de Oliveira *et al. BMC Plant Biology* (2025) 25:608
https://doi.org/10.1186/s12870-025-06536-4

BMC Plant Biology

RESEARCH

Open Access

Random forest algorithms: a tool to identify the impact of arbuscular mycorrhizal fungi inoculation, seed maturation stage and geographic diversity of *Pimpinella anisum* L. accessions on the physicochemical composition of seeds

Bruno Rodrigues de Oliveira[1], Alan Mario Zuffo[2*], Francisco Charles dos Santos Silva[2], Fábio Steiner[3], Amal Mohamed AlGarawi[4], Mohammad K. Okla[4], Mousa Nhs[5], Yousef Alhaj Hamoud[6,7], Izabela Josko[8], Mohamed S. Sheteiwy[9*], Mohamed Salem Alyafei[9] and Saad Sulieman[9,10]

Abstract

Background A study using random forest (RF) algorithms and principal component analysis (PCA) was proposed to identify the effects of arbuscular mycorrhizal fungal inoculation, the seed maturation stage and the geographic diversity of *Pimpinella anisum* L. accessions on the physicochemical composition of seeds. Seeds of six anise varieties from North African and Middle Eastern accessions were inoculated or not inoculated with AMF (an arbuscular mycorrhizal fungus) and then grown under controlled conditions. Seeds were harvested at three different maturity stages: mature seeds (157 d after sowing), premature seeds (147 d after sowing), and immature seeds (137 d after sowing). Forty-nine variables related to physical properties, total nutrients, metabolic compounds, essential oils, and biological activity were measured in *P. anisum* seeds.

Results The RF algorithm allows the differentiation of *P. anisum* varieties inoculated with AMF from different countries in North Africa and the Middle East. This evidence proves that the geographic origin of *P. anisum* seeds significantly influences the efficiency of the symbiotic association between anise roots and AMF. In turn, no significant effects of the seed maturation stage on the symbiotic interaction of plants with mycorrhizae were observed. The chemical compounds related to the biological activity of seeds are not influenced by AMF, followed by chemical compounds related to metabolism, total nutrients, and oil components.

Conclusions The performance of classification models using RF is driven primarily by independent variables related to the chemical composition of anise seeds, overshadowing the effects of geographic diversity and the seed

*Correspondence:
Alan Mario Zuffo
alan_zuffo@hotmail.com
Mohamed S. Sheteiwy
msalah@uaeu.ac.ae
Full list of author information is available at the end of the article

maturation stage. Among the chemical constituents of the seed, the variables belonging to the biological activity category best contain information (patterns) on the impacts of AMF inoculation.

Keywords Anise, Mycorrhizal association, Essential oil, Metabolic compounds, Biological activity

Introduction

Anise (*Pimpinella anisum* L.), native to the eastern Mediterranean region, is an annual herbaceous plant belonging to the Apiaceae family [1]. It is one of the oldest species used by humanity, being cultivated mainly for its fruits (seeds) with high levels of essential oils (1.5–6%) [2], which are widely used as flavorings, digestives, and carminatives and to relieve gastrointestinal spasms [3]. Anise essential oils are widely used as flavoring agents in culinary dishes, medicines, and perfumery because of their diverse therapeutic and aromatic properties [3]. The oil is highly valued for its antioxidant and antimicrobial capabilities, making it a key component in the formulation of natural remedies for digestive issues, respiratory conditions, and skin health. Its versatile applications across multiple industries, such as food, pharmaceutical, cosmetic, ice cream, sweet, fish, and personal care, underscore its economic significance [4, 5].

Among the sustainable strategies for enhancing aniseed crop production is the inoculation of arbuscular mycorrhizal fungi (AMF). When applied to the soil, these organisms can establish mutualistic associations with plant roots, benefiting fungi and plants. Mycorrhizae improve the efficiency of soil exploration by plants, as they develop an extra-root network formed by mycelia (hyphae), promote greater tolerance to environmental stresses, and improve nutrient and water uptake by plants [6–10]. Studies on anise crops inoculated with AMF have been limited to examining plant growth and morphological traits, seed production, and essential oil content [11–14]. However, further studies are still needed to elucidate the effects of the symbiosis between aniseed plants and AMFs on the chemical and nutritional compositions of their seeds. However, the accumulation of nutritional metabolic compounds in seeds can be influenced by several environmental factors during the crop growth cycle, such as the atmospheric CO_2 concentration [15] and light intensity [16]. Furthermore, continued gains in *P. anisum* seed production from AMF inoculation depend on a better understanding of the factors that can affect the efficiency of the symbiotic association and how this relationship evolves throughout the crop cycle. The effects of environmental factors and production systems, such as the soil water content [17], photoperiod [18], soil fertility and texture [19, 20], and cultivation practices [21, 22], on mycorrhizal associations with *P. anisum* roots are well understood. However, other factors still need to be investigated to assess the effects of the symbiotic interaction between mycorrhizal fungi and *P. anisum* varieties from different countries of origin. The chemical composition of *P. anisum* seeds can be modified by ecological conditions during the seed development and physiological maturity stages [23, 24]. However, for this plant species, few studies have investigated the effects of the symbiotic interaction between mycorrhizae and *P. anisum* varieties from different countries or of the seed maturation stage for harvesting on the chemical properties of the seeds.

Therefore, before new studies investigating the interaction between *P. anisum* genetic material from different countries of origin, AMF inoculation, and the seed maturation stage are conducted, exploratory studies should be performed to verify the existence of these interactive effects. However, studies involving different living organisms and the interaction of several sources of variation generate a significant amount of data, and these data do not have a linear relationship, which is a common phenomenon with most plant morphological traits and the physicochemical composition of seeds [25]. An excellent alternative for processing and identifying patterns in data of this nature is random forest (RF) algorithms. Owing to its excellent results and processing speed, the RF classifier has received increasing attention from researchers in the last two decades [26–28]. The RF algorithm produces reliable results using predictions from an ensemble of decision trees [29]. Furthermore, this technique can be successfully used to select and classify variables with the most remarkable ability to discriminate between target classes [26]. Therefore, the use of an RF algorithm can guide future studies on the variables that can be measured to analyze a given phenomenon.

The RF classifier is commonly applied using the original data collected in the study; however, because of the large number of variables that are analyzed, the occurrence of problems such as multicollinearity may be a relevant factor for the lower efficiency of RF algorithms. An alternative is the use of principal components of the data in RF algorithms. Principal component analysis (PCA) is widely used to develop machine learning models while minimizing information loss and reducing the complexity of numerous dependent variables [30]. This reduction simplifies the feature space and increases computational efficiency [31].

The primary objective of this study was to investigate the effects of AMF inoculation on the physicochemical composition and biological activity of *P. anisum* seeds. Specifically, this research seeks to quantify how AMF influences various traits, including nutrient composition, metabolic compounds, and essential oil content, under different environmental and genetic conditions. By employing advanced statistical methods such as PCA, we aim to identify key variables that differentiate AMF-inoculated plants from their noninoculated counterparts. This study innovatively employs the RF algorithm in conjunction with PCA to identify the impacts of AMF inoculation, the seed maturation stage, and the geographical diversity of *P. anisum* accessions on the physicochemical composition of seeds. While the impact of AMF inoculation on plant growth, morphology, and essential oil content has been studied, this study uniquely extends this research by focusing on the chemical and nutritional composition of aniseed seeds under various environmental and genetic conditions. By combining these powerful tools, this study provides a robust and comprehensive analysis of the intricate relationships among the variables involved. Overall, we hypothesize that these factors interact significantly, influencing seed quality through changes in nutrient absorption and genetic and environmental differences. In conjunction with RF, PCA enables the analysis of complex datasets with multiple variables, providing a clearer understanding of the interactive effects between these factors.

Materials and methods

Plant material, growth conditions, and seed harvest

Viable seeds of six aniseed (*P. anisum* L.) varieties from North African and Middle Eastern accessions, one variety from Morocco ('Halawa2', source: National Institute of Agricultural Research, Rebat), Tunisia ('Dulce'), Egypt ('Baladi', source: Agricultural Research Center, Giza), Yemen ('Fam', source: National research centers, Dhamar), Syria ('Ajmer Anise-1', source: General Commission for Scientific Agricultural Research, Damascus) and Turkey ('Gülsüm BOZTAŞ1 Emine BAYRAM1', source: Turkish Seed Gene Bank, Ankara), were used. The plants were sown in 35×25 cm pots filled with a mixture of loamy soil and organic compost at a 1:1 (v:v) ratio. The soil initially contained 9.74 mg C kg^{-1} soil, 13.1 mg N-NO_3 kg^{-1} soil, 0.98 mg N-NH_4 kg^{-1} soil, and 10.36 mg P kg^{-1} soil, with a moisture content at field capacity of 0.41 g water g^{-1} soil. The pH was 7.08, the EC was 3.7 dS m^{-1}, and the K was 2.75 meq L^{-1} soil. These properties influence nutrient availability, microbial activity, and plant growth.

The substrate's water content was adjusted to 60% of the field capacity. Field capacity was determined by saturating the soil and allowing excess water to drain. We then set the water content to 60% of this value and measured it at various points during drying to maintain the desired level. The pots were divided into two groups: the first group was inoculated with AMF, and the second group was not inoculated (control). AMF (*Rhizophagus irregularis* (MUCL 41,833 obtained from Glomeromycota in vitro harvest [GINCO]) were previously selected from the rhizosphere of medicinal plants and demonstrated high potential to increase seed yield and the accumulation of bioactive metabolites in aniseed plants [32]. The AMF inoculum was applied before sowing and was placed at a depth of 5.0 cm beneath the soil surface. Each pot received 10 g of inoculum, consisting of trapped soil with approximately 50 spores g^{-1}. An equal amount of autoclaved inoculum for the control group was added to ensure the same nutrient levels, excluding the mycorrhizal spores. The control pots without AMF received equivalent amounts of autoclaved inoculum to ensure the same nutrient supply. Five seeds were sown per pot.

A total of 96 pots (8 pots per replicate) were transferred to the growth chamber and kept for up to 157 d under controlled conditions [05:00–21:00 h (22 °C)/21:00–05:00 h (18 °C), photoperiod of 16/8 h (light/darkness), relative humidity of 60% and light intensity of 150 µmol m^{-2} s^{-1}]. The plants were watered daily to maintain the soil water content at 65% field capacity. The experiment was replicated twice. *P. anisum* seeds were harvested at three maturity stages, including immature seeds (harvested 137 d after sowing), premature seeds (harvested 147 d after sowing), and mature seeds (harvested 157 d after sowing).

After harvest, 49 variables related to physical properties, total nutrients, metabolic compounds, essential oils, and biological activity were measured in anise seeds. Datasets from studies by Balkhyour et al. [15] and Mahmoud et al. [32] were tested via various multivariate analysis techniques. A total of 72 experimental units (samples) were arranged in a completely randomized design in a 3×6×2 factorial scheme (three seed maturation stages, six countries of origin of anise seeds, and two treatments with AMFs [noninoculated plants and plants inoculated with AMF]) with two replicates were used in this study.

Machine learning, RF, and PCA

Pattern classification algorithms, including RF, typically involve training and testing steps. During training, the algorithm learns from a labeled dataset, identifying patterns and relationships within the data [33]. This trained model is then tested on an unseen subset of the data to evaluate its performance. A common technique to ensure robust evaluation is tenfold cross-validation [34], where ten parts divide the dataset. To minimize bias and variance in performance estimation, the algorithm is tested on one of nine parts while being trained on the remaining parts, with this process rotating through all parts.

During training, the RF algorithm, a robust ensemble learning method, constructs multiple decision trees [35]. In an RF, each tree is constructed from a random subset of the data and features, and the final classification is derived by aggregating the predictions from all the trees. A key feature of this method is the out-of-bag (OOB) score, which offers an unbiased estimate of the model's performance by using samples not included in the training subset of each tree [36]. The RF measures the impact of each variable on the model's accuracy to evaluate the importance of variables [37], which aids in identifying the most influential features.

Performance metrics are crucial for assessing model effectiveness. Providing a general sense of model performance, accuracy measures the proportion of correctly classified instances [38]. The receiver operating characteristic (ROC) curve and the area under the curve (AUC) are used to evaluate the model's ability to distinguish between classes. The F1 score, which is the harmonic mean of precision and recall, balances the trade-off between these two metrics and is particularly useful in imbalanced datasets [39], with the ROC curve plotting the true positive rate against the false positive rate and the AUC providing a single value summary of the model's discriminatory power.

PCA is a machine learning technique for dimensionality reduction [40]. It captures the most significant variance in the data by transforming a large set of variables into a smaller set of uncorrelated variables known as principal components [30]. PCA identifies the principal components and the directions along which the data's variance is maximized. PCA effectively compresses the dataset by retaining only the top principal components, reducing its dimensionality while preserving the most critical information [41]. This reduction simplifies the feature space, enhances computational efficiency [31], and often improves the performance of subsequent analysis or machine learning tasks by mitigating the curse of dimensionality.

Proposed approach

The proposed approach aims to differentiate between countries, maturation stages, and a linear combination of the physicochemical composition of *P. anisum* seeds, with an explicit focus on the effects of AMF inoculation. To achieve this, we utilize the RF, renowned for its efficacy in pattern classification and its ability to assess variable importance via decision tree learning schemes [42, 43]. The selection of the RF for classifying sample patterns between inoculated and noninoculated plants with AMF was a deliberate and thoughtful choice. Among the algorithms that evaluate variable importance, the RF is recognized for its superior performance in classification functions [37].

A tenfold cross-validation approach was employed to train and validate the models using the RF algorithm. The hyperparameters of the algorithm, as implemented in the Python library called Scikit Learn [44], were optimized for the best performance: bootstrap = True, criterion = entropy, oob_score = True, n_estimators = 1000, max_features = sqrt, random_state = 1. The remaining hyperparameters were used with the standard values. The performance of the models in the training phase was evaluated for each fold utilizing the OOB score, accuracy, F1 score, and ROC metrics.

In addition, we employ PCA to address the high dimensionality of the dataset, which comprises 49 dependent variables related to the physicochemical composition (physical properties, total nutrients, metabolic components, essential oils, and biological activity) of *P. anisum* seeds. PCA captures most of the variability in the data by reducing the dataset to two principal linear combinations (components), simplifying the feature spaces. This dimensionality reduction enhances model interpretability and improves classification performance, aligning with established findings in the literature [31, 41].

Statistical analysis and machine learning

For statistical analysis, analyses of variance (ANOVA) and tests of normality and homogeneity of variance (Levene, Bartlett, Shapiro–Wilk, Kolmogorov–Smirnov) were performed on all the data using Jamovi® software (version 2.5) for Windows [45], with a significance level of 5%. The RF and PCA algorithms were implemented in Python with the Scikit Learn [44] library on the Google Colab platform [46].

Results

The results of the ANOVA revealed that the dependent variables exhibited significant responses to different sources of variation (Tables 1 and 2). To effectively combine these variables for further analysis, PCA, which is crucial for reducing dimensionality and enhancing the interpretability of the data, was employed.

The p values for the statistical tests of normality and homogeneity of variance, shown in Table 1, reveal a complex data distribution. Many dependent variables, such as seed dry weight, seed water content, saponin content, and several monoterpene hydrocarbons, significantly deviate from normality and homogeneity of variance, as indicated by p values less than 0.05 in Levene's, Bartlett's, Shapiro–Wilk, and Kolmogorov–Smirnov tests.

Table 1 *P* values were calculated for statistical tests of normality and homogeneity of variances. Number of replicates ($n=5$)

Variable	Statistical test			
	Levene	Bartlett	Shapiro–Wilk	Kolmogorov–Smirnov
	Physical properties			
Seed dry weight	0.037	< 0.001	< 0.001	0.007
Seed water content	< 0.001	< 0.001	< 0.001	< 0.001
	Total nutrients			
Saponin	0.005	< 0.001	< 0.001	0.013
Steroid	0.015	< 0.001	< 0.001	0.019
Total protein	0.230	< 0.001	0.001	0.013
Total sugar	0.167	0.069	0.002	0.037
Ash	0.003	< 0.001	< 0.001	0.006
Crude fiber	0.020	< 0.001	< 0.001	0.032
Total phenols	1.000	1.000	< 0.001	0.032
Total flavonoids	0.001	< 0.001	< 0.001	< 0.001
Total alkaloid	0.284	0.123	0.017	0.037
Tannin	0.404	0.011	0.004	0.026
	Metabolic compounds			
Phenylalanine	0.005	0.002	0.009	0.345
L-phenylalanine ammonia-lyase	< 0.001	< 0.001	< 0.001	0.106
3-Deoxy-D-arabinoheptulosonate 7-phosphate synthase	< 0.001	0.008	0.011	0.181
Cinnamic acid	< 0.001	< 0.001	< 0.001	0.002
Shikimic acid	< 0.001	< 0.001	0.010	0.161
O-methyltransferase	< 0.001	< 0.001	< 0.001	0.062
	Oil compounds			
Oil yield	0.069	0.437	0.654	0.684
Essential oil	0.069	0.437	0.654	0.684
Phenylpropanoids trans-anethole	0.008	0.305	0.384	0.668
Phenylpropanoids o-isoeugenol	0.866	0.999	0.010	0.138
Phenylpropanoids p-anisaldehyde	< 0.001	< 0.001	0.002	0.110
Phenylpropanoids anisole	0.521	0.565	0.389	0.321
Phenylpropanoids p-anisaldehyde	< 0.001	< 0.001	< 0.001	< 0.001
Phenylpropanoids estragole	0.255	0.603	0.198	0.735
Monoterpene hydrocarbons α-pinene	< 0.001	< 0.001	< 0.001	< 0.001
Monoterpene hydrocarbons limonene	0.009	0.031	0.002	0.211
Monoterpene hydrocarbons myrcene	0.006	0.018	0.112	0.326
Monoterpene hydrocarbons linalool	0.471	0.080	0.014	0.400
Monoterpene hydrocarbons cis-β-ocimene	0.335	0.595	0.349	0.554
Monoterpene hydrocarbons sabinene	< 0.001	< 0.001	0.247	0.377
Monoterpene hydrocarbons p-cymene	0.013	0.003	< 0.001	0.177
Monoterpene hydrocarbons α-phellandrene	0.206	0.616	0.076	0.472
Monoterpene hydrocarbons fenchone	0.359	0.458	0.791	0.902
Monoterpene hydrocarbons 1,8-cineole	< 0.001	< 0.001	< 0.001	0.016
Monoterpene hydrocarbons α-fenchyl acetate	< 0.001	< 0.001	< 0.001	< 0.001
Monoterpene hydrocarbons α-terpinene	< 0.001	< 0.001	< 0.001	< 0.001
Sesquiterpene hydrocarbons γ-himachalene	< 0.001	< 0.001	< 0.001	< 0.001
Sesquiterpene hydrocarbons isolongifolene	< 0.001	< 0.001	< 0.001	< 0.001
Sesquiterpene hydrocarbons β-elemene	0.007	0.046	0.016	0.270
Sesquiterpene hydrocarbons zingiberene	< 0.001	< 0.001	0.002	0.055
	Biological activity			

Table 1 (continued)

Variable	**Statistical test**			
	Levene	**Bartlett**	**Shapiro–Wilk**	**Kolmogorov–Smirnov**
Amylase	< 0.001	< 0.001	< 0.001	< 0.001
Lipase	< 0.001	< 0.001	< 0.001	< 0.001
Anti-cholesterol	< 0.001	< 0.001	< 0.001	0.002
2,2-Diphenyl-1-picrylhydrazyl	0.264	0.995	0.008	0.114
Total antioxidant capacity	< 0.001	0.370	0.006	0.848
Anti-lipid peroxidation	< 0.001	0.003	< 0.001	0.061

This complexity underscores the challenge of analyzing data using traditional statistical methods, necessitating advanced techniques.

Table 2 shows the p values from the variance analysis for the effects of seed maturity stage, country of origin, AMF treatment, and their interactions on the physicochemical composition of *P. anisum* accessions. Several chemical compounds of *P. anisum* seeds, including saponin, crude fiber, total sugar, ash, total flavonoids, total phenols, and multiple phenylpropanoids and monoterpene hydrocarbons, significantly affected the sources of variation. For example, saponin, total sugar, and ash contents are significantly affected by country of origin, seed maturity stage, and AMF treatment, indicating a complex interaction between these sources of variation for these variables.

In summary, the ANOVA results underscore the significant and complex effects of country of origin, seed maturity stage, and AMF treatment on various physicochemical variables of *P. anisum* accessions. These findings are crucial for understanding the factors influencing the physicochemical composition of *P. anisum* seeds. Using PCA to combine these variables into principal components is instrumental in reducing dimensionality, addressing multicollinearity, and enhancing the interpretability and classification performance of the data. This approach enables more effective differentiation of noninoculated and AMF-inoculated plants, thereby facilitating the identification of the most important variables in the classification process.

Table 3 presents the explained variance ratios for each component derived from the PCA. These ratios, segmented by variable categories, demonstrate the effectiveness of reducing the dimensionality of the dataset while retaining most of the variability. This reduction, facilitated by PCA, paves the way for the subsequent application of the RF algorithm for classifying and differentiating *P. anisum* accessions, underscoring the utility of this technique in this study.

The results of the RF models used to evaluate the importance of the physicochemical variables of *P. anisum* seeds in the differentiation of plants inoculated with AMF from noninoculated plants (control) are shown in Figs. 1, 2, 3, 4 and 5. These algorithms show that the greater capacity of the variables to differentiate plants inoculated with AMF represents the more significant influence of the symbiotic interaction between AMF and *P. anisum* roots in this set of dependent variables. Figures 1, 2, 3, 4 and 5 (a) show the variations in the performance metrics of the classification models obtained. Figures 1, 2, 3, 4 and 5 (b) show the importance of the variable's country of origin, seed maturity stage, and the PCA 1 and PCA 2 components of the PCA.

The PCA combination of different variables was considered for each result. As shown in Fig. 1, all variables related to the physicochemical composition of *P. anisum* seeds were used for linear combination with PCA. The results in Fig. 1 reveal that the variability represented by the principal components of the PCA, followed by the geographic origin of the seeds, is the most relevant factor for classification. The variable "stage" has the least importance, with values close to zero, indicating that the stage of seed maturation has little or no influence on the classification of the samples by the model, considering all the variables collected. In Fig. 2, only the variables related to the composition of total nutrients are combined. These results demonstrate that when considering only the total nutrient composition variables, the RF model performs significantly better in classifying samples than when combining all physicochemical variables (Fig. 1). The performance metrics (accuracy, F1 score, and ROC) are relatively high and have relatively low variability, indicating a relatively robust and reliable model. The principal components (PCA 1 and PCA 2) remain the most important variables for classification, with a more balanced contribution between them. The variable "Country" maintains an intermediate influence, whereas "Stage" remains unimportant. Figure 3 shows the results of the

Table 2 *P* values of ANOVA for the impacts of country of origin, maturation stage, and treatment with AMF on the physicochemical properties of *P. anisum* seeds

Variable	Standard error	Source of Variation						
		Country (C)	Maturation Stage (M)	AMF treatment (T)	Interaction C × M	Interaction C × T	Interaction M × T	Interaction C × M × T
		Physical properties						
Seed dry weight	2.93	< 0.001	< 0.001	0.299	0.255	0.170	0.005	0.156
Seed water content	3.01	0.516	< 0.001	0.347	0.955	0.882	0.674	0.899
		Total nutrients						
Saponin	1.26	< 0.001	< 0.001	< 0.001	0.931	0.003	0.454	0.062
Steroid	14.6	0.722	< 0.001	0.153	0.948	0.324	0.698	0.739
Total protein	4.87	0.882	0.546	0.180	1.000	0.984	0.997	1.000
Total sugar	7.89	< 0.001	< 0.001	< 0.001	0.120	0.502	0.016	0.910
Ash	3.22	< 0.001	< 0.001	< 0.001	0.837	0.492	0.694	0.983
Crude fiber	0.32	< 0.001	< 0.001	< 0.001	0.959	< 0.001	0.183	0.887
Total phenols	0.46	< 0.001	0.168	0.014	0.035	0.098	0.998	1.000
Total flavonoids	1.34	< 0.001	< 0.001	< 0.001	< 0.001	0.003	< 0.001	< 0.001
Total alkaloid	6.90	< 0.001	< 0.001	0.002	0.162	< 0.001	0.138	0.060
Tannin	1.75	< 0.001	< 0.001	< 0.001	0.058	0.005	0.402	0.703
		Metabolic compounds						
Phenylalanine	0.07	< 0.001	< 0.001	< 0.001	0.125	< 0.001	0.584	0.407
L-phenylalanine ammonia-lyase	2.73	0.033	< 0.001	< 0.001	0.039	< 0.001	< 0.001	0.459
3-Deoxy-D-arabinoheptulosonate 7-phosphate synthase	0.02	< 0.001	< 0.001	< 0.001	0.003	< 0.001	0.002	< 0.001
Cinnamic acid	0.14	0.014	< 0.001	< 0.001	< 0.001	0.013	0.178	< 0.001
Shikimic acid	1.97	< 0.001	< 0.001	0.228	< 0.001	0.291	0.157	0.032
O-methyltransferase	1.10	0.004	< 0.001	< 0.001	0.871	0.011	< 0.001	0.922
		Oil compounds						
Oil yield	0.23	< 0.001	< 0.001	< 0.001	0.442	< 0.001	0.072	0.594
Essential oil	0.07	< 0.001	< 0.001	< 0.001	0.442	< 0.001	0.072	0.594
Phenylpropanoids trans-anethole	1.19	< 0.001	< 0.001	< 0.001	0.388	< 0.001	0.091	0.944
Phenylpropanoids o-isoeugenol	0.11	0.007	0.099	0.010	1.000	< 0.001	0.807	1.000
Phenylpropanoids p-anisaldehyde	0.02	0.987	< 0.001	0.527	0.991	0.016	0.211	0.358
Phenylpropanoids anisole	0.121	< 0.001	< 0.001	0.316	0.874	< 0.001	0.515	0.993
Phenylpropanoids p-anisaldehyde	0.276	< 0.001	< 0.001	0.229	< 0.001	< 0.001	0.093	< 0.001
Phenylpropanoids estragole	0.249	0.027	< 0.001	< 0.001	0.322	0.091	0.518	0.988
Monoterpene hydrocarbons α-pinene	0.160	0.007	< 0.001	0.189	< 0.001	0.001	0.118	< 0.001
Monoterpene hydrocarbons limonene	0.07	0.014	< 0.001	0.381	0.155	< 0.001	0.139	0.062
Monoterpene hydrocarbons myrcene	0.185	< 0.001	< 0.001	< 0.001	0.116	< 0.001	0.042	< 0.001

Table 2 (continued)

Variable	Standard error	Source of Variation						
		Country (C)	**Maturation Stage (M)**	**AMF treatment (T)**	**Interaction C × M**	**Interaction C × T**	**Interaction M × T**	**Interaction C × M × T**
Monoterpene hydrocarbons linalool	0.09	0.004	< 0.001	< 0.001	0.573	< 0.001	0.791	0.741
Monoterpene hydrocarbons cis-β-ocimene	0.17	0.133	< 0.001	< 0.001	0.488	< 0.001	< 0.001	< 0.001
Monoterpene hydrocarbons sabinene	0.05	< 0.001	< 0.001	0.275	0.273	0.002	0.042	0.404
Monoterpene hydrocarbons p-cymene	0.07	< 0.001	< 0.001	< 0.001	0.006	< 0.001	< 0.001	< 0.001
Monoterpene hydrocarbons α-phellandrene	0.11	< 0.001	< 0.001	0.003	0.908	0.004	0.313	0.506
Monoterpene hydrocarbons fenchone	0.15	< 0.001	< 0.001	0.001	0.112	0.696	0.102	0.651
Monoterpene hydrocarbons 1,8-cineole	0.14	0.798	< 0.001	0.023	0.901	0.050	0.059	0.934
Monoterpene hydrocarbons α-fenchyl acetate	0.11	0.819	< 0.001	0.527	0.994	0.068	0.699	0.137
Monoterpene hydrocarbons α-terpinene	0.02	0.308	< 0.001	0.769	0.506	0.040	0.255	0.587
Sesquiterpene hydrocarbons γ-himachalene	0.02	0.156	< 0.001	0.198	0.024	< 0.001	0.702	0.047
Sesquiterpene hydrocarbons isolongifolene	0.05	0.313	< 0.001	0.058	0.064	< 0.001	0.019	< 0.001
Sesquiterpene hydrocarbons β-elemene	0.16	0.008	0.063	0.181	0.990	< 0.001	0.148	0.860
Sesquiterpene hydrocarbons zingiberene	0.08	0.601	< 0.001	0.101	1.000	0.012	0.630	0.744
		Biological activity						
Amylase	1.45	0.003	< 0.001	< 0.001	0.869	< 0.001	0.026	0.884
Lipase	1.00	< 0.001	< 0.001	< 0.001	< 0.001	0.006	< 0.001	< 0.001
Anti-cholesterol	4.34	< 0.001	< 0.001	0.816	0.391	0.188	< 0.001	0.050
2,2-Diphenyl-1-picrylhydrazyl	2.86	< 0.001	< 0.001	0.342	0.732	0.971	0.543	0.986
Total antioxidant capacity	2.67	0.025	< 0.001	0.069	0.850	0.347	0.928	0.978
Anti-lipid peroxidation	1.13	0.006	< 0.001	0.002	0.381	0.285	0.644	0.589

variables related to the composition of the seed metabolic compounds. The results show that the exclusive use of metabolic compound variables results in less robust classification performance than the use of total nutrient variables (Fig. 2), resulting in greater variability in the performance metrics. Although the ROC still indicates a reasonable discrimination capacity, the greater variation in accuracy and F1 score suggests that the metabolic

Table 3 Variance explained by the PCA components, considering the different categories of physicochemical compositions of *P. anisum* seeds

Variable Set	Explained variance (%)	
	PCA 1	**PCA 2**
All	69.87	9.40
Physical Properties	97.68	2.31
Total Nutrients	76.40	10.89
Metabolic compounds	73.73	24.90
Oil compounds	84.73	7.46
Biological Activity	59.19	25.90

compound composition is not as effective in differentiating samples as the nutrient composition. The principal components, PCA 1 and PCA 2, remain the most important, with a more balanced contribution between them. The variable "Country" maintains a low influence, and "Stage" has no significant influence. Figure 4 presents the results related to the composition of the oil compounds, showing that the exclusive use of oil compound variables results in a classification performance with distinct characteristics. Although the discrimination ability between classes (measured by ROC) remains reasonable, the greater variability, mainly in the F1 score, indicates a potential imbalance between precision and recall, suggesting that the model has more difficulty correctly classifying samples based on only oil compounds. The principal components PCA 1 and PCA 2 continue to be the most important, with a more balanced contribution between them. The variable "Country" maintains a low influence, and "Stage" has no significant influence. In Fig. 5, the variables related to the components of the biological activity of the seeds are shown. The results in Fig. 5 reveal that using biological activity compound variables results in good classification performance, with high consistency (low variation in accuracy) and good discrimination ability between classes (high ROC). The F1 score suggests a slight imbalance between precision and recall but to a lesser extent than that observed for oil compounds. PCA 1 stands out as the primary determinant for classification, indicating that the variability captured by this component is crucial for differentiating samples based on biological activity. The variables "Country" and "Stage" have little or no influence.

Table 4 presents the OOB score values for each fold of the tenfold cross-validation, categorized by variable categories and computed for all the variables. These results are consistent with the patterns depicted in the previous graphs, offering a detailed overview of the RF model's performance across different subsets of variables. The scores highlight the model's performance across different variable contexts, demonstrating the utility of each subset in contributing to the overall classification accuracy. Figures 6 (a)-(f) display the ROC curve and the corresponding AUC values for each cross-validation. Assessments were conducted across various variable categories and for the entire dataset. Each figure includes the mean ROC curve across folds, the mean AUC, and the corresponding standard deviation.

Discussion

Studying complex biological data, especially when multiple interacting factors, such as AMF inoculation of *P. anisum* seeds of different varieties, often results in large datasets with nonlinear relationships between variables. This is common in terms of plant morphological traits and seed physicochemical composition, the complexity of which can hinder effective analysis. To address this, the PCA and RF algorithms are valuable tools. PCA simplifies data by reducing dimensionality, transforming correlated variables into uncorrelated principal components, and retaining the data's variability. Other studies have used PCA to analyze *P. anisum* accessions, although with different approaches. Balkhyour et al. [15] employed biplot graphs to analyze the impact of high CO_2 on anise seed chemical composition. Moreover, based on SNP data, Mehrayi et al. [51] used PCA to assess species diversity in *Pimpinella* species.

In addition to PCA and RF, an ensemble of decision trees complements PCA by offering reliable classification results and quickly processing large datasets. In our study, to assess how AMF inoculation affects the chemical composition of *P. anisum* seeds, PCA was applied to reduce dimensionality and simplify data complexity for effective classification (Tables 1, 2, and 3). By transforming correlated variables into uncorrelated PCs, PCA helps address multicollinearity, retaining most of the data's variance. This simplification is crucial for models such as RF, as it improves classification accuracy.

The first components (PC1 and PC2), alongside the country and stage variables, enhance the understanding and visualization of the dataset, enabling more precise insights into the interaction effects of the studied factors. This method is widely utilized to reduce the complexity of numerous variables and develop machine learning models while simultaneously minimizing information loss [30]. Our results revealed significant variations, with some compounds, such as total flavonoids and monoterpene hydrocarbons, showing consistent deviations from normality and homogeneity assumptions, which are important for subsequent analyses (Table 1). ANOVA results for the effects of country of origin, maturation

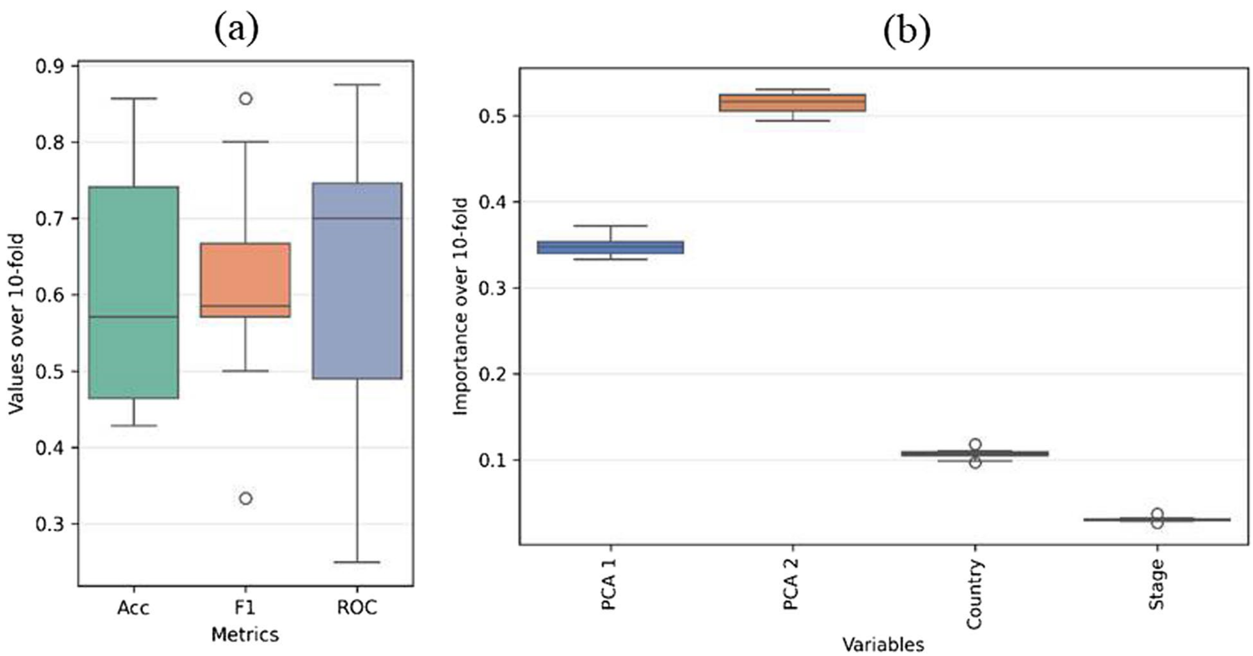

Fig. 1 Results considering the combination of all variables related to the physicochemical composition of *P. anisum* seeds by the PCA. **a** Performance metrics. **b** Importance of variables

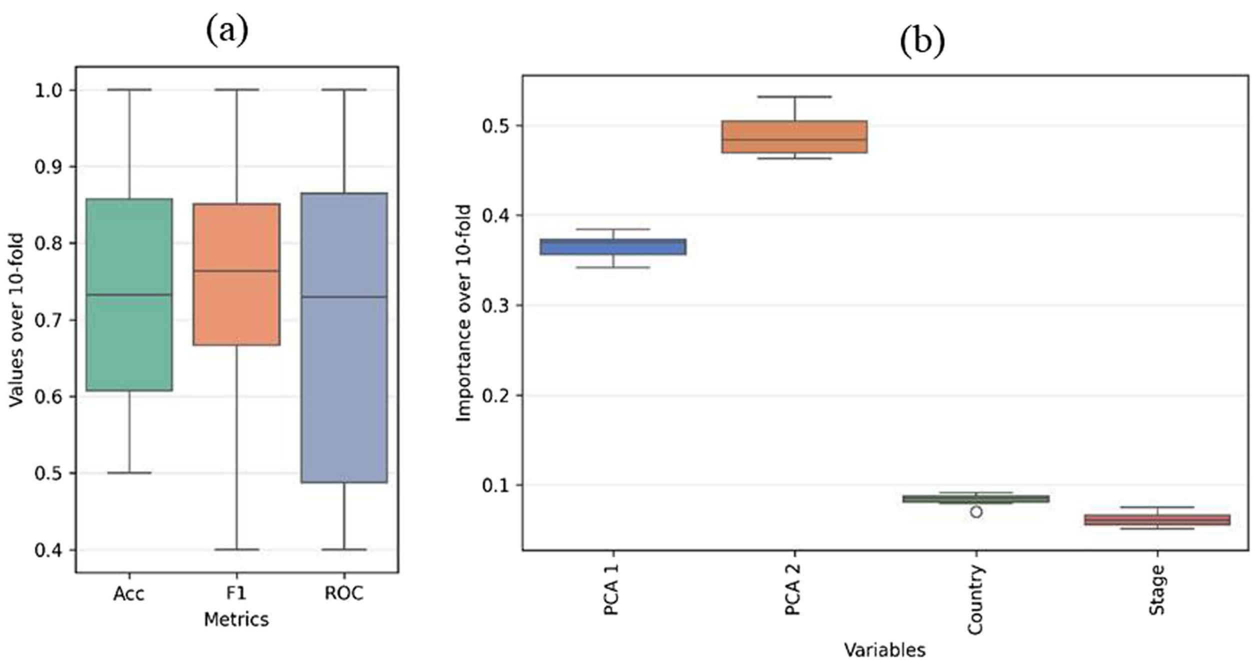

Fig. 2 Results of the PCA for variables of the total nutrient composition of *P. anisum* seeds. **a** Performance metrics, **b** Importance of variables

stage, AMF treatment, and their interactions. Significant *p* values for the country of origin and maturation stage highlight their impact on seed composition. Notably, compounds such as total flavonoids and monoterpene hydrocarbons significantly interact with these factors, emphasizing their role in biochemical variation, as also observed in other studies linking the origin and maturity stage to aniseed composition [23]. Inoculated vs.

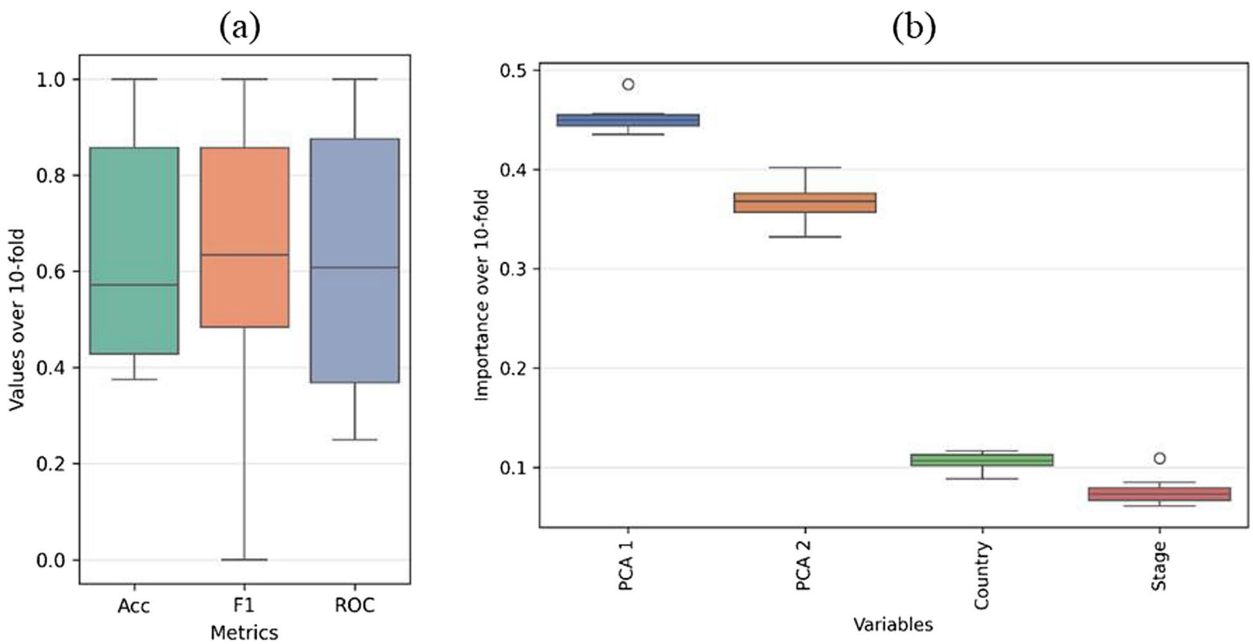

Fig. 3 Results of the PCA for variables associated with the metabolic compound composition of *P. anisum* seeds. **a** Performance metrics, **b** Importance of variables

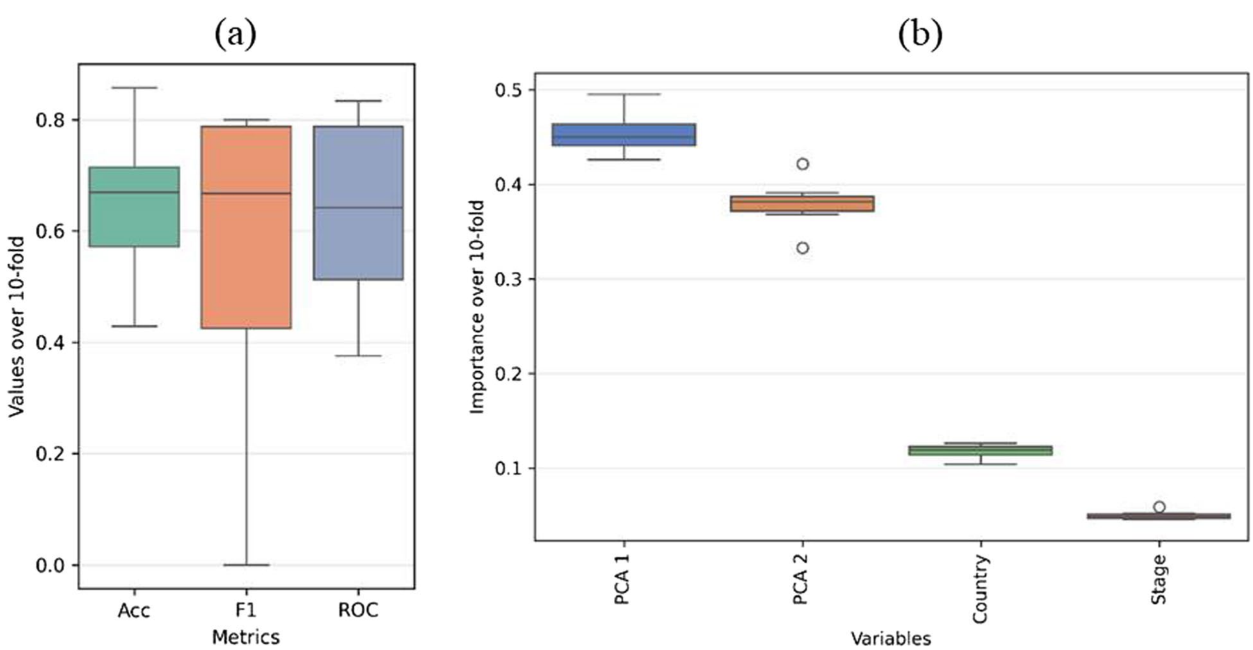

Fig. 4 Results of the PCA for variables associated with the oil compounds of *P. anisum* seeds. **a** Performance metrics, **b** Importance of variables

noninoculated plants were classified using PCA, with a median accuracy and F1 score of approximately 60%. Some folds exceeded 80%, whereas others were lower, with variability linked to limited sample sizes, affecting model accuracy and overfitting (Fig. 1). The PCA weight over country of origin and maturation stage in distinguishing plants was measured, and the boxplot shows consistent variable importance across folds despite performance variations (Fig. 1b). Our data also revealed greater variability in the importance of variables, with the country and maturation stage being more closely related (Fig. 2b), highlighting the significant role of geographic

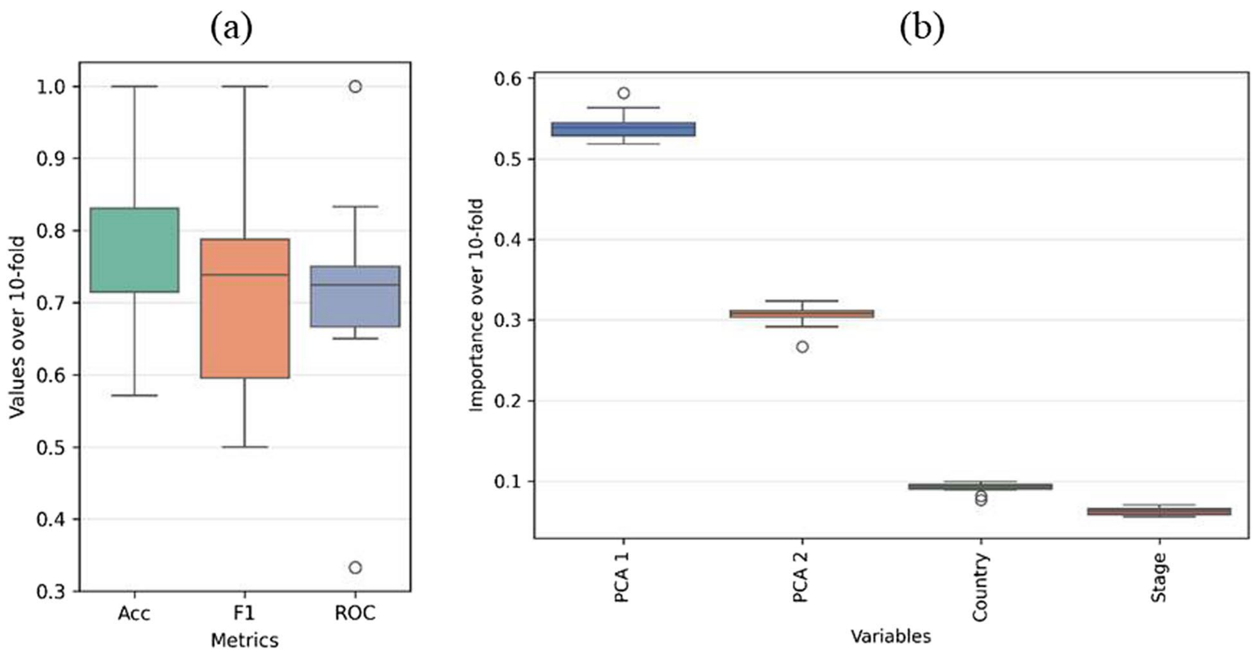

Fig. 5 Results of the PCA of the variables related to the biological activity of *P. anisum* seeds. **a** Performance metrics, **b** Importance of variables

diversity in AMF inoculation efficiency, with previous studies supporting the impact of origin on the chemical composition of seeds (Figs. 3, 4 and 5b). In this context, anise quality is determined mainly by its essential oil content, composition, and total flavonoid content [15]. Overall, the analyses in Tables 1, 2, and 3 and Figs. 1, 2, 3, 4 and 5 highlight the effectiveness of PCA in simplifying datasets for classifying AMF-inoculated vs. noninoculated plants. The physicochemical variables of seeds consistently exceeded those of the cultivation country or maturation stage in the RF-based models. Table 2 shows that, compared with the seed properties, only 12% of the variables are significantly affected by the interaction of these factors, indicating their limited impact.

In this study, compared with noninoculated plants, AMF increased essential oil accumulation. This finding suggests increased phenylpropanoid and terpene biosynthesis, although direct measurements were not made. Phenylpropanoid compounds are the main components of essential oils and are synthesized from phenylalanine through the cinnamic and shikimic acid pathways [52]. The essential oil content and chemical composition are

Table 4 Random forest out-of-bag (OOB) score results for each of the folds considered and variable contexts

Fold Cross-validation	Variables of the seed physicochemical composition					
	All	Physical properties	Total nutrients	Metabolic compounds	Oil compounds	Biological activity
1	0.7031	0.5156	0.7344	0.6719	0.6094	0.7500
2	0.6875	0.5938	0.6250	0.6875	0.6719	0.7188
3	0.6615	0.5846	0.6308	0.7538	0.6308	0.7385
4	0.7077	0.6308	0.6923	0.6000	0.6308	0.6923
5	0.6308	0.6000	0.5846	0.7231	0.6615	0.7385
6	0.6923	0.5846	0.5846	0.6923	0.6462	0.7231
7	0.6769	0.5846	0.6462	0.6462	0.6923	0.7231
8	0.6615	0.6000	0.6308	0.7077	0.6923	0.7077
9	0.7692	0.5692	0.7385	0.7231	0.6462	0.7385
10	0.6769	0.6308	0.6462	0.6000	0.5846	0.7231
Average	0.6868	0.5894	0.6513	0.6806	0.6466	0.7253

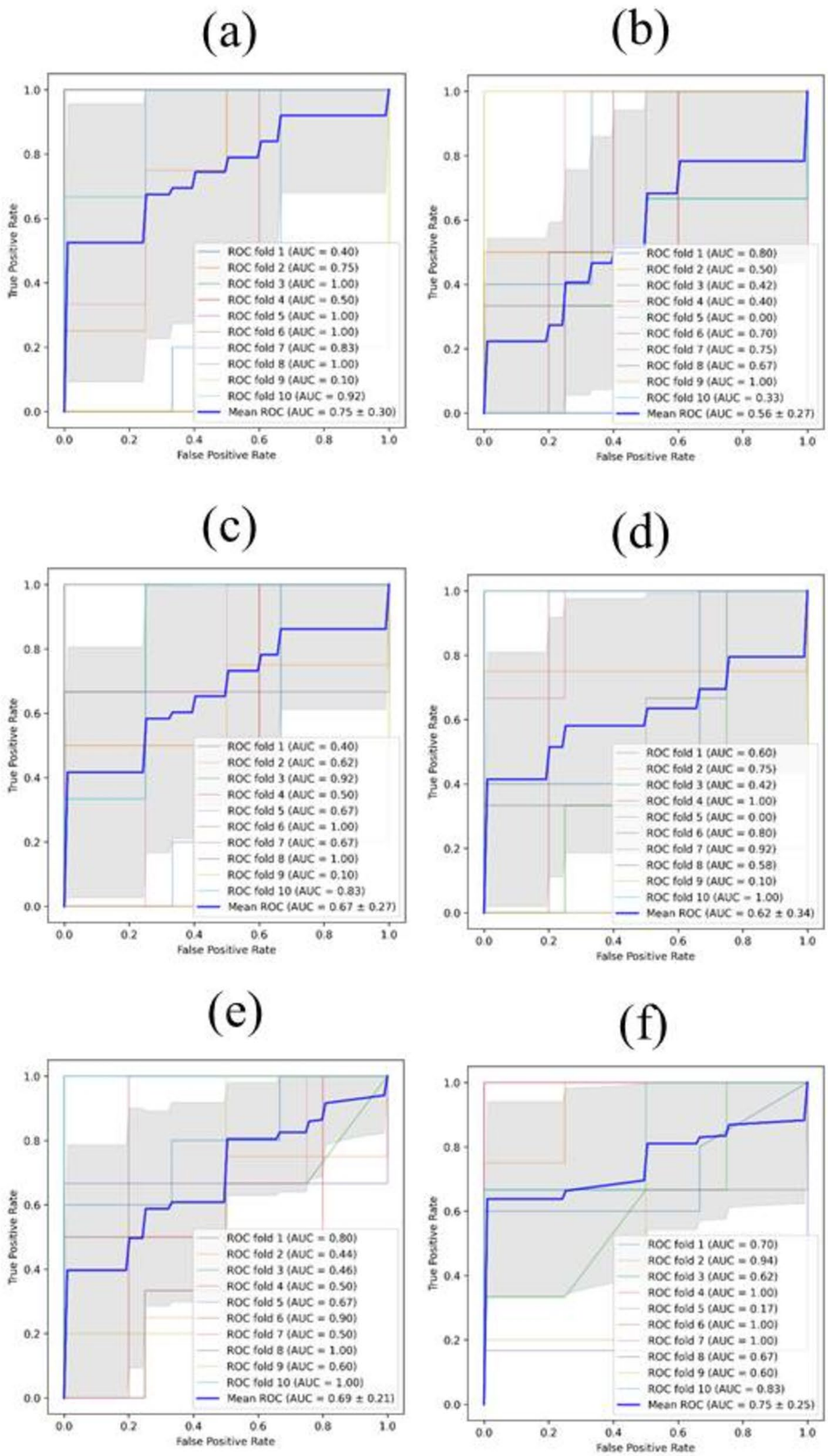

Fig. 6 The ROC curve and the AUC were used to evaluate the model's capacity to distinguish between AMF-inoculated plants and noninoculated (control) plants and were computed for each fold. The evaluations were conducted for the following variable categories: (**a**) All variables, (**b**) physical properties, (**c**) total nutrients, (**d**) metabolic compounds, (**e**) oil compounds, and (**f**) biological activity

critical factors, and anise quality is also influenced by a seed's developmental stage and geographical origin [15]. In this context, significant interactions between environmental factors, seed maturity, and origin play crucial roles in shaping anise's chemical composition and pharmaceutical properties. Several factors may explain these results [15]. Aćimović et al. [53] noted that climatic conditions significantly affect chemical composition and AMF symbiosis. Each country's specific conditions influence the chemical composition of the seeds, affecting their symbiotic associations with AMF. Genetic differences between *P. anisum* accessions from different countries also play a role, as geographic distance can lead to unique genetic traits within populations. The colonization process of AMF involves stages, pre-colonization, penetration/colonization, and establishment, each of which is regulated by species-specific genes [54]. Variations in these genes among plants from different countries can impact symbiosis efficiency, as observed in sorghum crops by Cobb et al. [55].

AMF significantly interferes with total nutrients, metabolic compounds, many oil compounds, and all variables associated with biological activity in anise seeds (Table 2). Because many chemical compounds in plants are involved in their adaptation to environmental conditions, these results can be altered by external environmental factors [15]. In this study, the variable environmental factor was the inoculation of AMF, which colonizes the plant root system. The degree of colonization by AMF was found to vary across the different varieties used in the study, with colonization ranging from 45 to 57% among the accessions. AMF colonization has been shown to increase plant nutrient value by increasing the photosynthesis rate, which supports overall plant growth [48, 49]. Photosynthesis is the primary driver of biomass accumulation and provides the carbon skeleton for the biosynthesis of all metabolites. Improved photosynthesis increases sugar accumulation, directing more sugars to seeds and increasing sink strength [56]. High sugar accumulation under AMF growth conditions may supply carbon skeletons for other secondary metabolites, such as essential oils [57]. In this context, increased sugar content acts as a precursor for synthesizing different metabolites, such as phenolic compounds [58].

AMF also helps plants access water more efficiently, promote higher chlorophyll contents, and improve light capture and energy conversion, which supports their photosynthetic function [59]. Overall, the effects of AMF on photosynthesis and mineral uptake increase plant yield and increase the chemical composition and biological activity of plants. For example, the effects of AMF inoculation on essential oil composition, plant growth, and antioxidant and lipoxygenase activity were observed in *Piper aduncum* L. [50]. It has also been reported that aniseeds contain high amounts of total phenolics and flavonoids, which are the main contributors to their antioxidant and other bioactivities [60]. Moreover, AMF has been demonstrated to significantly increase a plant's biological activity [61, 62], facilitating nutrient exchange between the two partners, which offers several benefits, including the accumulation of bioactive metabolites.

The findings highlighted the hierarchical importance of variables, with biological activity being the most critical, followed by metabolic compounds, total nutrients, oil compounds, and physical properties. This hierarchy should guide future feature selection and model optimization to improve classification accuracy. Integrating multiple variable contexts, particularly biological activity, enhances model performance. Individual ROC and AUC results for each fold are depicted in Fig. 6. The physical properties had the lowest average performance, although fold 9 achieved an AUC of 100%, indicating inconsistent patterns in this variable set. The best results were obtained for the biological activity variables, which were only slightly lower than those obtained when all variables were used, underscoring their crucial role in improving classification accuracy.

Agronomically, the ability of various groups to differentiate between treatments inoculated with AMF and those not inoculated with AMF correlates with the influence of mycorrhizal associations. Biological activity variables are most impacted by this interaction, followed by metabolic compounds, total nutrients, and oil compounds, whereas physical properties remain unaffected. PCA components also explain variance in physical properties, nutrients, metabolic compounds, oil compounds, and biological activity (Table 3). PCA1 and PCA2 capture 59.19% to 97.68% of the variance, reducing dimensionality while retaining key information for further analysis [31]. This approach simplifies the feature space, improving the interpretability and efficiency of studies by identifying critical factors, such as the country of origin and seed maturation stage, in shaping the biochemical profiles of *P. anisum*.

To further emphasize the significance of our findings, it is essential to recognize that the enhanced understanding of the effects of AMF inoculation on the physicochemical composition of *P. anisum* offers valuable insights for agricultural practices. These results indicate that AMF can improve essential oil accumulation and nutrient content, which is critical for the commercial and medicinal applications of anise. These findings suggest that incorporating AMF into cultivation strategies could lead to increased crop yields, improved quality of essential oils, and increased nutritional value. These practices contribute to agricultural sustainability and align with organic farming principles that prioritize natural soil health and microbial activity. Furthermore, the ecological implications of our study are noteworthy; promoting mycorrhizal associations might enhance soil health and nutrient cycling within agroecosystems [63, 64]. By fostering beneficial microbial relationships, farmers can achieve a more resilient agricultural system that reduces chemical fertilizer and pesticide dependency.

In summary, our findings advance the scientific understanding of AMF interactions with *P. anisum* and provide actionable insights for enhancing cultivation practices that could lead to sustainable agricultural development and improved food quality.

Conclusion

Integrating the RF algorithm with dimensionality reduction techniques such as PCA can improve our understanding of plant–microorganism interactions, particularly with respect to AMF inoculation in *P. anisum* L. accessions. This approach underscores the importance of considering seed origin in optimizing symbiotic associations and offers valuable insights for enhancing sustainable agricultural practices. Future research can build upon this foundation by exploring advanced machine learning methods to further unravel the complexities of plant–microorganism dynamics across diverse agricultural systems.

Acknowledgements
The authors extend their appreciation to the Researchers Supporting Project number (RSPD2025R931) King Saud University, Riyadh, Saud Arabia.

Clinical Trial Number
Not applicable.

Authors' contributions
Bruno Rodrigues de Oliveira, Alan Mario Zuffo, Francisco Charles dos Santos Silva, Fábio Steiner: conceptualization, methology, investigation and writing; Amal Mohamed AlGarawi, Mohammad K. Okla, Mousa nhs, Yousef Alhaj Hamoud: methology, investigation; Izabela Josko, Mohamed S. Sheteiwy, Mohamed Salem Alyafei, Saad Sulieman: Writing the first draft of the paper—review & editing the final version of this manuscript. Bruno Rodrigues de Oliveira, Alan Mario Zuffo: supervision, conceptualization, writing — review & editing, methology, resources, funding acquisition.

Funding
The authors extend their appreciation to the Researchers Supporting Project number (RSPD2025R931) King Saud University, Riyadh, Saud Arabia. King Saud University.

Declarations

Ethics approval and consent to participate
We declare that the manuscript reporting studies do not involve any human participants, human data, or human tissue. Therefore, it is not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Editora Pantanal, Nova Xavantina, MT, Brazil. [2]Department of Agronomy, State University of Maranhão, Balsas, MA, Brazil. [3]Department of Agronomy, State University of Mato Grosso Do Sul, Cassilândia, MS, Brazil. [4]Botany and Microbiology Department, College of Science, King Saud University, P.O. Box 2455, 11451 Riyadh, Saudi Arabia. [5]Botany & Microbiology Department, Faculty of Science, Assiut University, P.O. Box 71516, Assiut, Egypt. [6]The National Key Laboratory of Water Disaster Prevention, and College of Hydrology and Water Resources, Hohai University, Nanjing 210098, China. [7]Research Centre for Horticultural Crops (FGK), Fachhochschule Erfurt, 99090 Erfurt, Germany. [8]Institute of Plant Genetics, Breeding and Biotechnology, Faculty of Agrobioengineering, University of Life Sciences, 20-950 Lublin, Poland. [9]Department of Integrative Agriculture, College of Agriculture and Veterinary Medicine, United Arab Emirates University, P.O. Box 15551, Al Ain, Abu Dhabi, United Arab Emirates. [10]Department of Agronomy, Faculty of Agriculture, University of Khartoum, Khartoum North 13314, Shambat, Sudan.

Received: 29 September 2024 Accepted: 10 April 2025
Published online: 09 May 2025

References
1. Mehravi S, Hanifei M, Gholizadeh A, Khodadadi M. Water deficit stress changes in physiological, biochemical, and antioxidant characteristics of anise (*Pimpinella anisum* L.). Plant Physiol Biochem. 2023;201:107806. https://doi.org/10.1016/j.plaphy.2023.107806.
2. Bettaieb Rebey I, Bourgou S, Ben Kaab S, Aidi Wannes W, Ksouri R, Saidani Tounsi M, Fauconnier M-L. On the effect of initial drying techniques on essential oil composition, phenolic compound, and antioxidant properties of anise (*Pimpinella anisum* L.) seeds. J Food Measurement Character. 2020;14:220–8. https://doi.org/10.1007/s11694-019-00284-4.
3. Shojaii A, Abdollahi FM. Review of pharmacological properties and chemical constituents of *Pimpinella anisum*. International Scholarly Research Notices. 2012;2012: 510795. https://doi.org/10.5402/2012/510795.
4. Özcan MM, Chalchat JC. Chemical composition and antifungal effect of anise (*Pimpinella anisum* L.) fruit oil at ripening stage. Annals of Microbiology. 2006;56:353–8. https://doi.org/10.1007/BF03175031.
5. De Feo V, Senatore F. Medicinal plants and phytotherapy in the Amalfitan Coast, Salerno Province, Campania, Southern Italy. J Ethnopharmacol. 1993;39:39–51. https://doi.org/10.1016/0378-8741(93)90049-B.
6. Wang Q, Liu M, Wang Z, Li J, Liu K, Huang D. The role of arbuscular mycorrhizal symbiosis in plant abiotic stress. Front Microbiol. 2024;14. https://doi.org/10.3389/fmicb.2023.1323881
7. Hozzein WN, Abuelsoud W, Wadaan MA, Shuikan AM, Selim S, Al Jaouni S, AbdElgawad H. Exploring the potential of actinomycetes in improving soil fertility and grain quality of economically important cereals. Sci Total Environ. 2019;651:2787–98.
8. Sun W, Shahrajabian MH. The application of arbuscular mycorrhizal fungi as a microbial biostimulant, sustainable approaches in modern agriculture. Plants. 2023;12:3101. https://doi.org/10.3390/plants12203101.
9. Khaliq A, Perveen S, Alamer KH, Zia Ul Haq M, Rafique Z, Alsudays IM, Althobaiti AT, Saleh MA, Hussain S, Attia H. Arbuscular mycorrhizal fungi symbiosis to enhance plant–soil Interaction. Sustain. 2022;14:7840. https://doi.org/10.3390/su14137840
10. Etesami H, Jeong BR, Glick BR. Contribution of arbuscular mycorrhizal fungi, phosphate-solubilizing bacteria, and silicon to P uptake by plants. Front Plant Sci. 2021;12. https://doi.org/10.3389/fpls.2021.699618
11. Ebrahimi Z, Fateh E, Soorestani MM, Ghorbanpour M. Inoculating mycorrhiza fungi and growth-promoting bacteria and mulch of plant residues improve yield and essential oil production of anise (*Pimpinella anisum* L.). Acta Scientiarum Polonorum Hortorum Cultus. 2023;22:55–67. https://doi.org/10.24326/asphc.2023.5162.
12. Gholami Ganje S, Salehi A, Shahrasbi S, Amiri Nasab AM. Evaluation the effect of vermicompost different levels and mycorrhiza fungi on yield and essential oil production of Anise (*Pimpinella anisum* L.). Iranian J Horticult Sci. 2019;50:325–35. https://doi.org/10.22059/ijhs.2018.243998.1335.
13. Masomi Zavarian AF, Yousefirad M, Asghari M. Effects of mycorrhizal fungi on quantitative and qualitative characteristics of Anise plant (*Pimpinella anisum*) under salt stress. J Med Plants Res. 2015;14:139–48. https://dor.isc.ac/dor/20.1001.1.2717204.2015.14.56.3.0.

14. Zand A, Darzi MT, Hadi MRHS. Effects of phosphate solubilizing microorganisms and plant density on seed yield and essential oil content of anise (*Pimpinella anisum*). Middle-East J Sci Res. 2013;14:940–6. https://doi.org/10.5829/idosi.mejsr.2013.14.3.2050.
15. Balkhyour MA, Hassan AHA, Halawani RF, Summan AS, AbdElgawad H. Effect of elevated CO_2 on seed yield, essential oil metabolism, nutritive value, and biological activity of *Pimpinella anisum* L. accessions at different seed maturity stages. Biol. 2021;10:979. https://doi.org/10.3390/biology10100979.
16. Okla MK, Abdel-Mawgoud M, Alamri SA, Abbas ZK, Al-Qahtani WH, Al-Qahtani SM, Al-Harbi NA, Hassan AHA, Selim S, Alruhaili MH, et al. Developmental stages-specific response of anise plants to laser-induced growth, nutrients accumulation, and essential oil metabolism. Plants. 2021;10:2591. https://doi.org/10.3390/plants10112591.
17. Zehtab-Salmasi S, Javanshir A, Omidbaigi R, Alyari H, Ghassemi-Golezani K. Effects of water supply and sowing date on performance and essential oil production of anise (*Pimpinella anisum* L.). Acta Agronom Hungarica. 2001;49:75–81. https://doi.org/10.1556/AAgr.49.2001.1.10
18. Liu M, Zheng R, Bai S, Bai Ye, Wang J. Slope aspect influences arbuscular mycorrhizal fungus communities in arid ecosystems of the Daqingshan Mountains, Inner Mongolia, North China. Mycorrhiza. 2017;27:189–200. https://doi.org/10.1007/s00572-016-0739-7
19. Yousef REM, Khaul AM. Response of *Pimpinella anisum* to organic fertilizer treatments under new lands conditions. Egyptian Journal of Agricultural Research. 2015;93:1–15.
20. Sallam D. Effect of chemical fertilization, biofertilizer and thidiazuron on growth and yield of anise (*Pimpinella anisum* L.). Al-Azhar J Pharmaceut Sci. 2015;52:14–37. https://doi.org/10.21608/ajps.2015.12530
21. Wang P, Wang Y, Wu QS. Effects of soil tillage and planting grass on arbuscular mycorrhizal fungal propagules and soil properties in citrus orchards in southeast China. Soil and Tillage Research. 2016;155:54–61. https://doi.org/10.1016/j.still.2015.07.009.
22. Zhang S, Li Q, Lü Y, Sun X, Jia S, Zhang X, Liang W. Conservation tillage positively influences the microflora and microfauna in the black soil of Northeast China. Soil and Tillage Research. 2015;149:46–52. https://doi.org/10.1016/j.still.2015.01.001.
23. Bettaieb Rebey I, Aidi Wannes W, Kaab SB, Bourgou S, Tounsi MS, Ksouri R, Fauconnier ML. Bioactive compounds and antioxidant activity of *Pimpinella anisum* L. accessions at different ripening stages. Scientia Horticul. 2019;246:453–61. https://doi.org/10.1016/j.scienta.2018.11.016.
24. Bettaieb Rebey I, Bourgou S, Aidi Wannes W, Hamrouni Selami I, Saidani Tounsi M, Marzouk B, Fauconnier ML, Ksouri R. Comparative assessment of phytochemical profiles and antioxidant properties of Tunisian and Egyptian anise (*Pimpinella anisum* L.) seeds. Plant Biosyst. 2018;152:971–8. https://doi.org/10.1080/11263504.2017.1403394
25. Ling B, Goodin DG, Raynor EJ, Joern A. Hyperspectral analysis of leaf pigments and nutritional elements in tallgrass prairie vegetation. Front Plant Sci. 2019;10:142. https://doi.org/10.3389/fpls.2019.00142.
26. Belgiu M, Drăguţ L. Random forest in remote sensing: A review of applications and future directions. ISPRS J Photogramm Remote Sens. 2016;114:24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011.
27. Du P, Samat A, Waske B, Liu S, Li Z. Random forest and rotation forest for fully polarized SAR image classification using polarimetric and spatial features. ISPRS J Photogramm Remote Sens. 2015;105:38–53. https://doi.org/10.1016/j.isprsjprs.2015.03.002.
28. Rodriguez-Galiano VF, Ghimire B, Rogan J, Chica-Olmo M, Rigol-Sanchez JP. An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS J Photogramm Remote Sens. 2012;67:93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002.
29. Breiman L. Random forests. Mach Learn. 2001;45:5–32. https://doi.org/10.1023/A:1010933404324.
30. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374:20150202. https://doi.org/10.1098/rsta.2015.0202.
31. Salih Hasan BM, Abdulazeez AM. A review of principal component analysis algorithm for dimensionality reduction. J Soft Comput Data Mining. 2021;2:20–30. https://doi.org/10.30880/jscdm.2021.02.01.003
32. Mahmoud AM, Reyad AM, Khalaf MH, Sheteiwy MS, Dawood MFA, El-Sawah AM, et al. Investigating the endophyte *Actinomycetota* sp. JW0824 strain as a potential bioinoculant to enhance the yield, nutritive value, and chemical composition of different cultivars of anise (*Pimpinella anisum* L.) seeds. Biol. 2024, 13:553. https://doi.org/10.3390/biology13080553
33. Kubat M. *An introduction to machine learning*. Springer; 2017.
34. Raschka S. Model evaluation, model selection, and algorithm selection in machine learning. *ArXiv* 2018, abs/1811.12808
35. Tin KH. The random subspace method for constructing decision forests. IEEE Trans Pattern Anal Mach Intell. 1998;20:832–44. https://doi.org/10.1109/34.709601.
36. Shaik AB, Srinivasan S. A brief survey on random forest ensembles in classification model. In: *Proceedings of the International Conference on Innovative Computing and Communications*; Singapore. 2019; p. 253–60.
37. Louppe G, Wehenkel L, Sutera A, Geurts P. Understanding variable importances in forests of randomized trees. Adv Neural Inf Process Syst. 2013, 26.
38. Carvalho LO, Ribeiro DB, Monteiro FAC. Comparing artificial neural networks with support vector machines for soil classification. In: *Proceedings of the XL Ibero-Latin American Congress on Computational Methods in Engineering*; Natal, Brazil. 2019.
39. Nahm JM, Miller SM, Urpelainen J. G20's US $14-trillion economic stimulus reneges on emissions pledges. Nature. 2022;603:28–31.
40. Lee LC, Jemain AA. On overview of PCA application strategy in processing high dimensionality forensic data. Microchem J. 2021;169: 106608. https://doi.org/10.1016/j.microc.2021.106608.
41. Nguyen LH, Holmes S. Ten quick tips for effective dimensionality reduction. PLoS Comput Biol. 2019;15(6):e1006907. https://doi.org/10.1371/journal.pcbi.1006907.
42. Govindarajan S, Ardila-Rey JA, Krithivasan K, Subbaiah J, Sannidhi N, Balasubramanian M. Development of hypergraph-based improved random forest algorithm for partial discharge pattern classification. IEEE Access. 2021;9:96–109. https://doi.org/10.1109/ACCESS.2020.3047125.
43. Marques Ramos AP, Prado Osco L, Elis Garcia Furuya D, Nunes Gonçalves W, Cordeiro Santana D, et al. A random forest ranking approach to predict yield in maize with UAV-based vegetation spectral indices. Comput Electronics Agricult. 2020;178:105791. https://doi.org/10.1016/j.compag.2020.105791.
44. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. The Journal of Machine Learning Research. 2011;12:2825–30.
45. The Jamovi Project. Jamovi (Versão 2.5) [Software de Computador]. Recuperado de https://www.jamovi.org. 2024.
46. Bisong E. Google Colaboratory. In: Building machine learning and deep learning models on Google Cloud Platform: A comprehensive guide for beginners. Berkeley, CA: Apress; 2019. p. 59–64. https://doi.org/10.1007/978-1-4842-4470-8
47. Acimovic MG, Korac J, Jacimovic G, Oljaca S, Djukanovic L, Vuga-Janjatov V. Influence of ecological conditions on seeds traits and essential oil contents in anise (*Pimpinella anisum* L.). Notulae Botanicae Horti Agrobotanici Cluj-Napoca. 2014;42:232–8. https://doi.org/10.15835/nbha4219492
48. Neumann E, Schmid B, Römheld V, George E. Extraradical development and contribution to plant performance of an arbuscular mycorrhizal symbiosis exposed to complete or partial root-zone drying. Mycorrhiza. 2009;20:13–23. https://doi.org/10.1007/s00572-009-0259-9.
49. Tanaka A, Fujita K. Growth, photosynthesis and yield components in relation to grain yield of the field bean. Journal of the Faculty of Agriculture. 1979;59:145–238.
50. de Oliveira JSF, Xavier LP, Lins A, Andrade EH, Maia JGS, de Mello AH, et al. Effects of inoculation by arbuscular mycorrhizal fungi on the composition of the essential oil, plant growth, and lipoxygenase activity of *Piper aduncum* L. AMB Express. 2019;9:29. https://doi.org/10.1186/s13568-019-0756-y.
51. Mehravi S, Ranjbar GA, Mirzaghaderi G, Severn-Ellis AA, Scheben A, Edwards D, Batley J. De Novo SNP discovery and genotyping of Iranian *Pimpinella* species using ddRAD sequencing. Agronomy. 2021;11:1342. https://doi.org/10.3390/agronomy11071342.
52. Maeda H, Dudareva N. The shikimate pathway and aromatic amino acid biosynthesis in plants. Annu Rev Plant Biol. 2012;63:73–105. https://doi.org/10.1146/annurev-arplant-042811-105439.

53. Aćimović M, Lončar B, Pezo M, Stanković Jeremić J, Cvetković M, Rat M, Pezo L. Volatile compounds of *Nepeta nuda* L. from Rtanj Mountain (Serbia). *Horticulturae*. 2022;8:85. https://doi.org/10.3390/horticulturae8020085
54. Enebe MC, Erasmus M. Symbiosis—a perspective on the effects of host traits and environmental parameters in arbuscular mycorrhizal fungal richness, colonization, and ecological functions. Agriculture. 2023;13:1899. https://doi.org/10.3390/agriculture13101899.
55. Cobb AB, Wilson GWT, Goad CL, Bean SR, Kaufman RC, Herald TJ, Wilson JD. The role of arbuscular mycorrhizal fungi in grain production and nutrition of sorghum genotypes: enhancing sustainability through plant-microbial partnership. Agr Ecosyst Environ. 2016;233:432–40. https://doi.org/10.1016/j.agee.2016.09.024.
56. Thompson M, Gamage D, Hirotsu N, Martin A, Seneweera S. Effects of elevated carbon dioxide on photosynthesis and carbon partitioning: A perspective on root sugar sensing and hormonal crosstalk. Front Physiol. 2017;8:578. https://doi.org/10.3389/fphys.2017.00578.
57. Najar B, Zrig A, Alsherif EA, Selim S, Aloufi AS, Korany SM, Nhs M, Aldilam M, Bouqellah NA. Synergistic effect of arbuscular mycorrhizal fungi and germanium on the growth, nutritional quality, and health-promoting activities of *Spinacia oleracea* L. Plants. 2024;13(20):2869. https://doi.org/10.3390/plants13202869.
58. Al Jaouni, S.; Saleh, A.M.; Wadaan, M.A.M.; Hozzein, W.N.; Selim, S.; AbdElgawad, H. Elevated CO_2 induces a global metabolic change in basil (*Ocimum basilicum* L.) and peppermint (*Mentha piperita* L.) and improves their biological activity. J. Plant. Physiol. 201. https://doi.org/10.1016/j.jplph.2018.03.016
59. Amaya-Carpio, L.; Davies, F.T.; Fox, T.; HE, C. Arbuscular mycorrhizal fungi and organic fertilizer influence photosynthesis, root phosphatase activity, nutrition, and growth of *Ipomoea carnea* ssp. *Fistulosa*. Photosynthetica. 2009, 47(1):1–10. https://doi.org/10.1007/s11099-009-0003-x
60. Shehabeldine AM, Doghish AS, El-Dakroury WA, Hassanin MM, Al-Askar AA, AbdElgawad H, Hashem AH. Antimicrobial, antibiofilm, and anti-cancer activities of syzygium aromaticum essential oil nanoemulsion. Molecules. 2023;28(15):5812.
61. Sheteiwy MS, El-Sawah AM, Korany SM, Alsherif EA, Mowafy AM, Chen J, Jośko I, Selim S, AbdElgawad H. Arbuscular mycorrhizal fungus "*Rhizophagus irregularis*" impacts on physiological and biochemical responses of ryegrass and chickpea plants under beryllium stress. Environ Pollut. 2022;315: 120356.
62. Sheteiwy MS, El-Sawah AM, Kobae Y, Basit F, Holford P, Yang H, El-Keblawy A, Abdel-Fattah GG, Wang S, Araus JL, Korany SM, Alsherif EA, AbdElgawad H. The effects of microbial fertilizers application on growth, yield and some biochemical changes in the leaves and seeds of Guar (*Cyamopsis tetragonoloba* L.). Food Res Int. 2023;172:113122.
63. Hozzein WN, Abuelsoud W, Wadaan MA, Shuikan AM, Selim S, Al Jaouni S, AbdElgawad H. Exploring the potential of actinomycetes in improving soil fertility and grain quality of economically important cereals. Sci Total Environ. 2019;651:2787–98.
64. Versieren L, Evers S, AbdElgawad H, Asard H, Smolders E. Mixture toxicity of copper, cadmium, and zinc to barley seedlings is not explained by antioxidant and oxidative stress biomarkers. Environ Toxicol Chem. 2017;36(1):220–30.

Publisher's Note