2013) Hyperspectral Image Classification Based on Structured Sparse Logistic Regression and Three-Dimensional Wavelet Texture Features. IEEE Trans Geosci Remote Sens 51:2276–2291. https://doi.org/10.1109/TGRS.2012.2209657

R Core Team (2022) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/. Accessed 01 Feb 2024

Ran L, Zhang Y, Wei W, Zhang Q (2017) A Hyperspectral Image Classification Framework with Spatial Pixel Pair Features. Sensors 17. https://doi.org/10.3390/s17102421

Rewald B, Ephrath JE (2013) Minirhizotron Techniques. In: Eshel A, Beeckman T (eds) Plant roots: The hidden half, 4th edn. CRC Press, pp 735–750. https://doi.org/10.1201/b14550-50

Rewald B, Meinen C (2013) Plant roots and spectroscopic methods - analyzing species, biomass and vitality. Front Plant Sci 4. https://doi.org/10.3389/fpls.2013.00393

Rewald B, Meinen C, Trockenbrodt M, Ephrath JE, Rachmilevitch S (2012) Root taxa identification in plant mixtures–current techniques and future challenges. Plant Soil 359:165–182. https://doi.org/10.1007/s11104-012-1164-0

Rewald B, Rechenmacher A, Godbold DL (2014) It's Complicated: Intraroot System Variability of Respiration and Morphological Traits in Four Deciduous Tree Species. Plant Physiol 166:736–745. https://doi.org/10.1104/pp.114.240267

Rossel RAV, Jeon YS, Odeh IOA, McBratney AB (2008) Using a legacy soil sample to develop a mid-IR spectral library. Soil Research 46:1–16. https://doi.org/10.1071/SR07099

Rossi C, Kneubühler M, Schütz M, Schaepman ME, Haller RM, Risch AC (2022) Spatial resolution, spectral metrics and biomass are key aspects in estimating plant species richness from spectral diversity in species-rich grasslands. Remote Sensing in Ecology and Conservation 8:297–314. https://doi.org/10.1002/rse2.244

Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. Proceedings of the IEEE International Conference on Computer Vision, Barcelona, Spain. pp 2564–2571. https://doi.org/10.1109/ICCV.2011.6126544

Schmidt H, Nunan N, Höck A, Eickhorst T, Kaiser C, Woebken D, Raynaud X (2018) Recognizing Patterns: Spatial Analysis of Observed Microbial Colonization on Root Surfaces. Frontiers in Environmental Science 6. https://doi.org/10.3389/fenvs.2018.00061

Serbin S, Townsend P (2020) Scaling functional traits from leaves to canopies. In: Gamon JA, Townsend PA (eds) J Cavender-Bares. Remote Sensing of Plant Biodiversity. Springer, Cham, pp 43–82. https://doi.org/10.1007/978-3-030-33157-3_3

Sethy PK, Pandey C, Sahu YK, Behera SK (2022) Hyperspectral imagery applications for precision agriculture - a systemic survey. Multimedia Tools and Applications 81:3005–3038. https://doi.org/10.1007/s11042-021-11729-8

Shen X, Darmon F, Efros AA, Aubry M (2020) RANSAC-Flow: generic two-stage image alignment. arXiv [csCV]. https://doi.org/10.48550/arXiv.2004.01526

Straková P, Larmola T, Andrés J, Ilola N, Launiainen P, Edwards K, Minkkinen K, Laiho R (2020) Quantification of Plant Root Species Composition in Peatlands Using FTIR Spectroscopy. Frontiers in Plant Science 11. https://doi.org/10.3389/fpls.2020.00597

Stuart MB, McGonigle AJS, Willmott JR (2019) Hyperspectral Imaging in Environmental Monitoring: A Review of Recent Developments and Technological Advances in Compact Field Deployable Systems. Sensors 19. https://doi.org/10.3390/s19143071

Tang Y, Horikoshi M, Li W (2016) ggfortify: Unified Interface to Visualize Statistical Result of Popular R Packages. R J 8: 474–485. https://doi.org/10.32614/RJ-2016-060

Taylor K (2009) Biological flora of the British Isles: Urtica dioica L. J Ecol 97:1436–1458. https://doi.org/10.1111/j.1365-2745.2009.01575.x

Thenkabail P, Mariotto I, Gumma M, Middleton E, Landis D, Huemmrich K (2013) Selection of Hyperspectral Narrowbands (HNBs) and Composition of Hyperspectral Twoband Vegetation Indices (HVIs) for Biophysical Characterization and Discrimination of Crop Types Using Field Reflectance and Hyperion/EO-1 Data. Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of 6:427–439. https://doi.org/10.1109/JSTARS.2013.2252601

Thrupp LA (2000) Linking Agricultural Biodiversity and Food Security: the Valuable Role of Agrobiodiversity for Sustainable Agriculture. Int Aff 76:265–281. https://doi.org/10.1111/1468-2346.00133

Tong J, Xiang W, Lei P, Liu J, Tian D, Deng X, Fang X, Peng C (2016) Prediction of tree species composition in fine root mixed samples using near-infrared reflectance spectroscopy. Plant Biosystems - an International Journal Dealing with All Aspects of Plant Biology 150:412–419. https://doi.org/10.1080/11263504.2014.976291

Turker-Kaya S, Huck CW (2017) A review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis. Molecules 22. https://doi.org/10.3390/molecules22010168

Ustin SL, Gamon JA (2010) Remote sensing of plant functional types. New Phytol 186:795–816. https://doi.org/10.1111/j.1469-8137.2010.03284.x

Vamerali T, Bandiera M, Mosca G (2012) Minirhizotrons in Modern Root Studies. In: Mancuso S (ed) Measuring Roots: An Updated Approach. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp 341–361. https://doi.org/10.1007/978-3-642-22067-8_17

Vetterlein D, Doussan C (2016) Root age distribution: how does it matter in plant processes? A focus on water uptake. Plant Soil 407:145–160. https://doi.org/10.1007/s11104-016-2849-6

Vinay A, Rao AS, Shekhar VS, Kumar A, Murthy KB, Natarajan S (2015) Feature Extractionusing ORB-RANSAC for Face Recognition. Procedia Computer Science 70:174–184. https://doi.org/10.1016/j.procs.2015.10.068

Violle C, Navas M-L, Vile D, Kazakou E, Fortunel C, Hummel I, Garnier E (2007) Let the concept of trait be functional! Oikos 116:882–892. https://doi.org/10.1111/j.0030-1299.2007.15559.x

White KE, Reeves JB, Coale FJ (2011) Mid-infrared diffuse reflectance spectroscopy for the rapid analysis of plant root composition. Geoderma 167–168:197–203. https://doi.org/10.1016/j.geoderma.2011.08.009

Wickham H (2016) ggplot2: Elegant Graphics for Data Analysis. Springer, Cham. https://doi.org/10.1007/978-3-319-24277-4

Wojciechowska N, Marzec-Schmidt K, Kalemba EM, Ludwików A, Bagniewska-Zadworna A (2020) Seasonal senescence of leaves and roots of Populus trichocarpa—is the scenario the same or different? Tree Physiol 40:987–1000. https://doi.org/10.1093/treephys/tpaa019

Workman JJ Jr (1996) Interpretive Spectroscopy for Near Infrared. Appl Spectrosc Rev 31:251–320. https://doi.org/10.1080/05704929608000571

Wu Z, Ouyang G, Shi X, Ma Q, Wan G, Qiao Y (2014) Absorption and quantitative characteristics of C-H bond and O-H bond of NIR. Opt Spectrosc 117:703–709. https://doi.org/10.1134/S0030400X1411023X

Xu J-L, Gowen AA (2020) Spatial-spectral analysis method using texture features combined with PCA for information extraction in hyperspectral images. J Chemom 34. https://doi.org/10.1002/cem.3132

Yang X, Guo X (2014) Quantifying Responses of Spectral Vegetation Indices to Dead Materials in Mixed Grasslands. Remote Sensing 6:4289–4304. https://doi.org/10.3390/rs6054289

Yang X, Yu Y, Li M (2019) Estimating soil moisture content using laboratory spectral data. Journal of Forestry Research 30:1073–1080. https://doi.org/10.1007/s11676-018-0633-6

Yu K, Van Geel M, Ceulemans T, Geerts W, Ramos MM, Serafim C, Sousa N, Castro PML, Kastendeuch P, Najjar G, Ameglio T, Ngao J, Saudreau M, Honnay O, Somers B (2018) Vegetation reflectance spectroscopy for biomonitoring of heavy metal pollution in urban soils. Environ Pollut 243:1912–1922. https://doi.org/10.1016/j.envpol.2018.09.053

Zahir SADM, Omar AF, Jamlos MF, Azmi MAM, Muncan J (2022) A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. Sensor Actuat a-Phys 338. https://doi.org/10.1016/j.sna.2022.113468

Zhang C, Zhou X, Jiang J, Wei Y, Ma J, Hallett PD (2019) Root moisture content influence on root tensile tests of herbaceous plants. CATENA 172:140–147. https://doi.org/10.1016/j.catena.2018.08.012

Zhang W, Kasun LC, Wang QJ, Zheng Y, Lin Z (2022) A Review of Machine Learning for Near-Infrared Spectroscopy. Sensors 22. https://doi.org/10.3390/s22249764

Zhao Y, Zeng Y, Zhao D, Wu B, Zhao Q (2016) The Optimal Leaf Biochemical Selection for Mapping Species Diversity Based on Imaging Spectroscopy. Remote Sensing 8. https://doi.org/10.3390/rs8030216

Zhao J, Kumar A, Naik B, Balram M, Rajalakshmi P, Rewald B, Ninomiya S, Guo W (2022) Deep-Learning-Based Multispectral Image Reconstruction from Single Natural Color RGB Image—Enhancing UAV-Based Phenotyping. Remote Sensing 14. https://doi.org/10.3390/rs14051272

Zhou X, Armitage A, Prasad S (2016) Mapping mangrove communities in coastal wetlands using airborne hyperspectral data. 2016 8th Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), Los Angeles, CA, USA. pp 1–5. https://doi.org/10.1109/WHISPERS.2016.8071659