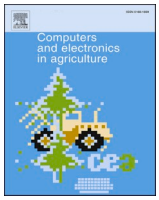
Computers and Electronics in Agriculture 229 (2025) 109746

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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



Original papers

Field-scale UAV-based multispectral phenomics: Leveraging machine learning, explainable AI, and hybrid feature engineering for enhancements in potato phenotyping

Janez Lapajne [a,*], Andrej Vončina [a], Ana Vojnović [b], Daša Donša [a], Peter Dolničar [b], Uroš Žibrat [a]

[a] Plant Protection Department, Agricultural Institute of Slovenia, Hacquetova ulica 17, Ljubljana 1000, Ljubljana, Slovenia
[b] Crop Science Department, Agricultural Institute of Slovenia, Hacquetova ulica 17, Ljubljana 1000, Ljubljana, Slovenia

ARTICLE INFO

Keywords:
Multispectral imaging
Potato research
Machine learning
Interpretability techniques

ABSTRACT

Fast and accurate identification of potato plant traits is essential for formulating effective cultivation strategies. The integration of spectral cameras on Unmanned Aerial Vehicles (UAVs) has demonstrated appealing potential, facilitating non-invasive investigations on a large scale by providing valuable features for construction of machine learning models. Nevertheless, interpreting these features, and those derived from them, remains a challenge, limiting confident utilization in real-world applications. In this study, the interpretability of machine learning models is addressed by employing SHAP (SHapley Additive exPlanations) and UMAP (Uniform Manifold Approximation and Projection) to better understand the modeling process. The XGBoost model was trained on a multispectral dataset of potato plants and evaluated on various tasks, i.e. variety classification, physiological measures estimation, and detection of early blight disease. To optimize its performance, nearly 100 vegetation indices and over 500 auto-generated features were utilized for training. The results indicate successful separation of plant varieties with up to 97.10% accuracy, estimation of physiological values with a maximum R^2 and rNRMSE of 0.57 and 0.129, respectively, and detection of early blight with an F1 score of 0.826. Furthermore, both UMAP and SHAP proved beneficial for comprehensive analysis. UMAP visual observations closely corresponded to computed metrics, enhancing confidence for variety differentiation. Concurrently, SHAP identified the most informative features – green, red edge, and NIR channels – for most tasks, aligning tightly with existing literature. This study highlights potential improvements in farming efficiency, crop yield, and sustainability, and promotes the development of interpretable machine learning models for remote sensing applications.

1. Introduction

Potatoes (*Solanum tuberosum* L.) rank fourth among the world's food crops, following wheat, corn, and rice. Their adaptability, ease of cultivation, short production cycle, and high yield make them a critical component of national food security (DeFauw et al., 2012; Wijesinha-Bettoni and Mouillé, 2019). Conventional farming practices, using synthetic pesticides for pest management, support food security and economic stability and help meet current demand (Cristache et al., 2018). However, in the European Union (EU), organic farming practices have been rapidly increasing since the 1990 s. (Offermann and Nieberg, 2002). These adhere to strict guidelines, reducing the use of synthetic pesticides, and emphasizing soil health, crop rotation, and integrated pest management (Geissen et al., 2021; Reganold and Wachter, 2016). Organic potato farming addresses consumer preferences for sustainably produced food and promotes biodiversity conservation (Pacifico and Paris, 2016). However, strict EU guidelines for organic farming limit the choice of crop varieties, as only selected ones meet all requirements (Röös et al., 2018). Advanced high-throughput phenotyping, a key aspect of phenomics, can aid in selecting suitable genotypes in breeding programs (Pasala and Pandey, 2020).

Potato varieties can be easily distinguished by numerous morphological characteristics (phenotypes) (Rozentsvet et al., 2024). Phenotyping techniques provide spectrally and spatially accurate data, enabling the quantification of agronomically relevant traits, such as leaf area, canopy architecture, chlorophyll content, and stress responses,

* Corresponding author.
E-mail address: janez.lapajne@kis.si (J. Lapajne).

https://doi.org/10.1016/j.compag.2024.109746
Received 8 August 2024; Received in revised form 24 October 2024; Accepted 29 November 2024
Available online 7 December 2024

including disease detection (Lammerts Van Bueren et al., 2011; Shakoor et al., 2017). The Distinctness, Uniformity, and Stability (DUS) testing protocols by the Community Plant Variety Office (CPVO) describe 10 different plant characteristics and 10 flower characteristics that define the growth of a particular variety. This provides a conventional approach using visual evaluations. Important plant characteristics for this type of phenotyping include foliage structure (leafy or stem type), growth habit (from spreading to erect), size and openness of the leaves, intensity of green color, and anthocyanin coloration of the foliage (stems and leaves) (Community Plant Variety Office, 2024; Dolničar et al., 2017).

Aforementioned traditional methods can fall short in disease detection over large areas. They require significant human effort by extensively trained experts, and struggle to provide spatial diagnostic results for plant diseases over large areas (Abbas et al., 2023). These issues are compounded when symptoms are not species-specific or vary significantly between varieties. In these cases, molecular methods are required for accurate identification, increasing human effort and overall costs of assessments.

Several of these issues are addressed by precision agriculture methods, such as multispectral imaging. Coupled with unmanned aerial vehicles (UAVs) it offers numerous benefits, including reduced human effort, objective and quantitative measurements, non-destructive monitoring, early detection of plant health issues and cost-effectiveness (Abbas et al., 2023; León-Rueda et al., 2022). Optical imaging sensors, capable of measuring spectral alterations in plant information due to stress exposure, enable the assessment of various stressors by observing changes in the plant's spectral signature, though these sensors do not directly measure plant physiological characteristics. Measurements of biochemical characteristics, such as pigment content, are hindered by complex plant tissue and presence of accessory pigments, proteins, carbohydrates and other molecules (Fernández-Marín et al., 2018; Sujatha, 2015). This underscores the crucial role of developing new interpretable methods for precise prediction making.

Machine learning methods are increasingly utilized to model various plant characteristics, stress presence, and physiological parameters, based on remote sensing data. This trend is driven by the growing volume of available data, and the complex, nonlinear interactions between variables. In potato plant research, the studies focused on disease detection (Lizarazo et al., 2023; Rodríguez et al., 2021; Sun et al., 2023), chlorophyll content (Yang et al., 2022; Yin et al., 2023), biomass estimation (Liu et al., 2022a, 2022b), yield estimation (Li et al., 2021) and accessing other characteristics of plants (Yu et al., 2023). While various models were utilized, multiple studies have identified XGBoost (eXtreme Gradient Boosting) as one of the most performant (Attia et al., 2022; Li et al., 2023; Shwartz-Ziv and Armon, 2021).

Features extracted from remote sensing images, such as spectral information, have been proven useful in enhancing the performance of machine learning models (Wei et al., 2022; Zeng et al., 2023). Empirically devised vegetation indices were demonstrated to be beneficial for analyzing various crop stress effects. They offer rapid processing capabilities, and can establish robust correlations between plant stress or illness and the corresponding patterns of reflected light (Han et al., 2022; Yin et al., 2023). Despite their utilization as effective proxies for physiology measurements, stress presence and overall crop condition (Abdelbaki et al., 2021), the performance of these models can be further boosted by generating additional features, such as new calculations or transformations, which can capture more complex patterns and relationships in the data (Lu et al., 2020; Sethy et al., 2022; Yang et al., 2023; Yu et al., 2023).

Explainable AI has emerged as a critical field in machine learning, aiming to address the 'black box' nature of many models, which are often criticized for their lack of interpretability (Jones et al., 2022). This is not only crucial for understanding model behaviors but also for extracting knowledge from modeling efforts (Li et al., 2023). Methods such as SHapley Additive exPlanations (SHAP) (Lundberg and Lee, 2017) offer deeper insights into trained models and enhance the capabilities of machine learning models (Nohara et al., 2022; Rodríguez-Pérez and Bajorath, 2020). Dimensionality reduction methods, e.g., UMAP (Uniform Manifold Approximation and Projection) (McInnes et al., 2018), can be used in conjunction with SHAP to visualize high-dimensional data. This aids in the interpretation of complex machine learning models, providing a more comprehensive understanding of the model's behavior (Armstrong et al., 2021). SHAP has been utilized in various crops for disease detection (Zeng et al., 2023), water status assessment (Wei et al., 2022), maize lodging (Han et al., 2022) and cotton yield (Jones et al., 2022). But the combination of SHAP and UMAP has not yet been utilized for multispectral imaging-based phenotyping and plant health assessment.

This study introduces a modern pipeline stack designed to process high-resolution multispectral UAV-based images, addressing requirements for assessing the characteristics and health conditions of potato plants. The focus was classification of varieties, prediction of physiological measurements, and detection of early blight (i.e., *Alternaria* sp. fungi), while also considering the explainability of the trained models. To enhance the model's performance, extensive feature engineering was employed, resulting in a vast array of calculated and generated features.

The main contributions of this paper are:

- We investigate the potential of using multispectral UAV-based imaging for high-throughput phenotyping and detection of diseases. This exploration is conducted on self-acquired datasets and includes an examination of the disparities between organically and conventionally grown potato plants.
- We adopt a novel approach that auto-generates features through simple mathematical formulations and combines these with a multitude of calculated vegetation indices.
- This study represents the first known attempt to integrate XGBoost, SHAP, and UMAP methods in spectral high-throughput phenotyping research of potato plants.
- Lastly, we propose a general analysis pipeline for similar problems and commit to open science by making the entire codebase and datasets publicly available for use by other researchers.

2. Materials and methods

2.1. Experiment setup

The experiment was conducted in Central Slovenia (46°08′31.3″N, 14°33′20.4″E) at an elevation of 320 m above sea level with a subalpine climate (average temperature: 9.5 °C, average precipitation: 1352 mm). Eight potato varieties (Table 1) were planted in a randomized block design with four replications (blocks). Each block contained one microplot of a single potato variety. Each plot consisted of a 6x3 meter section containing 70 ± 2 potato plants distributed into four rows. The experiment was conducted in two different fields, one organic ("eko") and one conventional production ("konv"). Potato seeds were planted in mid-April and the crop was grown during the 2022 growing season, which typically runs from April to September in Central Slovenia. An illustrative instance of the experimental design is presented in Fig. 1. The numbered blocks within the field and a section depicting the spatial occupation of an arbitrary variety within a given block are delineated in the figure.

2.2. Data acquisition

Multiple aerial scans and ground measurements were conducted at three distinct time points, on June 15th, 2022, July 11th, 2022, and July 20th, 2022, for both experiments. This ensured a direct correlation between the acquired airborne scans and the corresponding ground measurements. The acquisition time steps were selected to capture distinct

Table 1
Morphological characteristics of potato plant varieties. The columns represent the variety name, foliage structure, growth habit, plant height, and foliage color; as defined by CPVO (Community Plant Variety Office, 2024).

Variety	Foliage structure type	Growth habit	Height	Color
KIS Tamar	leafy semi prostrate	late	Low	dark green
KIS Blegoš	leafy prostrate	medium early	Low	light green colour
Carolus	upright	medium late	Medium	medium dark green
Alouette	upright	medium late	Tall	dark green + purple pigmentation
Levante	semi upright	medium late	Medium	light green colour
Twister	semi prostrate	medium late	Medium	medium green
Otolia	upright stem	medium early	Tall	medium dark green
KIS Kokra	leafy semi prostrate	medium late	medium	light green colour

stages of plant development, starting when the potato plants had at least five fully developed leaves (but typically more), ensuring clear identification in multispectral images with sufficient spatial coverage.

2.2.1. *Airborne multispectral imaging*

Multispectral images of potato field crops were acquired using a SKYHERO SPYDER X4-850 GEO Edition Unmanned Aerial Vehicle (UAV) equipped with a Micasense Rededge MX 5-band multispectral camera (Table 2). A Stonex S9i GNSS device was used in conjunction with ground control points (GCP) for georeferencing and to increase spatial accuracy of acquired images. Positioning and distribution of GCPs followed best practice guidelines, to ensure spatial accuracy in the acquired imagery (Fig. 1) (D S et al., 2023; Pepe et al., 2018). Imaging was performed around noon, maintaining a UAV altitude of 50 m above ground level, and a flight speed of 5 m/s, capturing GSD (ground sampling distance) of 3.5 cm/pixel. The flight plan incorporated an 80 % bidirectional overlap and a 25 % buffer to ensure data completeness and redundancy. Following acquisition, a standard series of processing procedures was executed for the images using PIX4DMapper ("Pix4Dmapper," 2024) and QGIS (QGIS Development Team, 2024). This included georeferencing, orthorectification, radiometric correction, conversion to reflectance, and the construction of mosaics to ensure accurate and calibrated multispectral data for analysis.

2.2.2. *Field measurements and plant health assessment*

Ground measurements encompassed various parameters, with leaf chlorophyll concentration (labeled SPAD) quantified using SPAD-502 (Konica Minolta, Japan) and Li 600 (LI-COR Biosciences, US) for plant physiology, such as stomatal conductance (gsw), transpiration (E), quantum efficiency in light (PhiPS2), and electron transport rate (ETR). The presence of early blight was determined through visual assessment by qualified evaluators following established protocols for potato health assessment (Fig. 2). Measurements were conducted within the central area of each plot (Fig. 1) to mitigate border effects. Four plants of each variety were selected, measurements taken on three leaves of each plant, and the average calculated. This resulted in 16 measurements per variety per image and a cumulative 128 measurements per imaging session across all eight varieties. This yielded a total of 768 measurements (384 for organic and 384 for conventional) across all three time points. Violin plots for all parameters are depicted in Figures A 1–5. Early blight visual assessments were exclusively conducted during the final time point, yielding 256 measurements (128 for organic and 128 for conventional). Additionally, coordinates were recorded for all 256 plants during the

Table 2
Micasense Rededge MX spectral specification.

Acronym	Name	Centre Wavelength [nm]	Bandwidth [nm]
B	Blue	475	20
G	Green	560	20
R	Red	668	10
RE	Red-edge	717	10
N	Near-infrared	840	40

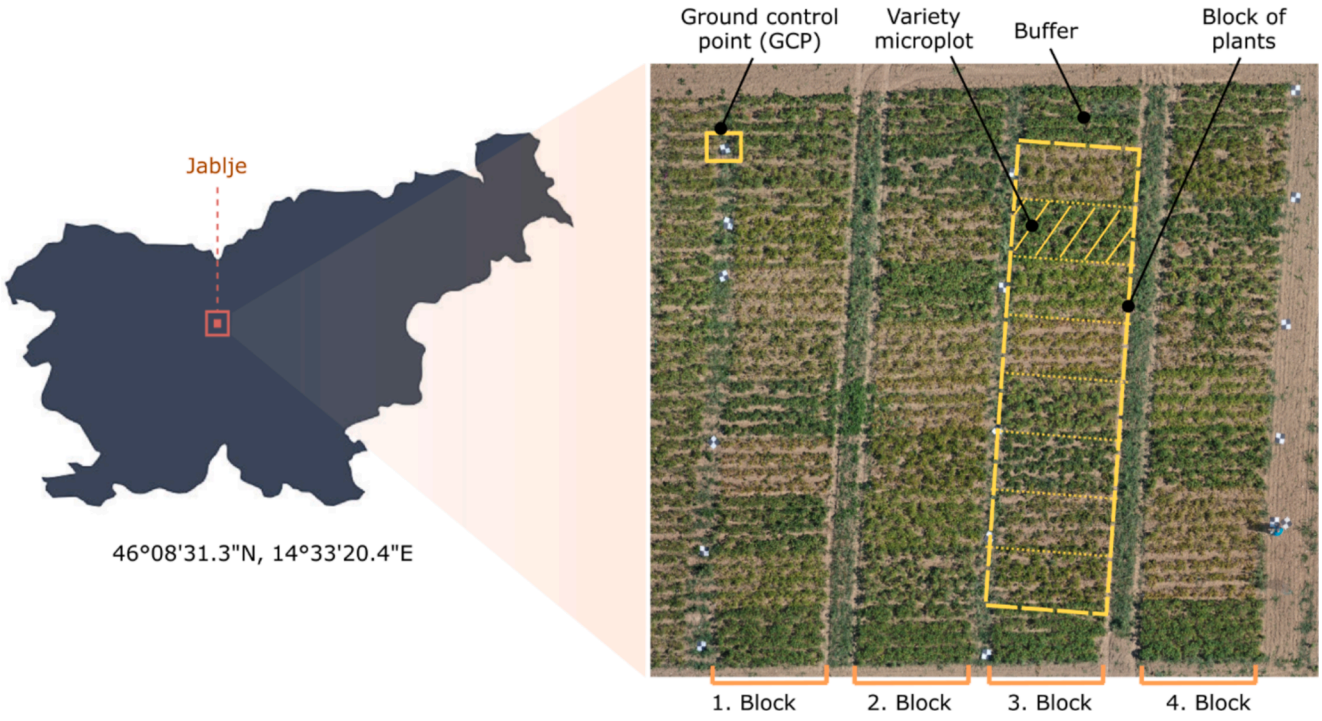


Fig. 1. Organic potato field, acquired on July 20th, 2022. The GCPs, blocks and microplots labeled.

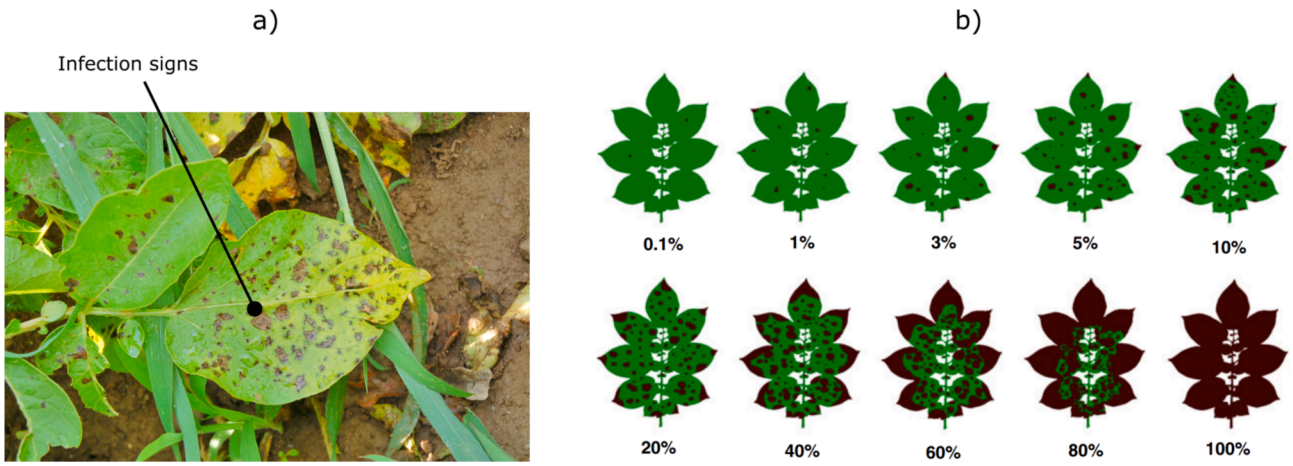


Fig. 2. a) visual signs of early blight presence and b) standard area diagrams to estimate severity of potato early blight (Duarte et al., 2013).

first time point, enabling direct pixel-to-measurement matching in the image. These coordinates were then also used for the two subsequent time points.

2.3. Processing pipeline

2.3.1. Overall pipeline description

The processing pipeline (Fig. 3) was structured into five distinct stages with the overarching goal of transforming raster images, GNSS coordinates, and field measurements into a format suitable for further analysis. The initial phase involved data loading from multiple sources i.e. spectral images and measurement files. Then the data manipulation stage formatted the data according to the analytical problem at hand, be it classification or regression, and partitioned it into training and testing datasets (train and test data) for model training and evaluation. In the feature engineering phase, new features were calculated for both the training and testing datasets, derived from initial reflectance values to augment the model's performance. The modeling stage utilized the training dataset to train the model and optimize its hyperparameters, with the best performing model being registered and saved into a model registry for later use. Finally, the evaluation stage involved deploying the optimized model from the registry to assess its performance on the testing dataset. The code of the pipeline was implemented in Python 3.10 programming language. This systematic pipeline ensured a methodical and organized approach to the analysis of multispectral images. Subsequent subsections provide a more comprehensive

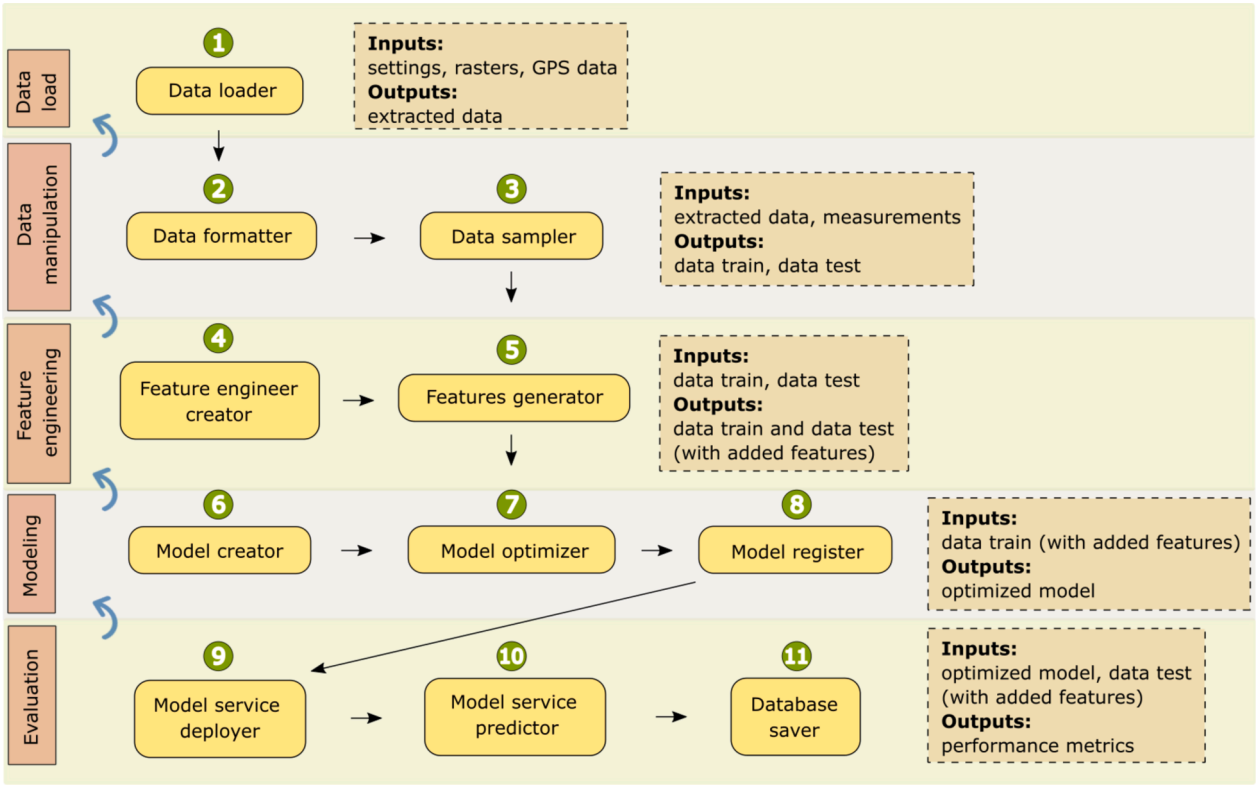


Fig. 3. Processing pipeline. Red rectangles signify the five stages, delineating steps based on their functionality. Yellow bubbles, sequentially numbered, represent individual steps. Brief input and output descriptions accompany each stage on the right. Blue, curved feedback arrows indicate optional backtracking through saved logs and artifacts.

description of each stage.

2.3.2. Data load

The georeferenced, orthorectified, and reflectance-transformed images, as detailed in Section 2.2.1, encapsulated the spectral information across entire fields. Extraction of leaf area pixels was guided by GNSS locations provided in a separate file. The data loader computed the coordinates of the ten closest points for each plant in the raster via Euclidean distance and extracted the spectral channel values for each of these points. The points were thoroughly analyzed, ensuring that mostly the extent of the potato plants was included, excluding majority of background pixels. This was achieved using QGIS, to ensure that the central pixel accurately fell in the middle of each plant.

These values were then averaged along the spectral channel, resulting in a singular value per channel, i.e. a mean reflectance of each plant. These data were fused with metadata, such as date, treatment, block, plant, and variety, into a new row within a table. Table 3 provides a representative illustration of this process, although the presented data is fabricated for illustrative purposes.

2.3.3. Data manipulation

The dataset underwent partitioning into training (75 %) and testing (25 %) subsets facilitated by a data sampler. Partitioning was guided by metadata variables, specifically date, variety, treatment, and block, ensuring a stratified approach in all instances, aimed to prevent overfitting, particularly with respect to individual varieties. This split ensured a balanced representation of metadata categories in both training and testing datasets, contributing to the robustness of subsequent model training and evaluation. Diverse datasets were generated (see subsections 2.3.3.1, 2.3.3.2, and 2.3.3.3) to evaluate the applicability of machine learning on relevant challenges in precision agriculture: 1. Classification of varieties, 2. Estimation of physiology measurements and 3. Detection of diseases (i.e. early blight).

2.3.3.1. Data for varieties classification. The target variable (variety name) was directly constructed from metadata, resulting in an encoded label that represented each variety across all cases. Data instances (i.e. plants) were evenly distributed across all blocks, varieties, treatments, and imaging sessions to achieve balanced data stratification. As a result, the dataset used for varietal classification consisted of 2976 instances for testing and 8928 instances for training; all plants from each microplot were included. Variety classifications were performed in a systematic approach with several iterations. The initial evaluation was focused on distinguishing between two varieties (KIS Krka and KIS Blegoš). This was followed by the addition of one variety at a time in each subsequent iteration. The selection of the first two varieties and a new variety to include was based on algorithm's ability to provide the best separation, as determined by the highest F1 scores (see sections 2.3.5 and 2.3.6). This stepwise integration of varieties allowed metrics to be calculated at each stage, and to evaluate the maximum achievable metrics for the investigated varieties. This approach was deemed necessary due to the potential presence of varieties with highly similar spectral signatures. Their inclusion in the initial classification model could have hindered successful class separation.

2.3.3.2. Data for regression analysis. Since physiological measurements were conducted on four plants of each variety within each block, a smaller number of instances were available, as for variety classification. The measurements were performed during every imaging session on predefined plants, maintaining an equal number of measurements across varieties and blocks (as described in 2.2.2 *Field measurements and plant health assessment*). This consistency in following measurement protocols provided an equivalent number of instances for predicting any physiological value, maintaining uniformity across all variables. The dataset comprised a total of 192 instances for testing and 576 for training. The descriptive statistics of either target values are written in Table 4.

2.3.3.3. Disease detection. Detection of early blight was formulated as a binary classification problem. Given the uncontrolled nature of early blight appearance, an imbalance existed between the two classes, with a higher prevalence of infected plants. Three varieties (Twister, Otolia, and KIS Blegos) were excluded from the dataset as they lacked healthy plants. The test set comprised 40 data instances (Fig. 4), consisting of 21 healthy and 19 infected plants, whereas for training 120 instances were available, with 63 healthy and 57 infected plants. To further address the imbalanced dataset issue, SMOTE (Synthetic Minority Oversampling Technique) (Chawla et al., 2011) was applied to the training dataset. This method was selected to foster a more equitable representation of both minority and majority classes in the machine learning model. It tackles this by creating synthetic samples for the underrepresented class, aiming to improve model performance on the minority class without overfitting. As a result, the training dataset consisted of 240 instances, evenly distributed with 120 instances each for healthy and infected plants. This balanced allocation ensured that, from each variety, 24 instances belonged to healthy plants, and 24 instances belonged to infected plants.

2.3.4. Feature engineering

The importance of feature engineering in influencing and substantially enhancing results is well-established (Heaton, 2017; Shin and Oh, 2021; Verdonck et al., 2024). In our study, a conscientious approach to feature engineering was adopted, wherein the process was individually applied to the training and testing datasets to prevent any potential data leakage (Apicella et al., 2024). Features were systematically generated and appended to the corresponding rows in the data table.

A hybrid feature engineering strategy was employed, encompassing

Table 4
Descriptive statistics of all available data instances. Mean – average, Std. – standard deviation, Min. – minimum value and Max. – maximum value.

Target	Units	Mean	Std.	Min	Max
E	mmol m^{-2} s^{-1}	2.88	1.69	0.04	10.48
ETR	μmol electrons m^{-2} s^{-1}	251.67	77.29	56.43	466.55
gsw	mmol m^{-2} s^{-1}	0.11	0.07	0.01	0.53
PhiPS2	/	0.38	0.12	0.08	0.73
SPAD	/	4.37	0.64	2.62	6.09

Legend: Abbreviations: E – transpiration; ETR – electron transport rate, gsw – stomatal conductance, PhiPS2 – quantum efficiency in light, SPAD – leaf chlorophyll concentration.

Table 3
Illustration of the initial rows produced during the data loading phase.

Index	Reflectance					Metadata				
	Blue	Green	Red	Red edge	NIR	Date	Treatment	Block	Plant ID	Variety
1	0.33	0.47	0.41	0.12	0.03	2022_06_15	eko	1	3	Carolus
2	0.17	0.43	0.47	0.36	0.25	2022_06_15	eko	1	2	Twister
3	0.43	0.20	0.43	0.48	0.28	2022_06_15	konv	2	1	KIS_Tamar
4	0.24	0.24	0.48	0.05	0.47	2022_06_15	konv	3	1	Levante
…	…	…	…	…	…	…	…	…	…	…

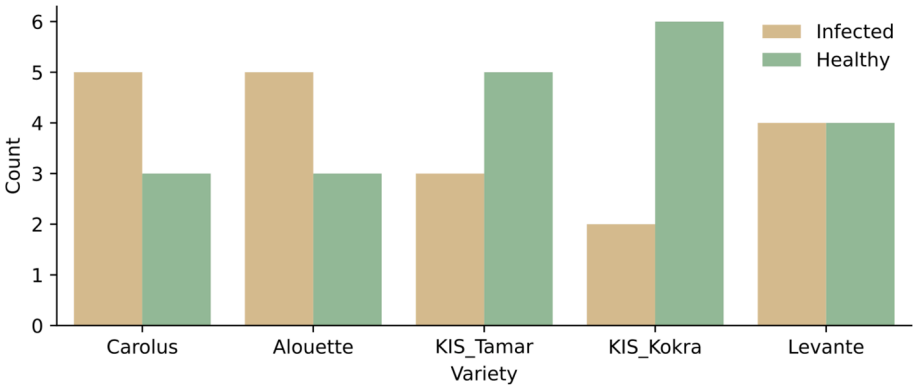


Fig. 4. Distribution of data instances in case of early blight classification for test set.

three categories of features:

I. Non-modified reflectance values (black label): these were the reflectance values of the five bands (blue, green, red, red-edge, and near-infrared).
II. Automatically generated features (green label): These additional features were derived from the reflectance values using elementary mathematical formulations, as described in (Horn et al., 2019) (e.g., red-edge/green, sqrt(blue)/NIR). This process aimed to extract more intricate information from the raw data, generating in total 530 new features.
III. Calculated spectral indices (red label): These indices are designed to capture specific aspects of the material properties under investigation (e.g., NDVI, CVI (Arevalo-Ramirez et al., 2022)). Combining multiple indices enhances the ability to discriminate between plant health conditions, improving overall inference performance. A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research was used for selection (Montero et al., 2023). Employed (98) spectral indices are included in the supplementary material.

Each practice retained only those features that demonstrated a performance boost, determined by Sequential feature selector (SFS) (Pudil et al., 1994), typically resulting in fewer than 20 features per practice. To remove similar features, pairwise Pearson's correlations were calculated between all features. Those with correlations exceeding an empirically predefined threshold of 0.99 were eliminated. This high threshold was chosen to remove only those features that were nearly identical in the information they provide. By eliminating features with correlations exceeding 0.99, the redundancy was reduced (Akoglu, 2018), thereby improving computational efficiency without sacrificing valuable information from less correlated features (Dormann et al., 2013).

2.3.5. Modeling

The XGBoost (eXtreme Gradient Boosting) algorithm (Chen and Guestrin, 2016; Shwartz-Ziv and Armon, 2021) was used for both classification and regression. It was trained on a dataset enriched with engineered features, ensuring the model's ability to learn complex relationships within the data. Hyperparameter tuning (Table 5) was performed using an automated hyperparameter optimization software framework Optuna (Akiba et al., 2019), conducting 200 trials. It uses the Tree-structured Parzen Estimator (TPE) algorithm (Bergstra et al., 2011; Watanabe, 2023) for hyperparameter tuning. TPE is a Bayesian optimization method that models the objective function to efficiently explore the hyperparameter space. The optimization objective involved 3-times repeated 5-fold cross-validation (Bates et al., 2021), yielding either an averaged F1 score for classification or RMSE for regression. The model was always initiated with the same random seed to ensure experiment reproducibility. The tuning process was tailored to the small dataset, aiming to optimize model configuration for enhanced performance. The version of the model achieving the highest performance score (F1 or RMSE) was archived in the model registry for future deployment and evaluation on the test set.

Table 5
Hyperparameters of XGBoost explored in optimization procedure.

Parameter	Range	Default Value
N estimators	100 to 1000	100
Max depth	3 to 6	3
Learning rate	0.01 to 0.2	0.1
Min child weight	1.0 to 10.0	1.0
Subsample	0.5 to 0.8	1.0
Colsample bytree	0.5 to 0.8	1.0
Reg lambda	1.0 to 10.0	1.0

2.3.6. Evaluation

For final evaluation, each model was initially deployed from the model registry and assessed against the test dataset. An array of metrics, including accuracy, balanced accuracy (bAccuracy), F1 score, precision, and recall for classification and R^2, mean absolute error (MAE), root mean square error (RMSE), normalized root mean square error (NRMSE), RMSE standardized by the range of values (rNRMSE) (Abdelbaki et al., 2021), and max error (ME) for regression, were utilized to comprehensively evaluate the model's performance. In the case of the multiclass classification problem, the F1 score, precision, and recall were calculated using a weighted approach, where the metrics for each class were averaged, weighted by the number of true instances (support) for each class. All metrics and data visualization plots were methodically stored, portraying both data integrity and the decision-making process of the model, along with its achieved performance. To facilitate easy retrieval, these results were systematically stored in a database.

2.4. Statistical analysis of physiology parameters

Linear Mixed Models (LMM) (Gałecki and Burzykowski, 2013) fitted by Restricted Maximum Likelihood (REML) were used to analyze the effects of varieties and treatments on physiological parameters. The model formula used was:

$$y \sim \text{variety} + \text{treatment} + (1|\text{block}) + (1|\text{date})$$

where the variables were defined as follows:

- Date: encoded as an ordinal variable, with values ranging from 0 to 2 representing each consecutive date.

- Treatment: categorical variable indicating the type of treatment, either “conventional” or “organic.”
- Block: categorical variable representing different blocks by number
- Variety: categorical variable indicating different varieties by their names.

The model included fixed effects for variety and treatment, and random intercepts for block and date, allowing for the assessment of the impact of these factors on the response variable y. A Box-Cox transformation (Sakia, 1992) was applied to the measurements to stabilize variance and make the data more normally distributed. P-values were used to determine the statistical significance of pairwise comparisons between the varieties within each treatment category. A threshold of $\alpha = 0.05$ was applied, with p-values less than this value indicating a statistically significant difference. All statistical analyses were performed using R Statistical Software (R 4.4.1).

2.5. Exploratory data analysis and explainable AI techniques

A combination of exploratory data analysis (EDA) and explainable artificial intelligence (XAI) techniques was employed to enhance understanding of the dataset and model outcomes. Impacts of individual features in the XGBoost models were evaluated using SHAP (SHapley Additive exPlanations) values (Lundberg and Lee, 2017), enabling a more interpretable and transparent analysis. All three categories of features (I., II. and III) were used in the evaluation. Among all, 10 most prominent features were investigated further. To address the challenges posed by high-dimensionality datasets, comprising diverse features (Lapajne et al., 2022), Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2018) was utilized in an unsupervised fashion. This dimensionality reduction technique allowed for the visualization of intricate relationships within the dataset, aiding in the discrimination between different potato varieties. To objectively evaluate the success of UMAP’s dimensionality reduction, a Support Vector Machine (SVM) (Lapajne et al., 2022) with a Radial Basis Function (RBF) kernel was employed. Its hyperparameters were not optimized, since it was directly

Fig. 5. UMAP visualization for well separated potato plant varieties. Circles and triangles represent organic and conventionally grown potato plants. Colored areas depict SVM decision boundaries, reflecting separation quality between varieties, expressed with accuracy score. Full UMAP results are provided in Supplementary material.

used on the dimensionally reduced data obtained through UMAP. This allowed for the visualization of decision boundaries in lower-dimensional space and the calculation of accuracy scores, thus mitigating potential overfitting, as the boundaries can be visually inspected for irregularities or excessive complexity.

3. Results

3.1. Visual exploration of varieties and statistical results

A high degree of similarity in the underlying features was observed for each variety as most data points grouped together, regardless of treatment (Fig. 5). Exceptions were observed, e.g., within the organic treatment of KIS Kokra, where a subset of data points tends to cluster with instances from the Otalia variety. By combining UMAP and SVM, distinct boundaries were identified between KIS Tamar (late small dark green semi prostrate phenotype) and other very different looking varieties, such as KIS Blegoš (early, light green), Alouette (tall dark green with anthocyanin pigmentation), Levante (semi prostrate light green), Otolia (erect stem type), Carolus and Twister. Notably, the separation between KIS Tamar and KIS Blegoš appeared to be the most pronounced, even though they are both similar sized, prostrate type, but the opposite intensity of green coloration. Additionally, a clear separation boundary was observed between KIS Kokra and KIS Blegoš, and KIS Kokra and Otolia. Other differences were a bit more subtle, certain groups of data points set apart Alouette and KIS Kokra, Carolus and Twister, Carolus and KIS Tamar, KIS Kokra and Otolia, and KIS Kokra and KIS Tamar, among others (Figure A 7). Notably, there are instances where visual separations are indiscernible between varieties, such as KIS Blegoš and Levante, Carolus and Levante, Otolia and Levante, KIS Blegoš and Twister, Levante and Twister, highlighting instances of potential similarity between these pairs (Figure A 8). In this case, the visualized separation boundaries appear randomly positioned due to the absence of a clear distinction between the data clusters.

LMMs indicate that varieties had a significant effect for every physiological parameter (Tables A 2–16). The effect of treatment was significant only for the variables gsw and E, where conventional treatment had a negative impact. Among non-significant effects, ETR was also lower in the conventional treatment, while PhiPS2 and SPAD were higher. Estimates for variety Allouette in organic treatment (the reference) were not significant for any response variable, suggesting no substantial baseline effects. Variety KIS Tamar consistently exhibited statistically significantly higher values in all response variables, compared to variety Aloutte. The fewest significant differences were observed for response variable E, and most for SPAD, where all varieties except one (KIS Tamar) had a lower response. For SPAD, all varieties except KIS Kokra significantly differed from Allouette. For gsw, no significant differences were found between KIS Kokra, Otolia, and Twister compared to Allouette; KIS Tamar had a positive estimate while others had negative. For ETR, no significant differences were observed between KIS Kokra and Carolus relative to Allouette, with KIS Tamar showing a positive estimate among significant varieties. For PhiPS2, Carolus and Otolia showed no significant differences from Allouette, whereas KIS Kokra and KIS Tamar had higher values and other varieties had lower values. Regarding E, no significant differences were found between KIS Kokra, Carolus, and Twister compared to Allouette, with KIS Tamar again showing higher values. In all response variables there was a comparatively large amount of unexplained variability. Among the two random effects, date of data acquisition explained more of the variability than experimental block.

Statistical analysis of physiological values produced results consistent with the UMAP methodology, showing substantial overlap between the findings of both approaches. Specifically, no significant differences were observed between treatments for either variety or any physiological parameter. However, differences were noted between the varieties themselves. The most pronounced differences were observed in SPAD measurements (Fig. 6), where KIS Tamar, separated distinctly from all other varieties in both treatments. This finding aligns with the UMAP analysis, where KIS Tamar was identified as the most distinct variety, since SPAD values were generally higher compared to other varieties (Figure A 5). Similarly, SPAD measurements showed that KIS Blegoš also differed notably from most other varieties. In this case, the values were lower compared to other varieties. In the gsw measurements (Figure A 2), conventionally grown KIS Tamar variety remained the most distinct, while organically grown KIS Tamar was more similar to other varieties. Similarly to SPAD, gsw measurements were significantly higher for conventional treatment. Interestingly, the organically grown Levante variety showed significant differences compared to other varieties in gsw measurements (Fig. 6) and conventional grown in E measurements (Figure A 6). For measurements of ETR and PhiPS2 no distinct patterns could be observed.

3.2. Evaluation of machine learning models

Variety identification achieved the highest accuracy in the discrimination of two varieties (F1 = 0.971). Conversely, the lowest score of 0.476 was obtained when distinguishing among eight varieties (Table 6). The varieties were integrated in the following order: KIS Tamar, KIS Blegoš, Otolia, Carolus, Alouette, KIS Kokra, Levante and Twister. The incorporation of new varieties led to an increase in the dataset size, resulting in the utilization of a larger number of data points for both training and testing phases. The scores in between exhibited a continuous decrease. Precision and Recall values closely mirrored the F1 score, suggesting that emphasis was not disproportionately placed on any particular variety by the model, maintaining a balanced performance across different varieties. The inclusion of varieties follows the separation ability of UMAP. KIS Blegoš and KIS Tamar are included first due to their clear separation. However, determining the third variety based solely on UMAP already becomes challenging, as overall similarity to both previously included varieties must be considered. It becomes a matter of estimation to predict which varieties will show the most significant discrimination when multiple are separated simultaneously. For instance, Otolia, Carolus, and Allouette generally exhibit good separation from almost all other varieties in one-to-one comparisons (Fig. 5, Figure A 7). Conversely, some varieties, such as Levante and Twister, tend to be less distinguishable. Levante, in particular, frequently appears in the diagram of poorly separated varieties (Figure A 8).

Comparing the RMSE from the test set with RMSE-CV from the train set showed no overfitting during the optimization run. Predictions always surpassed a random base model, with a minimum R^2 of 0.313 (Table 7). The most accurate predictions were observed for SPAD, achieving an R^2 of 0.570 and rNRMSE of 0.129. In all cases, the rNRMSE metric has reached values below 0.2, signifying a relatively modest RMSE within the overall range, and thus comparatively good prediction efficacy. Comparing MAE and RMSE, the former showed somewhat lower values, but the difference is comparatively small, indicating a uniform distribution of errors. On the other hand, ME shows significant deviations from both RMSE and MAE, with ME ranging from 1.7-times (gsw) to 3.8-times (ETR) higher than RMSE. This indicates that while the models generally perform well, there are a few outliers where the prediction errors are much larger.

Detection of early blight infected potato plants achieved an F1 score of 0.825, indicating that the model did not exhibit overfitting toward either class (Table 8). The highest accuracy was achieved for Levante, with an F1 of 1.000, while Alouette achieved the lowest performance, with an F1 of 0.620. It is imperative to acknowledge that certain metrics may be skewed due to the small test set (8 plants per variety), and the distribution difference between the two classes. For instance, although the metrics for the KIS Kokra variety indicated relatively high values, it is essential to note that there were twice as many instances of healthy data as there were infected ones available for testing (Fig. 4).

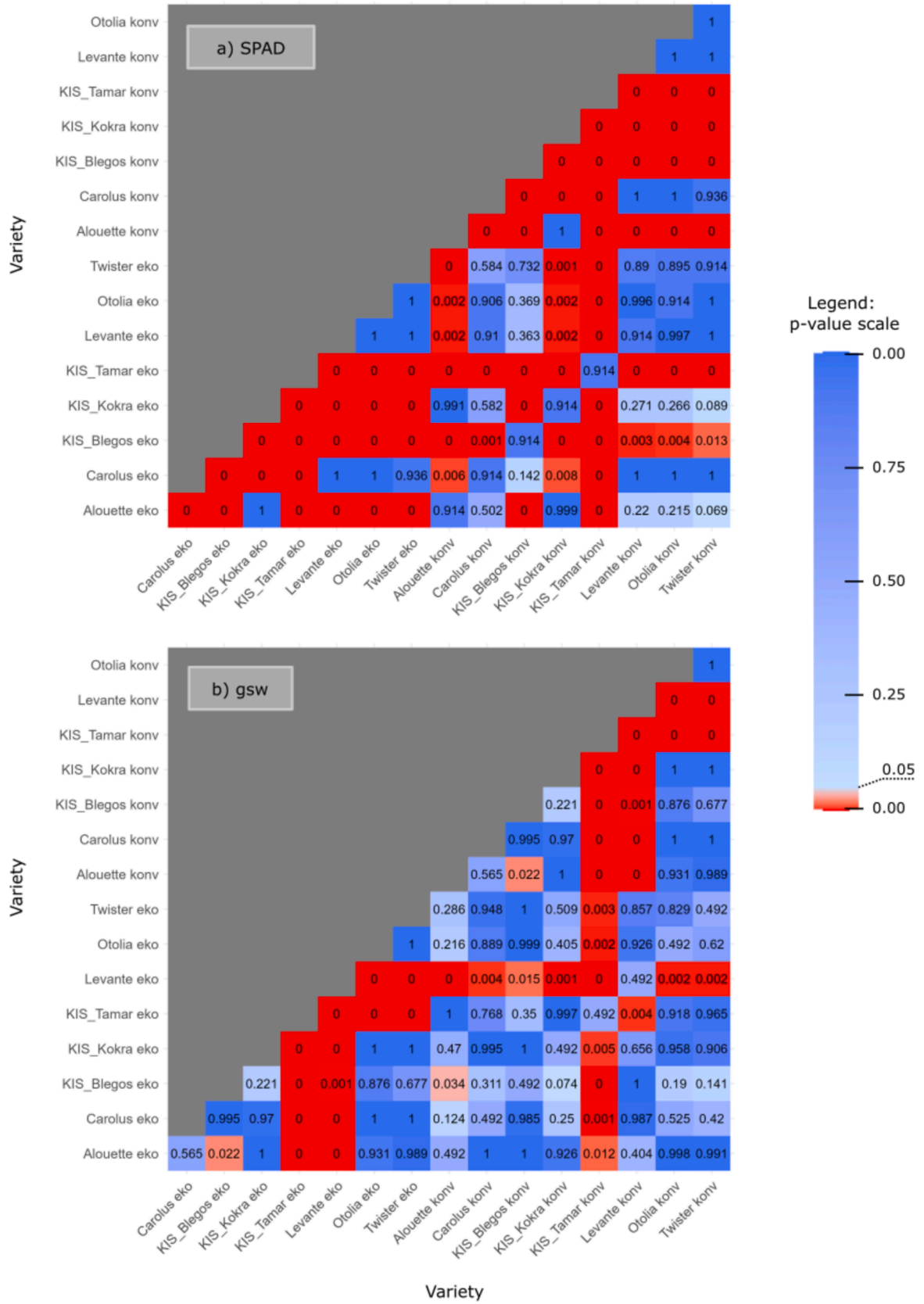


Fig. 6. Statistical Analysis of physiology parameters: a) SPAD and b) gsw. Significant differences between pairs are indicated in red, while non-significant differences are shown in blue. The shade intensity of both colors represents the p-value size, with lighter shades indicating values near the decision boundary ($\alpha = 0.05$) and stronger shades representing more extreme values (closer to 0 or 1). A bar plot has been added alongside, illustrating the shades and referencing the corresponding p-values. Full results are provided in Supplementary material.

Table 6
Varieties classification metrics. The term "# Varieties" denotes the total number of varieties included in the analysis. The columns labeled "# Train" and "# Test" represent the respective numbers of data instances (i.e. plants) employed for training and testing the model. The "F1 − CV" column showcases the performance score attained through model optimization using cross-validation (CV).

# Varieties	# Train	# Test	Accuracy	F1	Precision	Recall	F1 − CV
2	2268	756	0.971	0.971	0.971	0.971	0.968
3	3402	1134	0.800	0.800	0.801	0.800	0.805
4	4464	1488	0.710	0.709	0.709	0.710	0.706
5	5580	1860	0.651	0.653	0.658	0.651	0.646
6	6696	2232	0.587	0.586	0.587	0.587	0.579
7	7812	2604	0.548	0.549	0.551	0.548	0.539
8	8928	2976	0.476	0.474	0.472	0.476	0.484

Table 7
Summary of regression analysis evaluation on the test set. The "Target" column specifies the specific physiology value for which the corresponding metrics are presented. The RMSE − CV column showcases the performance score attained through model optimization using cross-validation (CV) on the train set.

Target	R2	MAE	ME	RMSE	RMSE − CV	rNRMSE
E	0,379	0,962	4,671	1,270	1,342	0,136
ETR	0,499	41,626	212,662	55,100	53,733	0,160
gsw	0,313	0,044	0,251	0,061	0,067	0,149
PhiPS2	0,383	0,078	0,289	0,100	0,103	0,159
SPAD	0,570	0,310	1,394	0,399	0,443	0,129

Legend: Abbreviations: E – transpiration; ETR – electron transport rate, gsw – stomatal conductance, PhiPS2 – quantum efficiency in light, SPAD – leaf chlorophyll concentration.

Table 8
Binary classification scores on the test set for discerning between heathy and infected potato plants. The F1 − CV column showcases the performance score attained through model optimization using cross-validation (CV) of the train set.

Variety	Accuracy	bAccuracy	F1	Precision	Recall	F1 − CV
Alouette	0.625	0.630	0.631	0.656	0.625	
Carolus	0.875	0.900	0.877	0.906	0.875	
KIS Kokra	0.750	0.815	0.767	0.875	0.750	
KIS Tamar	0.875	0.890	0.877	0.906	0.875	
Levante	1.000	1.000	1.000	1.000	1.000	
Pooled	0.825	0.826	0.826	0.825	0.825	0.862

3.3. Evaluation of most important spectral features

Generally, the most informative feature comprised green, red edge and NIR channels in conjunction with other channels for regression of physiology parameters (Fig. 7). For instance, these channels are included in II. Automatically generated features for prediction of E and ETR and in combinations of III. Calculated spectral indices (i.e. GOSAVI, MCARI, and OSAVI) for prediction of gsw and PhiPS2. Moreover, the preeminence of the red edge-to-green channel ratio is evident for E, gsw, and ETR. The incorporation of the NIR and red edge channels with other features further contributed significant information. Specifically, the NIR channel, in conjunction with the red channel, emerged as the most important feature for PhiPS2 estimation (OSAVI). The blue channel appeared comparatively less influential, consistently ranking lower in relevance across all cases. Nevertheless, its significance was not discounted, as it holds the third position, for example, in estimating E. Similar trends were observed in SPAD. The NIR channel proved to be the most informative spectral region, contributing significantly to the top-ranked spectral features. These features primarily involved ratios with the red-edge channel and indices such as CVI. In contrast, the blue channel demonstrated smaller contribution, ranking fifth in importance.

For the detection of early blight (Fig. 8), the NIR channel alone yielded the highest information, akin to SPAD regression analysis. Moreover, the absolute difference in mean spectral reflectance for this band, along with the variance of the values in this channel, is the largest between the two classes. Furthermore, in addition to the NIR band, other spectral channels, including the green and blue channels, were recognized for their contribution to prediction efficacy. Calculated spectral indices (III. features category) contributed relatively less information, with the first spectral index, RI4XS ranking fourth in importance. None of the features from the II. features category were present among the top 10 features.

For variety discrimination, two spectral indices provided the most information: CIRE, and BCC (Fig. 8). The most notable difference between varieties was observed NIR. Automatically generated features (II. Features category), such as the combination of NIR and red edge channels, contributed supplementary information. No single raw spectral channel (I. Features category) was ranked among the top 10 features. The sequence of relevant features may exhibit bias due to the suboptimal classification when discerning all eight varieties, yielding a modest F1 score of 0.474. SHAP analysis results consistently demonstrated that more information is provided by the combination of features (II. And III. Features categories) compared to the utilization of raw reflectance values (I. Feature category), aligning with other research (Huang et al., 2016; Rahman et al., 2022).

4. Discussion

Symptomatic detection of early blight in potatoes using multispectral imaging was successful, with an overall F1 score exceeding 0.82. The separation of plant varieties was feasible. Accuracy varied among varieties, but remained acceptable, even though the 8 varieties included in this study were spectrally too similar for accurate variety identification. Predictions of plant physiological status was also possible using exclusively spectral data, with best achieved R^2 and rNRMSE of 0.57 and 0.129, respectively.

Our F1-based integration approach objectively identified and prioritized the inclusion of potato varieties with the most distinct spectral features. This closely mirrored the visual separation achieved using UMAP. Varieties with the most evident visual differences, KIS Tamar and KIS Blegoš, were integrated first and showed the highest classification accuracy (F1 score of 0.97). Similarly, KIS Otolia, the next variety integrated, remained well-separated visually from KIS Tamar (F1 0.86) and moderately from KIS Blegoš (F1 0.71). This pattern continued with subsequent variety additions. However, some varieties exhibited highly similar spectral responses, hindering differentiation and impacting classification metrics. The close alignment of spectral signatures between certain varieties makes separation nearly impossible (Fig. 8). This is confirmed by the low F1 score (below 0.7) and the visually indistinct separation boundaries in the UMAP projection. Nevertheless, UMAP's strength lies in its ability to reveal these challenges upfront. By visualizing the data before modeling, UMAP allows for a more informed exploration of the dataset, highlighting potential limitations in classification accuracy for specific variety combinations. Furthermore, UMAP revealed clear distinctions between different varieties, corroborated by significant differences found in field measurements of plant physiology parameters. These are closely related to stomatal characteristics, such as density, size, and degree of opening (Haworth et al., 2021). UMAP

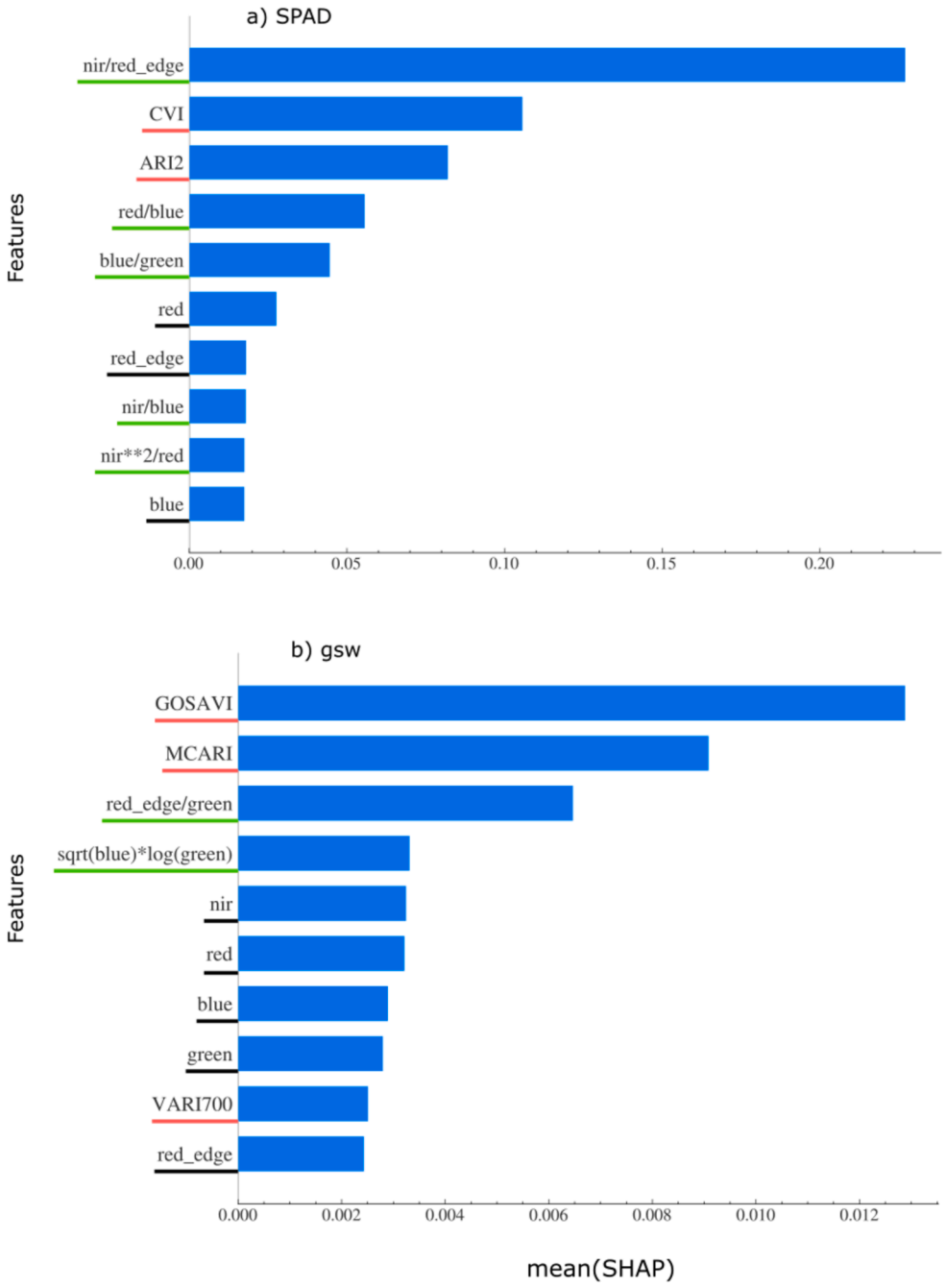


Fig. 7. SHAP bar plot of the most relevant features for regression analysis of: a) SPAD and b) gsw. Larger values indicate features that contribute more information. Different features are color-coded for easy identification: non-modified reflectance values (black), automatically generated features (green), and calculated spectral indices (red). Full results are provided in Supplementary material.

effectively distinguished variations in key plant processes related to photosynthesis, gas exchange, plant water status, and stress tolerance.

The inability to distinguish between certain plant varieties using multispectral imaging can be attributed to several interlinked factors. Plant varieties can have very similar genetic makeup, especially if they are from the same parental lines, leading to similar biochemical compositions and similar spectral responses (Lapajne et al., 2024; Maldia et al., 2023). In this study, varieties exhibited varying intensities of flowering, inflorescence, corolla sizes and coloration. They also differed in maturity, ranging from very early, determinate growth varieties to very late, indeterminate growth varieties. Consequently, these varieties can often be distinguished from one another based on their phenotypic traits, which are observable to the naked eye. However, in practice, this can be more challenging, as many varieties share similar traits.

The potato plant variety differentiation is significantly influenced by certain traits, with foliage structure and color being the most impactful. Varieties exhibiting different foliage structures, such as KIS Blegoš (leafy prostrate) and KIS Tamar (leafy semi-prostrate), achieve high classification scores (0.93). The differentiation between upright and semi-prostrate types, exemplified by Alouette (upright) and KIS Tamar (semi-prostrate), also shows significant classification accuracy (0.87). Growth habits (early, medium, late) further distinguish varieties, as seen with KIS Tamar (late) and KIS Blegoš (medium early) sharing a high differentiation score (0.93). Height, though less influential than foliage structure and color, helps in distinguishing tall varieties like Alouette and Otolia from medium and low-height varieties. Color enhances

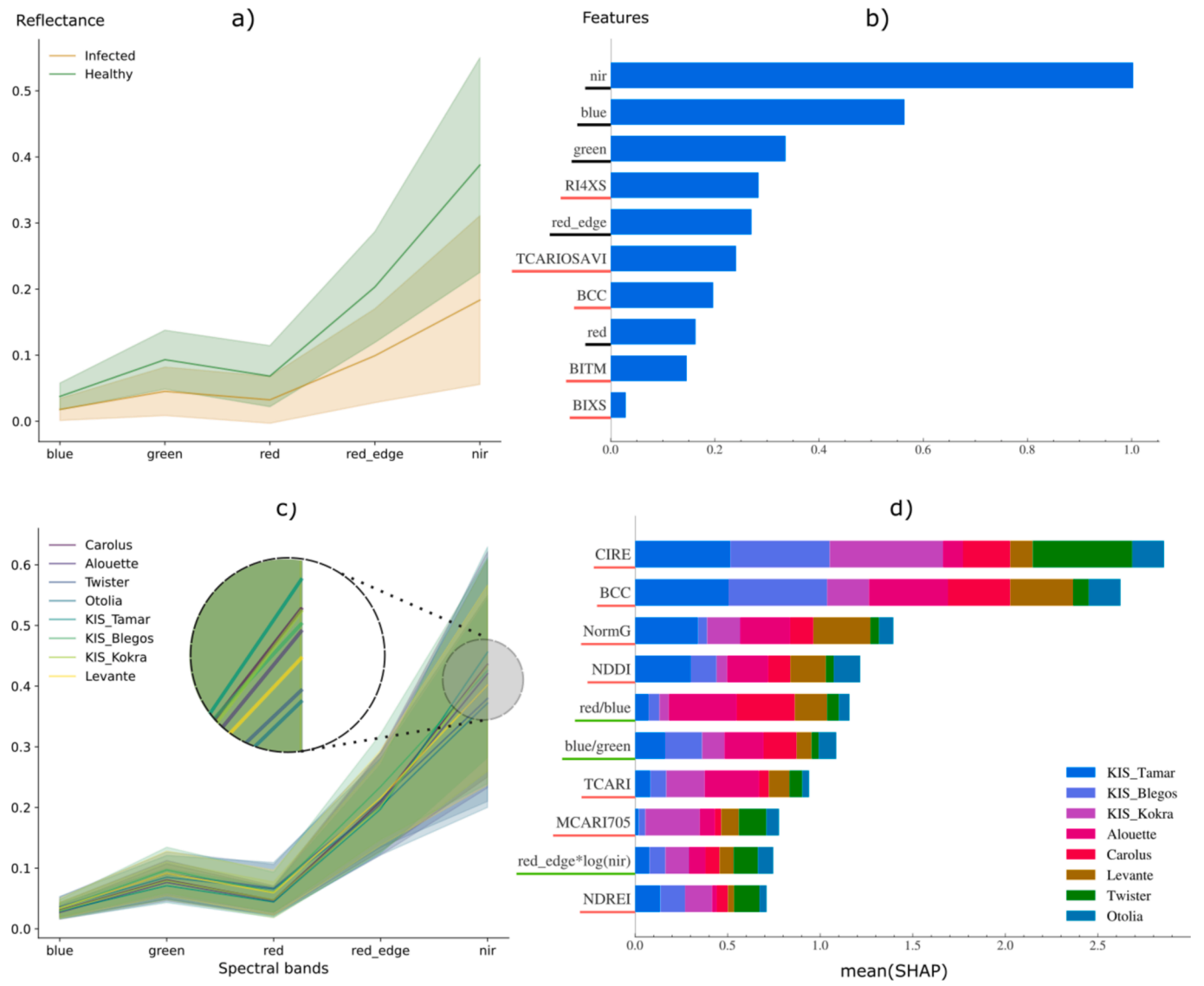


Fig. 8. Detection of early blight: a) Mean reflectance of either infestation condition; b) SHAP bar plot. Classification of varieties: c) Mean reflectance per variety; d) SHAP bar plot, where each bar's color corresponds to the contribution of the respective variety. In both cases, the reflectance's mean value and standard deviation are represented with solid line and ribbon, respectively. Larger values in SHAP bar plots indicate features that contribute more information. Different features are color-coded for easy identification: non-modified reflectance values (black), automatically generated features (green), and calculated spectral indices (red).

differentiation, particularly when combined with other traits; for example, Alouette's dark green foliage with purple pigmentation stands out against others. However, varieties with similar growth habits, heights, and colors, such as Carolus and Levante, exhibit lower differentiation scores (0.68). Combining foliage structure and color yields strong differentiation, as demonstrated by KIS Blegoš and KIS Tamar. Growth habit, in conjunction with foliage structure, also significantly influences classification. Pairs with the lowest scores, such as Alouette and Levante (0.61), suggest that despite distinct colors, similarities in other traits like growth habit and height diminish differentiation. In conclusion, the differentiation between potato plant varieties using spectral imaging appears to be influenced by factors such as foliage structure and color, with growth habit and height also potentially playing significant roles. However, further research and confirmation are needed to fully establish the reliability of these factors.

Potential differences in varieties' spectral signatures can be subtle, and multispectral imaging may not be sensitive enough to characterize them. This limitation can be overcome using hyperspectral imaging, capturing a broader range of spectral bands, including the short-wave infrared (1000——2500 nm) part of the spectrum, at narrower intervals (Feng et al., 2020). This provides more detailed information about the biochemical composition and structural characteristics of plants, allowing for better discrimination between similar varieties. Additional data sources can further improve variety classifications. Incorporating spatial information from point clouds or LiDAR, such as plant canopy structure and lodging strategies (Grebby et al., 2011; Han et al., 2022), and texture information extracted from images using gray-level co-occurrence matrices, can improve variety phenotyping and discrimination (Sun et al., 2023).

Prediction accuracy of physiological characteristics varies depending on the specific parameter being inferred. The metrics achieved in this study provide a comprehensive overview of model performance across different physiological parameters. R2 values range from 0.313 to 0.570, indicating that the models explain between 31.3 % and 57 % of the variability in the data, while rNRMSE values are consistently below 0.20, demonstrating relatively low normalized error. Since both R2 and rNRMSE are expressed on an absolute scale, they enable direct comparison across models estimating different physiological targets. In

contrast, MAE, RMSE, and ME are scale-dependent metrics, meaning their values are influenced by the units and scale of the target variables. This dependence limits their utility for direct cross-model comparisons but provides valuable insights into the magnitude of prediction errors within each specific context. MAE measures the average error magnitude and is less influenced by outliers, making it useful for assessing general model performance without skew from extreme deviations. RMSE, due to its quadratic nature, is more sensitive to large errors, highlighting significant deviations that could critically impact outcomes. ME identifies the largest individual deviation between predictions and actual values, assessing the model's worst-case performance when single large errors could have substantial consequences. For example, ETR exhibits the largest relative MAE (41.626) and RMSE (55.100), while gsw shows the smallest relative errors (MAE = 0.044, RMSE = 0.061). The evaluated ground measurements of SPAD aligned most closely with predictions, attaining an R^2 and rNRMSE of 0.57 and 0.129, respectively. Chlorophyll plays a crucial role in photosynthesis by absorbing sunlight and converting it into energy and is therefore directly correlated with the greenness of the plant (Björn et al., 2009).

The estimation of physiological variables in all cases was not entirely satisfactory, as a significant portion of variance was not covered by the model. This suggests that the five spectral bands were insufficient for accurately predicting physiological parameters. Further improvements can be achieved by incorporating more spectral bands, possibly by utilization of hyperspectral imaging, and by integrating combinations of additional technology such as LiDAR (Wang et al., 2022).

Linear mixed model analysis suggests that both variety and treatment have a significant effect on physiological parameters of plants, with specific varieties having different impacts. Treatment (organicand conventional production) had a significant effect only for variables gsw and E, where the response in the conventional setting was lower than inorganic. Stomatal conductance and transpiration rate are closely related because the opening and closing of stomata control both processes, and higher stomatal conductance typically leads to higher transpiration rates (Kirschbaum and McMillan, 2018). Organic farming practices often enhance soil health, including enhanced microbial activity, which can improve resource availability and plant health. Furthermore, plants in organic farming often develop deeper and more extensive root systems, which improves water uptake and supports higher transpiration rates. These factors work together and support more efficient gas exchange and water use in potato plants (Lawson et al., 2014; Martin-StPaul et al., 2017).

Analysis of random effects revealed that the impact of Block was relatively minor, indicating that it does not contribute significantly to additional variation. The timing of observations had a more pronounced effect, suggesting the importance of accounting for temporal variability. Regardless, the comparatively considerable amount of unexplained variability indicates there are other important factors influencing the response variable, or that there is considerable inherent randomness in the data. Several of the possible additional effects, such heterogeneity of resource distribution, are nested within the random effect Block, which was shown to have a limited effect. Nevertheless, further research and the inclusion of additional measured effects are necessary to better explore these relationships and potentially enhance the explanatory power of LMMs.

Identification of early blight in potatoes achieved a high success rate. Detection accuracies varied between varieties, with Levante achieving an almost perfect score (~100 %), while with Alouette accuracies were acceptable (>60 %). Even though spectral similarities between varieties hindered accurate detection of late blight using multispectral imaging, overall accuracies were acceptable. All varieties included are resistant to certain diseases, such as late blight, while their susceptibility to early blight varies. Exact disease resistance scores for early blight are not available, but according to potato variety evaluation data in Slovenia, Levante, Otolia, KIS Blegoš and Twister are among the more sensitive varieties (Agricultural institute of Slovenia, n.d.). Resistance to early blight cannot be deduced from the overall resistance of varieties to other diseases. Plant disease resistance is pathogen-specific; the genetic, physiological, and biochemical pathways that confer resistance to early blight are distinct from those that protect plants against other pathogens. Similarly, manifestation of symptoms, including spectral, can be highly specific for each host-pathogen combination. This specificity is essential for developing robust disease-resistant cultivars and for the precise application of integrated pest management.

Visible symptoms of early blight include leaf lesions, yellowing and browning on the margins of leaves. Symptoms start developing in the lower parts of plants and move upwards, spreading throughout the canopy. In resistant varieties, symptoms develop slower and are less pronounced, making detection more challenging and less accurate. Visual examinations data can partially overcome this limitation, detecting changes in pigment structure before these become visible or apparent. Additionally, hyperspectral imaging can provide more information and enable pre-symptomatic detection. In this study, multispectral data was acquired at only one time point, leading to a limited data set and reduced accuracy. Extraction of leaf-area pixels was performed using a 10 px circle, without within-circle segmentation, as this would require better spatial resolutions. Each plant was considered as a whole, i.e. mean reflectance was calculated from all 10 pixels within each circle, leading to reduced signal-to-noise ratios. The number of mixed signals depends on varieties' plant structure, information that can be gleaned from point clouds or using LiDAR. Of the early maincrop varieties, Alouette has the fastest initial foliage development, and good final development, while both Carolus and Levante have slower initial canopy development. Young leaves are generally resistant to early blight, and susceptibility increases with leaf senescence. However, rapid canopy development can shade lower, and therefore older, leaves, thus reducing disease severity. Several other factors influence the resistance of individual varieties to early blight, from expression of resistance genes and biochemical defenses, such as phytoalexins, to leaf morphology, photosynthetic activity, and environmental conditions (Xue et al., 2019). While these differences are not the main cause for the observed differences in early blight detection, they are one of the confounding factors, requiring additional research to fully explore their effects on disease detection.

Red-green–blue (RGB), red-edge and near-infrared bands (and their combinations) are commonly used in crop phenology determination (Guo et al., 2021). Red-edge is useful for its correlation with nitrogen abundance, water content, chlorophyll concentration and structural features due to its chlorophyll absorption and leaf internal scattering (Hennessy et al., 2020; Schlemmer et al., 2005). Pigments can be detected in visible spectrum (Beamish et al., 2018). The green part of the spectrum is less absorbed by chlorophyll *a* and b compared to blue and red. This makes features utilizing green reflectance, along with other bands, effective for detecting variations in chlorophyll content in leaves (Chen et al., 2018), which was also shown in this study. Previous studies have also shown that the red-edge, red, and green spectral bands, both individually and in combination, play an essential role in estimating crop transpiration coefficients (Shao et al., 2022). The red band is especially important for distinguishing among different plant varieties due to its relationship with chlorophyll (Hennessy et al., 2020). Previous studies have shown that detection of disease severity on potato plants can be performed solely from visible part of the spectrum (Gibson-Poole et al., 2017; Siebring et al., 2019). Study from Van De Vijver et. al. (Van De Vijver et al., 2020) showed that 750 nm, 550 nm and 680 nm wavelengths, corresponding to the NIR, red, and red edge bands, respectively, were significant for detection of early blight. Detection of this disease is possible because infection leads to changes in leaf pigment structure and cell structure, alternating spectral reflectance which can be detected by multispectral cameras (Van De Vijver et al., 2020). Atherton et. al. ("Remote Sensed Spectral Imagery to Detect Late Blight in Field Tomatoes," 2015) concluded that due to the effect this disease

has on chlorophyll, which has high energy absorption in red and blue bands, these parts of spectrum are important for early blight detection.

The ease of automatic feature ranking and interpretation underscores the efficacy of using SHAP for such problems. This provides deeper insights into model decisions (i.e. model transparency) and facilitates improved modeling and more informed decisions. Based on this foundation, the hybrid engineering strategy utilized in this study, proved to be beneficial for modeling of the analyzed problems. The importance of both II. automatically generated features and III. calculated spectral indices were relevant for the estimation of physiological parameters, as most of the top 10 features presented a combination of multiple spectral bands. Both strategies were found to be uniformly relevant among the top 10 features, suggesting they hold equal importance. However, for both classification problems – variety classification and detection of early blight – the II. automatically generated features were found to be less informative. Established spectral indices, refined and validated through extensive research and practical application, are specifically designed to enhance the spectral characteristics associated with vegetation. These indices are based on well-established biophysical principles, making them more directly interpretable and relevant for assessing plant health compared to generic statistical features, which may not capture these specific spectral relationships as effectively and lead to non-informative features (Fei et al., 2022; Zafari et al., 2019). This implies that the majority of variance could be covered by solely employing vegetation indices. Interestingly, for the classification of early blight, this was the only instance where non-modified reflectance values were found to be the most relevant. This suggests that diseased and healthy plants could be already well distinguished based on reflectance spectral bands. This is likely due to the developed visible symptoms (plant health assessment was performed visually) and hence a clear distinction in spectral signatures, where the values for infected plants were lower (Fig. 8). While this approach enables symptomatic detection, disease severeness and pre-symptomatic detection would require integration of additional data, such as hyperspectral imaging and LiDAR.

This study's findings bear implications for improvement of precision agriculture practices. The capability to precisely monitor various stress factors or characteristics of potato plants could encourage efficient farming practices, potentially improving crop yield and promoting sustainability. Furthermore, the results suggest the feasibility of non-invasive identification of early blight by using multi-spectral imaging and UAVs. This could enable the mitigation of potential losses, further underscoring the value of these technologies in modern agriculture.

Despite the promising results of the study, several noteworthy factors should be considered. The model was built considering only spectral features, either direct spectral values or their combinations. Other features, such as textural, structural, or statistical ones, were not included in the study, even though they could potentially enhance the model's performance in some cases (Sun et al., 2023). However, the focus of this study was primarily on the hybrid approach, which generated features based on extracted spectral signatures. Additionally, it's worth noting that the background was not removed before extracting the spectral signatures of individual potato plants, as done in some other studies (Rodríguez et al., 2021; Yang et al., 2022). This could potentially impact spectral signatures if an arbitrary pixel covered too much soil. However, this was mitigated by carefully studying the pinpointed locations of plants and ensuring that, in most cases, only the potato extents were extracted. Moreover, achieving excellent results in some cases does not necessarily imply perfect modeling in sense of causality. Changes in spectral features could be influenced by other factors such as water or nutrient deficiency stress, among others (Lapajne et al., 2024). The performance and generalizability of the model may also be affected by small datasets, limited number of varieties, and limited spatial and temporal distributions; common issues reported in many studies (Jin et al., 2022; Lapajne et al., 2022; Martens and Dardenne, 1998). The need to adapt data acquisition due to unfavorable weather conditions introduced an additional source of uncertainty in the study. Since data collection was performed whenever weather permitted, rather than at predefined intervals, this variation may have affected the consistency of the observations. Furthermore, the inclusion of multiple potato plant varieties with varying growth rates—some being early-maturing and others late—made it difficult to establish uniform data collection dates.

5. Conclusions

This research aimed to integrate modern modeling techniques to decode intricate relationships among features derived from UAV-acquired multispectral imaging of potato plants. It addressed multiple challenges in potato research, employing the XGBoost model and leveraging explainable machine learning tools (UMAP and SHAP) to investigate the intrinsic nature of the analyzed problems. The study employs a novel approach that auto-generates features by using simple mathematical formulations and integrates these with a wide array of calculated vegetation indices. Additionally, the paper presents a modern pipeline for processing multispectral images, offering a solution that can be extended or applied to other research domains. The results indicate the adequacy of the utilized methodology, as the analyzed problems yielded satisfactory and anticipated evaluation metrics. For all research questions, the most significant spectral bands and extracted features were evaluated to determine which channels provided the most information for successful modeling. In certain cases, such as the estimation of physiological parameters, the achieved metrics were moderate, suggesting the potential benefit of incorporating additional features, possibly through hyperspectral imaging and integration of LiDAR. This could enhance the results by integrating necessary additional information. To summarize, the paper offers numerous benefits: it promotes more interpretable machine learning techniques, encourages feature integration through automatic mathematical formulations, deepens knowledge in potato plant research, and fosters further research by making the entire codebase and datasets publicly available. These contributions could lead to more sustainable farming practices, and a deeper understanding of plant physiology, crucial for the future of agriculture.

6. Code and data availability

In line with our commitment to fostering open-science practices, we have made the data and code readily accessible. Interested researchers will be able to access the pre-processed dataset at https://doi.org/10.5281/zenodo.10934163 (accessed on 8 October 2024) and code at https://github.com/janezlapajne/manuscripts (accessed on 8 October 2024). Additionally, the optimization and feature extraction parts were separated and implemented in the Python library at https://github.com/siapy/siapy-lib (accessed on 15 October 2024). By providing transparent and unrestricted access to these resources, we aim to encourage collaborative replication, validation, and further exploration of the findings presented in this paper. The accompanying documentation ensures that scientists can seamlessly engage with the materials, promoting transparency and reproducibility.

CRediT authorship contribution statement

Janez Lapajne: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Andrej Vončina:** Visualization, Resources, Methodology, Investigation, Data curation. **Ana Vojnović:** Data curation. **Daša Donša:** Writing – original draft. **Peter Dolničar:** Writing – review & editing, Validation, Project administration, Methodology, Funding acquisition, Conceptualization. **Uroš Žibrat:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

This research was supported by funds from the Slovenian research and innovation agency (ARIS) (MR 54720, P4-0072), and Horizon Europe project ECOBREED (grant agreement ID 771367). Some of the research was conducted using equipment financed by EU-FP7 project CropSustain (FP7-REGPOT-CT2012-316205), and research infrastructure ELIXIR-SI (https://elixir-slovenia.org), funded by the European Regional Development Fund, the Slovenian Ministry of Education, Science and Sports, and by the Slovenian Research and Innovation Agency.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Janez Lapajne reports financial support was provided by Public Research Agency of the Republic of Slovenia.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compag.2024.109746.

Data availability

Data will be made available on request.

References

Abbas, A., Zhang, Z., Zheng, H., Alami, M.M., Alrefaei, A.F., Abbas, Q., Naqvi, S.A.H., Rao, M.J., Mosa, W.F.A., Abbas, Q., Hussain, A., Hassan, M.Z., Zhou, L., 2023. Drones in plant disease assessment, efficient monitoring, and detection: a way forward to smart agriculture. Agronomy 13, 1524. https://doi.org/10.3390/agronomy13061524.

Abdelbaki, A., Schlerf, M., Retzlaff, R., Machwitz, M., Verrelst, J., Udelhoven, T., 2021. Comparison of crop trait retrieval strategies using UAV-based VNIR hyperspectral imaging. Remote Sens. (Basel) 13, 1748. https://doi.org/10.3390/rs13091748.

Agricultural institute of Slovenia, n.d. Unpublished data, Yearly variety screening and testing results 2024.

Akiba, T., Sano, S., Yanase, T., Ohta, T., Koyama, M., 2019. Optuna: A Next-generation Hyperparameter Optimization Framework. 10.48550/ARXIV.1907.10902.

Akoglu, H., 2018. User's guide to correlation coefficients. Turk. J. of Emerg. Med. 18, 91–93. https://doi.org/10.1016/j.tjem.2018.08.001.

Apicella, A., Isgrò, F., Prevete, R., 2024. Don't Push the Button! Exploring Data Leakage Risks in Machine Learning and Transfer Learning. 10.48550/ARXIV.2401.13796.

Arevalo-Ramirez, T., Guevara, J., Rivera, R.G., Villacres, J., Menendez, O., Fuentes, A., Auat Cheein, F., 2022. Assessment of multispectral vegetation features for digital terrain modeling in forested regions. IEEE Trans. Geosci. Remote Sens. 60, 1–9. https://doi.org/10.1109/TGRS.2021.3109601.

Armstrong, G., Martino, C., Rahman, G., Gonzalez, A., Vázquez-Baeza, Y., Mishne, G., Knight, R., 2021. Uniform Manifold Approximation and Projection (UMAP) Reveals Composite Patterns and Resolves Visualization Artifacts in Microbiome Data. mSystems 6, e00691-21. 10.1128/mSystems.00691-21.

Attia, A., Govind, A., Qureshi, A.S., Feike, T., Rizk, M.S., Shabana, M.M.A., Kheir, A.M.S., 2022. Coupling process-based models and machine learning algorithms for predicting yield and evapotranspiration of maize in arid environments. Water 14, 3647. https://doi.org/10.3390/w14223647.

Bates, S., Hastie, T., Tibshirani, R., 2021. Cross-validation: what does it estimate and how well does it do it? 10.48550/ARXIV.2104.00673.

Beamish, A.L., Coops, N.C., Hermosilla, T., Chabrillat, S., Heim, B., 2018. Monitoring pigment-driven vegetation changes in a low-Arctic tundra ecosystem using digital cameras. Ecosphere 9, e02123. https://doi.org/10.1002/ecs2.2123.

Bergstra, J., Bardenet, R., Bengio, Y., Kégl, B., 2011. Algorithms for hyper-parameter optimization. NIPS'11: Proceedings of the 24th International Conference on Neural Information Processing Systems 2546–2554.

Björn, L.O., Papageorgiou, G.C., Blankenship, R.E., Govindjee, 2009. A viewpoint: Why chlorophyll a? Photosynth Res 99, 85–98. 10.1007/s11120-008-9395-x.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2011. SMOTE: Synthetic Minority Over-sampling Technique. 10.48550/ARXIV.1106.1813.

Chen, T., Guestrin, C., 2016. XGBoost: A Scalable Tree Boosting System. 10.48550/ARXIV.1603.02754.

Chen, A., Orlov-Levin, V., Meron, M., 2018. Applying High-Resolution Visible-Channel Aerial Scan of Crop Canopy to Precision Irrigation Management, in: The 2nd International Electronic Conference on Remote Sensing. Presented at the International Electronic Conference on Remote Sensing, MDPI, p. 335. 10.3390/ecrs-2-05148.

Cristache, S.-E., Vuță, M., Marin, E., Cioacă, S.-I., Vuță, M., 2018. Organic versus conventional farming—a paradigm for the sustainable development of the European countries. Sustainability 10, 4279. https://doi.org/10.3390/su10114279.

D S, D.S., N, M., S, S., D, S., R, J., G, M., 2023. Influence of quantity, quality, horizontal and vertical distribution of ground control points on the positional accuracy of UAV survey. Appl Geomat 15, 897–917. 10.1007/s12518-023-00531-w.

DeFauw, S.L., He, Z., Larkin, R.P., Mansour, S.A., 2012. Sustainable Potato Production and Global Food Security, in: He, Z., Larkin, R., Honeycutt, W. (Eds.), Sustainable Potato Production: Global Case Studies. Springer Netherlands, Dordrecht, pp. 3–19. 10.1007/978-94-007-4104-1_1.

Dolničar, P., Maras, M., Vladimir, M., Metka, Ž., Urek, G., Širca, S., Simončič, A., 2017. KIS Slavnik and KIS Savinja - new potato varieties bred at Agricultural Institute of Slovenia. Slovenian Soc. Agron. 2017, 58–65.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J., Münkemüller, T., McClean, C., Osborne, P.E., Reineking, B., Schröder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography 36, 27–46. https://doi.org/10.1111/j.1600-0587.2012.07348.x.

Duarte, H.S.S., Zambolim, L., Capucho, A.S., Júnior, A.F.N., Rosado, A.W.C., Cardoso, C. R., Paul, P.A., Mizubuti, E.S.G., 2013. Development and validation of a set of standard area diagrams to estimate severity of potato early blight. Eur. J. Plant Pathol. 137, 249–257. https://doi.org/10.1007/s10658-013-0234-3.

Fei, H., Fan, Z., Wang, C., Zhang, N., Wang, T., Chen, R., Bai, T., 2022. Cotton classification method at the county scale based on multi-features and random forest feature selection algorithm and classifier. Remote Sens. (Basel) 14, 829. https://doi.org/10.3390/rs14040829.

Feng, X., He, L., Cheng, Q., Long, X., Yuan, Y., 2020. Hyperspectral and multispectral remote sensing image fusion based on endmember spatial information. Remote Sens. (Basel) 12, 1009. https://doi.org/10.3390/rs12061009.

Fernández-Marín, B., García-Plazaola, J.I., Hernández, A., Esteban, R., 2018. Plant Photosynthetic Pigments: Methods and Tricks for Correct Quantification and Identification, in: Sánchez-Moreiras, A.M., Reigosa, M.J. (Eds.), Advances in Plant Ecophysiology Techniques. Springer International Publishing, Cham, pp. 29–50. 10.1007/978-3-319-93233-0_3.

Gałecki, A., Burzykowski, T., 2013. Linear Mixed-Effects Models Using R: A Step-by-Step Approach, Springer Texts in Statistics. Springer New York, New York, NY. 10.1007/978-1-4614-3900-4.

Geissen, V., Silva, V., Lwanga, E.H., Beriot, N., Oostindie, K., Bin, Z., Pyne, E., Busink, S., Zomer, P., Mol, H., Ritsema, C.J., 2021. Cocktails of pesticide residues in conventional and organic farming systems in Europe – legacy of the past and turning point for the future. Environ. Pollut. 278, 116827. https://doi.org/10.1016/j.envpol.2021.116827.

Gibson-Poole, S., Humphris, S., Toth, I., Hamilton, A., 2017. Identification of the onset of disease within a potato crop using a UAV equipped with un-modified and modified commercial off-the-shelf digital cameras. Adv. Anim. Biosci. 8, 812–816. https://doi.org/10.1017/S204047001700084X.

Grebby, S., Naden, J., Cunningham, D., Tansey, K., 2011. Integrating airborne multispectral imagery and airborne LiDAR data for enhanced lithological mapping in vegetated terrain. Remote Sens. Environ. 115, 214–226. https://doi.org/10.1016/j.rse.2010.08.019.

Guo, W., Carroll, M.E., Singh, A., Swetnam, T.L., Merchant, N., Sarkar, S., Singh, A.K., Ganapathysubramanian, B., 2021. UAS-Based Plant Phenotyping for Research and Breeding Applications. Plant Phenomics 2021, 2021/9840192. 10.34133/2021/9840192.

Han, L., Yang, G., Yang, X., Song, X., Xu, B., Li, Z., Wu, J., Yang, H., Wu, J., 2022. An explainable XGBoost model improved by SMOTE-ENN technique for maize lodging detection based on multi-source unmanned aerial vehicle images. Comput. Electron. Agric. 194, 106804. https://doi.org/10.1016/j.compag.2022.106804.

Haworth, M., Marino, G., Loreto, F., Centritto, M., 2021. Integrating stomatal physiology and morphology: evolution of stomatal control and development of future crops. Oecologia 197, 867–883. https://doi.org/10.1007/s00442-021-04857-3.

Heaton, J., 2017. An Empirical Analysis of Feature Engineering for Predictive Modeling. 10.48550/ARXIV.1701.07852.

Hennessy, A., Clarke, K., Lewis, M., 2020. Hyperspectral classification of plants: A review of waveband selection generalisability. Remote Sens. (Basel) 12, 113. https://doi.org/10.3390/rs12010113.

Horn, F., Pack, R., Rieger, M., 2019. The autofeat Python Library for Automated Feature Engineering and Selection. 10.48550/ARXIV.1901.07329.

Huang, M., He, C., Zhu, Q., Qin, J., 2016. Maize seed variety classification using the integration of spectral and image features combined with feature transformation based on hyperspectral imaging. Appl. Sci. 6, 183. https://doi.org/10.3390/app6060183.

Jin, J., Yin, F., Xu, Y., Zhang, J., 2022. Learning a Model with the Most Generality for Small-Sample Problems, in: Proceedings of the 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence. Presented at the ACAI 2022: 2022 5th International Conference on Algorithms, Computing and Artificial Intelligence, ACM, Sanya China, pp. 1–6. 10.1145/3579731.3579814.

Jones, E.J., Bishop, T.F.A., Malone, B.P., Hulme, P.J., Whelan, B.M., Filippi, P., 2022. Identifying causes of crop yield variability with interpretive machine learning. Comput. Electron. Agric. 192, 106632. https://doi.org/10.1016/j.compag.2021.106632.

Kirschbaum, M.U.F., McMillan, A.M.S., 2018. Warming and elevated CO2 have opposing influences on transpiration. which is more important? Curr. Forestry Rep. 4, 51–71. https://doi.org/10.1007/s40725-018-0073-8.

Lammerts Van Bueren, E.T., Jones, S.S., Tamm, L., Murphy, K.M., Myers, J.R., Leifert, C., Messmer, M.M., 2011. The need to breed crop varieties suitable for organic farming, using wheat, tomato and broccoli as examples: a review. NJAS: Wageningen J. Life Sci. 58, 193–205. https://doi.org/10.1016/j.njas.2010.04.001.

Lapajne, J., Knapič, M., Žibrat, U., 2022. Comparison of selected dimensionality reduction methods for detection of root-knot nematode infestations in potato tubers using hyperspectral imaging. Sensors 22, 367. https://doi.org/10.3390/s22010367.

Lapajne, J., Vojnović, A., Vončina, A., Žibrat, U., 2024. Enhancing water-deficient potato plant identification: assessing realistic performance of attention-based deep neural networks and hyperspectral imaging for agricultural applications. Plants 13, 1918. https://doi.org/10.3390/plants13141918.

Lawson, T., Simkin, A.J., Kelly, G., Granot, D., 2014. Mesophyll photosynthesis and guard cell metabolism impacts on stomatal behaviour. New Phytol. 203, 1064–1081. https://doi.org/10.1111/nph.12945.

León-Rueda, W.A., León, C., Caro, S.G.-, Ramírez-Gil, J.G., 2022. Identification of diseases and physiological disorders in potato via multispectral drone imagery using machine learning tools. Trop. plant pathol. 47, 152–167. 10.1007/s40858-021-00460-2.

Li, D., Miao, Y., Gupta, S.K., Rosen, C.J., Yuan, F., Wang, C., Wang, L., Huang, Y., 2021. Improving potato yield prediction by combining cultivar information and UAV remote sensing data using machine learning. Remote Sens. (Basel) 13, 3322. https://doi.org/10.3390/rs13163322.

Li, Y., Zeng, H., Zhang, M., Wu, B., Zhao, Y., Yao, X., Cheng, T., Qin, X., Wu, F., 2023. A county-level soybean yield prediction framework coupled with XGBoost and multidimensional feature engineering. Int. J. Appl. Earth Obs. Geoinf. 118, 103269. https://doi.org/10.1016/j.jag.2023.103269.

Liu, Y., Feng, H., Yue, J., Li, Z., Jin, X., Fan, Y., Feng, Z., Yang, G., 2022a. Estimation of aboveground biomass of potatoes based on characteristic variables extracted from UAV hyperspectral imagery. Remote Sens. (Basel) 14, 5121. https://doi.org/10.3390/rs14205121.

Liu, Y., Feng, H., Yue, J., Li, Z., Yang, G., Song, X., Yang, X., Zhao, Y., 2022b. Remote-sensing estimation of potato above-ground biomass based on spectral and spatial features extracted from high-definition digital camera images. Comput. Electron. Agric. 198, 107089. https://doi.org/10.1016/j.compag.2022.107089.

Lizarazo, I., Rodriguez, J.L., Cristancho, O., Olaya, F., Duarte, M., Prieto, F., 2023. Identification of symptoms related to potato Verticillium wilt from UAV-based multispectral imagery using an ensemble of gradient boosting machines. Smart Agric. Technol. 3, 100138. https://doi.org/10.1016/j.atech.2022.100138.

Lu, B., Dao, P., Liu, J., He, Y., Shang, J., 2020. Recent advances of hyperspectral imaging technology and applications in agriculture. Remote Sens. (Basel) 12, 2659. https://doi.org/10.3390/rs12162659.

Lundberg, S., Lee, S.-I., 2017. A Unified Approach to Interpreting Model Predictions. 10.48550/ARXIV.1705.07874.

Maldia, L.S.J., Combalicer, M.S., Tinio, C.E., 2023. Plants' Anatomical and Genetic Responses to Anthropogenic Climate Change and Human-Induced Activities, in: Ramamoorthy, S., Buot, I.E., Rajasekaran, C. (Eds.), Plant Diversity in Biocultural Landscapes. Springer Nature Singapore, Singapore, pp. 403–441. 10.1007/978-981-19-8649-9_18.

Martens, H.A., Dardenne, P., 1998. Validation and verification of regression in small data sets. Chemom. Intel. Lab. Syst. 44, 99–121. https://doi.org/10.1016/S0169-7439(98)00167-1.

Martin-StPaul, N., Delzon, S., Cochard, H., 2017. Plant resistance to drought depends on timely stomatal closure. Ecol. Lett. 20, 1437–1447. https://doi.org/10.1111/ele.12851.

McInnes, L., Healy, J., Melville, J., 2018. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. 10.48550/ARXIV.1802.03426.

Montero, D., Aybar, C., Mahecha, M.D., Martinuzzi, F., Söchting, M., Wieneke, S., 2023. A standardized catalogue of spectral indices to advance the use of remote sensing in Earth system research. Sci. Data 10, 197. https://doi.org/10.1038/s41597-023-02096-0.

Nohara, Y., Matsumoto, K., Soejima, H., Nakashima, N., 2022. Explanation of machine learning models using shapley additive explanation and application for real data in hospital. Comput. Methods Programs Biomed. 214, 106584. https://doi.org/10.1016/j.cmpb.2021.106584.

Offermann, F., Nieberg, H., 2002. Does organic farming have a future in Europe? EuroChoices 1, 12–17. https://doi.org/10.1111/j.1746-692X.2002.tb00082.x.

Community Plant Variety Office, 2024. CPVO Technical Protocols [WWW Document]. https://cpvo.europa.eu/en/applications-and-examinations/technical-examinations/technical-protocols/cpvo-technical-protocols. URL (accessed 7.26.24).

Pacifico, D., Paris, R., 2016. Effect of organic potato farming on human and environmental health and benefits from new plant breeding techniques. is it only a matter of public acceptance? Sustainability 8, 1054. https://doi.org/10.3390/su8101054.

Pasala, R., Pandey, B.B., 2020. Plant phenomics: high-throughput technology for accelerating genomics. J Biosci 45, 111. https://doi.org/10.1007/s12038-020-00083-w.

Pepe, M., Fregonese, L., Scaioni, M., 2018. Planning airborne photogrammetry and remote-sensing missions with modern platforms and sensors. Eur. J. Remote Sens. 51, 412–436. https://doi.org/10.1080/22797254.2018.1444945.

Pix4Dmapper, 2024.

Pudil, P., Novovičová, J., Kittler, J., 1994. Floating search methods in feature selection. Pattern Recogn. Lett. 15, 1119–1125. https://doi.org/10.1016/0167-8655(94)90127-9.

QGIS Development Team, 2024. QGIS Geographic Information System.

Rahman, A.S.B.A., Sebastian, P., Izhar, L.I., 2022. Potato Crop Health Assessment Using Multispectral Image Analysis, in: 2022 International Conference on Future Trends in Smart Communities (ICFTSC). Presented at the 2022 International Conference on Future Trends in Smart Communities (ICFTSC), IEEE, Kuching, Sarawak, Malaysia, pp. 151–157. 10.1109/ICFTSC57269.2022.10039849.

Reganold, J.P., Wachter, J.M., 2016. Organic agriculture in the twenty-first century. Nat. Plants 2, 15221. https://doi.org/10.1038/nplants.2015.221.

Remote Sensed Spectral Imagery to Detect Late Blight in Field Tomatoes, 2015. , in: 2015 ASABE International Meeting. Presented at the 2015 ASABE International Meeting, American Society of Agricultural and Biological Engineers. 10.13031/aim.20152186976.

Rodríguez, J., Lizarazo, I., Prieto, F., Angulo-Morales, V., 2021. Assessment of potato late blight from UAV-based multispectral imagery. Comput. Electron. Agric. 184, 106061. https://doi.org/10.1016/j.compag.2021.106061.

Rodríguez-Pérez, R., Bajorath, J., 2020. Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. J Comput. Aided Mol. Des. 34, 1013–1026. https://doi.org/10.1007/s10822-020-00314-0.

Röös, E., Mie, A., Wivstad, M., Salomon, E., Johansson, B., Gunnarsson, S., Wallenbeck, A., Hoffmann, R., Nilsson, U., Sundberg, C., Watson, C.A., 2018. Risks and opportunities of increasing yields in organic farming. a review. Agron. Sustain. Dev. 38, 14. https://doi.org/10.1007/s13593-018-0489-3.

Rozentsvet, O., Bogdanova, E., Nesterov, V., Bakunov, A., Milekhin, A., Rubtsov, S., Rozentsvet, V., 2024. Phenotyping of potato plants using morphological and physiological tools. Plants 13, 647. https://doi.org/10.3390/plants13050647.

Sakia, R.M., 1992. The Box-Cox transformation technique: a review. The Statistician 41, 169. https://doi.org/10.2307/2348250.

Schlemmer, M.R., Francis, D.D., Shanahan, J.F., Schepers, J.S., 2005. Remotely measuring chlorophyll content in corn leaves with differing nitrogen levels and relative water content. Agron. J. 97, 106–112. https://doi.org/10.2134/agronj2005.0106.

Sethy, P.K., Pandey, C., Sahu, Y.K., Behera, S.K., 2022. Hyperspectral imagery applications for precision agriculture - a systemic survey. Multimed. Tools Appl. 81, 3005–3038. https://doi.org/10.1007/s11042-021-11729-8.

Shakoor, N., Lee, S., Mockler, T.C., 2017. High throughput phenotyping to accelerate crop breeding and monitoring of diseases in the field. Curr. Opin. Plant Biol. 38, 184–192. https://doi.org/10.1016/j.pbi.2017.05.006.

Shao, G., Han, W., Zhang, H., Wang, Y., Zhang, L., Niu, Y., Zhang, Y., Cao, P., 2022. Estimation of transpiration coefficient and aboveground biomass in maize using time-series UAV multispectral imagery. Crop J. 10, 1376–1385. https://doi.org/10.1016/j.cj.2022.08.001.

Shin, H., Oh, S., 2021. Feature-weighted sampling for proper evaluation of classification models. Appl. Sci. 11, 2039. https://doi.org/10.3390/app11052039.

Shwartz-Ziv, R., Armon, A., 2021. Tabular Data: Deep Learning is Not All You Need. 10.48550/ARXIV.2106.03253.

Siebring, J., Valente, J., Domingues Franceschini, M.H., Kamp, J., Kooistra, L., 2019. Object-based image analysis applied to low altitude aerial imagery for potato plant trait retrieval and pathogen detection. Sensors 19, 5477. https://doi.org/10.3390/s19245477.

Sujatha, B., 2015. Photosynthesis, in: Bahadur, B., Venkat Rajam, M., Sahijram, L., Krishnamurthy, K.V. (Eds.), Plant Biology and Biotechnology. Springer India, New Delhi, pp. 569–591. 10.1007/978-81-322-2286-6_22.

Sun, H., Song, X., Guo, W., Guo, M., Mao, Y., Yang, G., Feng, H., Zhang, J., Feng, Z., Wang, J., Ma, Y., Zheng, C., Li, P., Pan, D., 2023. Potato late blight severity monitoring based on the relief-mRmR algorithm with dual-drone cooperation. Comput. Electron. Agric. 215, 108438. https://doi.org/10.1016/j.compag.2023.108438.

Van De Vijver, R., Mertens, K., Heungens, K., Somers, B., Nuyttens, D., Borra-Serrano, I., Lootens, P., Roldán-Ruiz, I., Vangeyte, J., Saeys, W., 2020. In-field detection of Alternaria solani in potato crops using hyperspectral imaging. Comput. Electron. Agric. 168, 105106. https://doi.org/10.1016/j.compag.2019.105106.

Verdonck, T., Baesens, B., Óskarsdóttir, M., Vanden Broucke, S., 2024. Special issue on feature engineering editorial. Mach. Learn. 113, 3917–3928. https://doi.org/10.1007/s10994-021-06042-2.

Wang, Q., Che, Y., Shao, K., Zhu, J., Wang, R., Sui, Y., Guo, Y., Li, B., Meng, L., Ma, Y., 2022. Estimation of sugar content in sugar beet root based on UAV multi-sensor data. Comput. Electron. Agric. 203, 107433. https://doi.org/10.1016/j.compag.2022.107433.

Watanabe, S., 2023. Tree-Structured Parzen Estimator: Understanding Its Algorithm Components and Their Roles for Better Empirical Performance.

Wei, H.-E., Grafton, M., Bretherton, M., Irwin, M., Sandoval, E., 2022. Evaluation of the use of UAV-derived vegetation indices and environmental variables for grapevine water status monitoring based on machine learning algorithms and SHAP analysis. Remote Sens. (Basel) 14, 5918. https://doi.org/10.3390/rs14235918.

Wijesinha-Bettoni, R., Mouillé, B., 2019. The contribution of potatoes to global food security, nutrition and healthy diets. Am. J. Potato Res. 96, 139–149. https://doi.org/10.1007/s12230-018-09697-1.

Xue, W., Haynes, K.G., Qu, X., 2019. Characterization of early blight resistance in potato cultivars. Plant Dis. 103, 629–637. https://doi.org/10.1094/PDIS-05-18-0794-RE.

Yang, H., Hu, Y., Zheng, Z., Qiao, Y., Zhang, K., Guo, T., Chen, J., 2022. Estimation of potato chlorophyll content from UAV multispectral images with stacking ensemble algorithm. Agronomy 12, 2318. https://doi.org/10.3390/agronomy12102318.

Yang, H., Yin, H., Li, F., Hu, Y., Yu, K., 2023. Machine learning models fed with optimized spectral indices to advance crop nitrogen monitoring. Field Crop Res. 293, 108844. https://doi.org/10.1016/j.fcr.2023.108844.

Yin, H., Huang, W., Li, F., Yang, H., Li, Y., Hu, Y., Yu, K., 2023. Multi-temporal UAV imaging-based mapping of chlorophyll content in potato crop. PFG 91, 91–106. https://doi.org/10.1007/s41064-022-00218-8.

Yu, T., Zhou, J., Fan, J., Wang, Y., Zhang, Z., 2023. Potato leaf area index estimation using multi-sensor Unmanned Aerial Vehicle (UAV) imagery and machine learning. Remote Sens. (Basel) 15, 4108. https://doi.org/10.3390/rs15164108.

Zafari, A., Zurita-Milla, R., Izquierdo-Verdiguier, E., 2019. Evaluating the performance of a random forest kernel for land cover classification. Remote Sens. (Basel) 11, 575. https://doi.org/10.3390/rs11050575.

Zeng, T., Fang, J., Yin, C., Li, Y., Fu, W., Zhang, H., Wang, J., Zhang, X., 2023. Recognition of rubber tree powdery mildew based on UAV remote sensing with different spatial resolutions. Drones 7, 533. https://doi.org/10.3390/drones7080533.