

Integrating Machine Learning and SHAP Analysis to Boost Callus Growth and Rutin Biosynthesis in *Capparis spinosa* L.

Marouane Mohaddab

mmohaddab@uliege.be

Universite de Liege Gembloux Agro-Bio Tech <https://orcid.org/0000-0002-7791-368X>

Younes EL Goumi

Sultan Moulay Slimane University Polydisciplinary Faculty of Beni Mellal: Universite Sultan Moulay Slimane Faculte Polydisciplinaire de Beni Mellal

Mohammed Elakrouh

Universite de Liege Gembloux Agro-Bio Tech

Soufiane Hasni

Abdelmalek Essaadi University Faculty of Science and Technology of Tangier: Universite Abdelmalek Essaadi Faculte des Sciences et Techniques de Tanger

Clément Burgeon

Universite de Liege Gembloux Agro-Bio Tech

Manon Genva

Universite de Liege Gembloux Agro-Bio Tech

Malika Fakiri

Universite Hassan 1er Faculte des Sciences et Techniques Settlat

Marie-Laure Fauconnier

Universite de Liege Gembloux Agro-Bio Tech

Research Article

Keywords: *Capparis spinosa* L., Plant tissue culture, Callogenesis, Machine Learning, Rutin

Posted Date: January 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8619922/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Plant Cell, Tissue and Organ Culture (PCTOC) on June 15th, 2026. See the published version at <https://doi.org/10.1007/s11240-026-03509-9>.

Abstract

Capparis spinosa L. is a Mediterranean medicinal species of high economic value, yet its large-scale propagation and metabolite production remain constrained by conventional approaches. A full factorial design was used to evaluate the effects of four plant growth regulators, 6-benzylaminopurine, kinetin, 2,4-dichlorophenoxyacetic acid, and 1-naphthaleneacetic acid, on fresh weight gain from leaf explants. Data from twenty hormonal treatments were modeled using four machine learning algorithms: Random Forest, Gradient Boosting, Extreme Gradient Boosting, and second-degree polynomial regression. Random Forest provided the highest predictive accuracy. SHapley Additive exPlanations analysis identified 2,4-dichlorophenoxyacetic acid as the dominant factor driving callogenesis, with 6-benzylaminopurine exerting a secondary synergistic effect, whereas kinetin and 1-naphthaleneacetic acid showed minimal or inhibitory influence. Experimental validation confirmed the five best Random Forest–predicted hormonal combinations, including the optimal mixture of 6-benzylaminopurine and 2,4-dichlorophenoxyacetic acid, which produced the highest increase in callus fresh weight gain. Rutin, the major bioactive flavonoid of *C. spinosa*, was identified by LC-QTOF-MS/MS tandem mass spectrometry and semi-quantified by LC-TQ-MS/MS under 6-benzylaminopurine and 2,4-dichlorophenoxyacetic acid combinations. A stacked Random Forest model integrating fresh weight gain predictions successfully estimated rutin accumulation, with maximal production at moderate hormone levels. This integrative machine learning and SHapley Additive exPlanations framework offers an interpretable and scalable strategy for optimizing callus culture and enhancing high-value metabolite production in *C. spinosa*. Moreover, callus culture represents a promising and sustainable alternative for large-scale production of valuable metabolites, reducing reliance on wild plant resources.

Introduction

Capparis spinosa L., an endemic species of arid Mediterranean regions, has attracted increasing scientific attention due to its medicinal, nutritional, and agroecological values. The species is extensively harvested and commercially exploited across several Mediterranean countries, including Morocco, Spain, Italy, Greece, and Turkey (Sozzi and Vicente 2006; Grimalt et al. 2018, 2022; Hazrati et al. 2025). *C. spinosa* is characterized by a rich phytochemical profile, encompassing glucosinolates, phenolic acids, alkaloids, and flavonoids, which are associated with a wide range of biological activities such as antioxidants, antidiabetic, anti-inflammatory, and anticancer effects (Moghadamnia et al. 2019; Esmaeilpour et al. 2020; Rahimi et al. 2020). These properties highlight its considerable potential for applications in the pharmaceutical, nutraceutical, and cosmetic industries. (Karous et al. 2021; Kirkan et al. 2021; Annaz et al. 2022). Among these compounds, rutin stands out for its high economic and therapeutic value, making it a compound of particular interest (Fotiadou et al. 2025; El Amri et al. 2025). The biosynthesis of rutin involves branched metabolic pathways from phenylalanine and is mediated by phenylalanine ammonia-lyase (PAL) and other key enzymes (Kianersi et al. 2020).

The exploitation of this plant supports the livelihoods of rural and marginalized communities while contributing to local economies (Sozzi et al. 2012). However, commercial harvesting relies mainly on

wild populations, increasing pressure on natural resources and threatening long-term sustainability (Zhang and Ma 2018). Due to its tolerance to water deficit and adverse soils, *C. spinosa* also represents an ideal candidate for climate-resilient agriculture (Yousefi et al. 2025). Despite these advantages, traditional propagation methods, whether sexual or vegetative, face significant limitations, including low efficiency, slow growth, and the potential loss of desirable traits (Gristina et al. 2014). Pretreatments such as scarification, sulfuric acid exposure, or hormone application partially improve germination and rooting but remain insufficient for rapid large-scale multiplication (Labbafi et al. 2018; Nowruzian and Aalami 2023).

In vitro culture techniques, particularly callus induction, offer an effective solution (Efferth 2019). They enable the regeneration of uniform, pathogen-free plants and may constitute effective reservoirs for the biosynthesis of secondary metabolites, thereby contributing to the conservation of natural populations and meeting the growing demand for phytotherapeutic resources (Yin et al. 2014; Mohaddab et al. 2022). Although research on *C. spinosa* remains limited (Caglar et al. 2005; Movafeghi et al. 2008; Al-Safadi and Elias 2011; Carra et al. 2012; Fahmideh et al. 2019; Gianguzzi et al. 2019; Mohaddab et al. 2024), callus tissue, being totipotent, serves as an ideal platform for targeted secondary metabolite production. Hagaggi et al. (2024) reported that callus extracts contain higher levels of phenolic and flavonoid compounds, a greater diversity of volatile metabolites, and superior antioxidant and antibacterial activities compared to leaf extracts. Thus, strict control of culture conditions enhances the biosynthesis of bioactive flavonoids, offering direct potential for industrial and pharmaceutical applications (Wijerathna-Yapa et al. 2025).

Callus formation and proliferation depend on intrinsic factors, including explant type, age, and medium composition, as well as extrinsic factors such as light, temperature, and humidity (Pasternak and Steinmacher 2024). Plant growth regulators (PGRs), particularly auxins and cytokinins, are essential for morphogenesis, biomass accumulation, and secondary metabolite production (Ikeuchi et al. 2013; Bravo-Vázquez et al. 2023). Commonly used phytohormones include 2,4-dichlorophenoxyacetic acid (2,4-D) and 1-naphthaleneacetic acid (NAA), applied alone or in combination with cytokinins such as 6-benzylaminopurine (BAP) or kinetin (KIN) (Koufan et al. 2022; Higazy et al. 2023). Studies indicate that 2,4-D at 1.0 mg/L combined with BAP at 1.5 mg/L effectively induces callus from *C. spinosa* leaf explants (Wang et al. 2007; Liu et al. 2011; Yin et al. 2014). Strategic manipulation of PGRs during callogenesis can therefore enhance both callus regeneration and flavonoid production, particularly rutin (Sobhy et al. 2025).

In vitro plant development is a complex process governed by multiple interconnected variables (Ma et al. 2018; Svolacchia and Sabatini 2023). Studying these factors in isolation is insufficient, as their interactions strongly influence outcomes. Classical statistical approaches are limited when handling nonlinear and multidimensional datasets and often exhibit high error rates (Aasim et al. 2023; Katirci et al. 2024). Mechanistic models, although grounded in biological knowledge, require precise measurement of numerous parameters that are difficult to standardize experimentally (Noordijk et al. 2024), reducing their predictive power and practical usefulness for rapid protocol optimization (Baker et al. 2018).

Artificial intelligence (AI), especially supervised Learning or Machine Learning (ML), offers a means to manage this complexity (Peng and Rajjou 2024). ML applications to callogenesis and organogenesis represent an emerging and promising field (Hesami and Jones 2021). Algorithms such as Random Forest (RF), Gradient Boosting (GB), and Extreme Gradient Boosting (XGBoost) have been successfully applied to predict callus induction, shoot regeneration, biomass yield, and secondary metabolite production in various species (Hesami and Jones 2021; Alcalde et al. 2022; Eren et al. 2023; Demirel et al. 2023; Sarabandi et al. 2024; Bozkurt et al. 2024; Tan et al. 2024; Zarbakhsh et al. 2024; Koçak et al. 2025; Sarmah et al. 2025). Despite their predictive performance, these models often remain “black boxes,” where relationships between biological variables and observed responses are opaque (Petch et al. 2022). SHapley Additive exPlanations (SHAP) values provide a solution to this opacity. Derived from game theory, they quantify the contribution of each variable to the final prediction by considering all possible factor combinations (Lundberg and Lee 2017; Lundberg et al. 2020). By transforming a black-box model into a “white-box” model, SHAP values allow identification of key variables and their interactions, offering biologically meaningful interpretation (Ekanayake et al. 2022).

Although interest in *C. spinosa* is increasing, investigations into callus culture and flavonoid biosynthesis, particularly rutin, remain limited and fragmented. The majority of studies have focused on isolated factors rather than the complex, non-linear interactions among growth regulators, explant type, and environmental conditions (Dar et al. 2021). Furthermore, while machine learning approaches have demonstrated the capacity to drive diverse tissue culture systems, their application to *C. spinosa*, especially using interpretable methods such as SHAP, has yet to be explored, leaving critical variables governing flavonoid production largely unexplored.

This study examined the effects of growth regulators, individually and in combination, on callus induction and proliferation, as well as on rutin biosynthesis. The collected data were evaluated using machine learning models, complemented by SHAP analyses, to quantify the relative influence of each regulator and identify the key factors modulating both callus development and rutin biosynthesis.

Materials and methods

Collection of Plant Material

Fresh leaves of *C. spinosa* L. were collected in April 2022 from the Taounate region (Morocco; latitude: 34°26'31.187" N, longitude: 4°42'11.97" W, altitude: 435 m). The leaves were harvested on the same day they were processed for further experiments.

Surface Sterilization

Harvested leaves were surface sterilized prior to culture. Sterilization was performed following the protocol described by Slimani et al. (2021), ensuring that explants were free from microbial contamination before placement on culture medium.

Culture Medium Preparation

Callus induction was carried out on Murashige and Skoog (MS) medium (Murashige and Skoog 1962), supplemented with 0.8% agar, 3% sucrose, and MES buffer (M0255, Duchefa). Leaf explants, approximately 1 cm in diameter, were placed abaxial side down on 20 mL of medium in each Petri dish, with five explants per dish. A total of 1,000 explants were used in the experiment. The pH of the medium was adjusted to 5.7 ± 0.1 before autoclaving at 121°C for 20 minutes under 1 bar of pressure.

Plant Growth Regulator Combination Design

The basal medium was supplemented with different concentrations of exogenous plant growth regulators (PGRs) to evaluate their effects on callus induction. The PGRs tested included 2,4-dichlorophenoxyacetic acid (2,4-D: 0.0–2.0 mg/L), 6-benzylaminopurine (BAP: 0.0–2.0 mg/L), 1-naphthaleneacetic acid (NAA: 0.0–2.0 mg/L), and kinetin (KIN: 0.0–2.0 mg/L). A full factorial design was applied with four PGR combinations: BAP/NAA, BAP/2,4-D, KIN/NAA, and KIN/2,4-D. For each combination, four concentrations (0.0–1.5 mg/L) were applied per regulator, resulting in 16 unique concentration pairs per combination. Each PGR was also tested individually at 2.0 mg/L, for a total of 20 distinct treatments, each replicated three times.

In vitro culture Conditions

All cultures were maintained in a phytotron under a 16-hour light/8-hour dark photoperiod at $24 \pm 2^{\circ}\text{C}$ for 30 days. Illumination was provided by Sylvania fluorescent tubes delivering a light intensity of approximately 1600 W.

Measurement of Callus Growth

After 30 days, callus Fresh Weight Gain (FWG) was measured by weighing callus immediately after removal from the medium. These data were then compiled into a dataset for machine Learning analysis.

FWG was calculated as: $\text{FWG} = \text{Final fresh weight} - \text{Initial weight of explant}$

Identification and semi-quantification of rutin in callus Cultures using LC–MS/MS

Approximately 150 mg of bone ash, dried at 30°C to constant weight, was extracted according to the protocol of Belkessam et al. (2025), with minor modifications, using 2 mL of methanol–water (80:20, v/v) and 5 μL of flavone as an internal standard. The samples were sonicated for 20 minutes, centrifuged for 10 minutes, and then filtered through 0.22 μm filters before analysis by LC–MS/MS. Rutin was qualitatively identified by LC-QTOF and semi-quantified by LC-TQ via the MRM transition $609 \rightarrow 300$, in accordance with the original method. Acquisition and processing of the data were carried out with MassHunter® Workstation software (version 10.0, Agilent Technologies, Santa Clara, CA, USA).

$$\text{Molecule content (A. U. /mg DW)} = \frac{\text{MRM peak area of rutin}}{\text{MRM flavone peak area} \times \text{Sample dry weight}}$$

A.U. = Arbitrary Units; DW = Dry Weight ; MRM = Multiple Reaction Monitoring.

Modeling procedures

The study methodology is presented in figure. 1. Callogenesis was induced using four PGR combinations: BAP/NAA, BAP/2,4-D, KIN/NAA, and KIN/2,4-D at varying concentrations to evaluate their effects on FWG and rutin accumulation. To ensure data quality, the data were normalized, outliers were removed, and the relationships between parameters were examined using a Pearson correlation test. The final dataset was partitioned into a training set (75%) and an independent test set (25%), while maintaining a balanced distribution of experimental conditions (Figure. 1 (A)).

Figure 1A schematic representation of the stepwise computational approach used in this study. (A) Modeling and prediction of callus Fresh Weight Gain (FWG) (output) based on 6-Benzylaminopurine (BAP), Kinetin (KIN), Naphthalene Acetic Acid (NAA), and 2,4-Dichlorophenoxyacetic Acid (2,4-D) as inputs using four models. (B) SHAP value analysis of the best-selected model to identify and quantify the key model features. (C) Stacking of the best model and using it as input for the production and accumulation of rutin (Model 2).

Four machine learning models, RF, GB, XGBoost, and second-degree polynomial regression (PL2), were evaluated for FWG prediction. Model performance was assessed using established statistical indicators: the coefficient of determination (R^2) and cross-validation (R^2_{CV}), moreover root mean square error (RMSE). Low RMSE values indicate that the model's predictions are reliable, with minimal error. R^2 quantifies the proportion of the total variance in the observed data captured by the model, while R^2_{CV} assesses its ability to generalize to new data. Both range from 0 to 1, with values closer to 1 reflecting strong agreement between predicted and observed values and good model stability. Overall, a model with high R^2 and R^2_{CV} combined with low RMSE can accurately and reliably reproduce the behavior of the studied system.

SHAP effectively serves as a sensitivity analysis (Fig. 1(B)), measuring the impact of each growth regulator on FWG while accounting for the combined effects of all other variables. This approach allows the identification of regulators with the greatest influence on callus growth and provides insight into the complex interactions among them.

Predictions from model 1 were then incorporated into a stacked model designed to predict rutin accumulation (model 2) (Fig. 1(C)). This model used both the initial concentrations of the growth regulators and the FWG predictions as input variables, directly linking callus growth to secondary metabolite production. To validate the efficacy and reliability of this optimized model, the predicted conditions for maximizing FWG and rutin accumulation were experimentally tested in triplicate. The results confirmed the accuracy of the predictions and demonstrated the practical utility of the model for optimizing callogenesis and secondary metabolite production.

Results

Influence of plant growth regulators on *C. Spinosa* callus Induction

This study investigated the influence of phytohormones matrix at different concentrations (0, 0.5, 1 and 1.5 mg/L) and hormonal combinations (BAP/NAA), (BAP/2,4-D), (KIN/NAA), and (KIN/2,4-D) on the FWG of callus derived from *C. spinosa* leaf explants after 30 days of culture (Table A). Notably, no callus formation occurred in the culture medium lacking PGRs, highlighting the essential role of exogenous hormones in inducing callogenesis in *C. spinosa* leaf explants.

Among the tested hormone combinations, 2,4-D and BAP showed an effect on callus induction and growth. The highest FWG (3639.5 mg ± 208.0) was recorded with a combination of 1 mg/L BAP with 1 mg/L 2,4-D. Similarly high values were observed with 1 mg/L BAP + 0.5 mg/L 2,4-D (3459.33 mg ± 84.68). Additionally, 2,4-D applied alone also promoted substantial callus proliferation, with FWG values of 3438.5 mg ± 493.5 and 3394 mg ± 31 at 0.5 and 1 mg/L, respectively.

In contrast, the (BAP/ NAA) and (KIN/ NAA) combinations produced significantly lower FWG, indicating less efficiency in promoting callus formation and proliferation. The maximum FWG in these groups was observed with 0.5 mg/L BAP combined with 1.5 NAA (1874.45 mg ± 25.66) and with 0.5 mg/L KIN combined with 1 mg/L NAA (1350.5 mg ± 215.5). To further investigate these complex hormonal interactions and their effects on callus induction, artificial intelligence was employed to analyze the dataset.

Model performance in predicting callus induction

The callogenesis of *C. spinosa* was quantitatively modeled using a set of explanatory variables representing four exogenously applied PGRs: BAP, KN, NAA, and 2,4-D. The target variable for modeling was the FWG of callus tissue. To capture the complex relationships between hormonal treatments (inputs) and FWG (outputs), four ML models were applied and compared: RF, GB, XGBoost, and a traditional second-degree Polynomial Regression.

Table 1
Machine learning model performance for predicting callus fresh weight gain.

ML Models	R ²	R ² CV	RMSE
RF	0.86	0.89	263.33
GB	0.83	0.86	285.59
XGBoost	0.85	0.86	274.46
PL2	0.56	0.63	471.98

R²: coefficient of determination, R²CV: cross-validated R², RMSE: root mean square error, RF: Random Forest, GB: Gradient Boosting, XGBoost: Extreme Gradient Boosting, PL2: second-degree polynomial regression.

The ensemble Learning models, RF, GB, and XGBoost, demonstrated superior and comparable performance (Table 1). RF had the highest predictive accuracy with an R^2 of 0.86, R^2CV of 0.89, and a low RMSE of 263.33, indicating precise predictions and strong generalization. XGBoost and GB also performed well, with R^2 values of 0.85 and 0.83, R^2CV of 0.86 for both, and RMSE values of 274.46 and 285.59, respectively. These results highlight the ability of ensemble methods to capture complex, non-linear relationships in FWG data effectively. In contrast, the PL2 model exhibited substantially lower performance, with an R^2 of 0.56, R^2CV of 0.63, and a markedly higher RMSE of 471.98. This indicates that simple polynomial relationships are insufficient to model the underlying biological complexity, resulting in larger prediction errors. Overall, the findings suggest that ensemble learning approaches, particularly RF and XGBoost, provide reliable and robust predictions for FWG.

Figure 2 Scatter plots presenting the values of real vs predictions of callus fresh weight gain for four models: Random Forest, Gradient Boosting, XGBoost, and second-degree Polynomial regression.

The combined analysis of numerical and graphical results highlights the superiority of ensemble tree-based models (RF, GB, and XGBoost) for predicting callus FWG. As shown in Fig. 2, these models exhibit regression slopes close to unity (RF: $0.98x + 42.04$; GB: $0.94x + 83.13$; XGBoost: $0.96x + 71.62$) and a tight clustering of points around the ideal $y = x$ line, indicating precise calibration and limited residual dispersion. These findings are consistent with the high coefficients of determination and low cross-validation errors (R^2CV) reported in Table 1, confirming the robustness and generalizability of these models across the full FWG range. In contrast, PL2 shows a markedly lower slope ($0.81x + 244.11$) and substantial dispersion, reflecting systematic underestimation of high FWG values and poor ability to capture the non-linearities inherent in biological data. Overall, these results indicate that ensemble models, particularly Random Forest, represent the most accurate and reliable approach for predicting callus fresh weight gain, whereas polynomial regression remains unsuitable for modeling such complex relationships.

The graphs show the overall importance of the variables (Figure. 2A) reveal that 2,4-D has the highest absolute mean SHAP value (~ 700), significantly surpassing those of BAP, KIN, and NAA, making it the most influential factor in predicting callus biomass accumulation. Analyzing the meaning of SHAP values (Figure. 3B), 2,4-D has a strongly positive effect on FWG ($+12.19$). Comparatively, BAP also contributes positively, but more moderately ($+6.39$). In contrast, KIN reduces predicted growth (-6.95), and NAA has a slightly negative impact (-2.01), highlighting its marginal role. Thus, these results confirm that 2,4-D is the key regulator of callus growth, BAP plays a secondary favorable role, while KIN may act as an inhibitor, and NAA remains of little influence.

Figure 3 Global importance (A) and directional effects (B) of hormones on predicted callus FWG using SHAP values in Random Forest models. 6-benzylaminopurine (BAP), kinetin (KIN), naphthaleneacetic acid (NAA), and 2,4-dichlorophenoxyacetic acid (2,4-D).

Experimental validation of callus fresh weight gain

Laboratory experiments were conducted to validate the five top phytohormone combinations predicted by the RF model and assess the reliability of the results. The validation experiment results (Figure. 4) demonstrated the reliability and accuracy of RF in predicting optimal combinations of plant growth regulators, enabling the achievement of maximal callus FWG. Results from the validation experiment provide robust evidence for the effectiveness of decision tree models, particularly Random Forest, as a powerful tool for optimizing callus induction. This approach also paves the way for suspension culture and large-scale production, while further enabling the optimization of micropropagation protocols for plants, such as vegetative propagation of *C. Spinosa*.

Callus culture as a platform for rutin accumulation

Hormonal regulation is a key factor in controlling the biosynthesis of secondary metabolites in *C. spinosa* callus cultures. In this study, rutin, the main metabolite of this species, was first identified using UPLC-QTOF-MS/MS, while relative concentration was then semi-quantified using UPLC-TQ-MS/MS in MRM mode (609 → 300). The BAP/2,4-D hormone combination was chosen for these experiments due to its favorable influence on FWG (Figure. 3), creating optimal conditions to assess rutin accumulation in callus tissues. The influence of hormonal treatments on rutin content was further analyzed using an RF model. Figure. 5 presents a comparison between experimentally determined and predicted rutin concentrations. The data indicate that rutin accumulation is strongly dependent on the relative concentrations of BAP and 2,4-D. Maximum rutin levels were observed under moderate BAP concentrations (0.5–1.0 mg/L) combined with low 2,4-D concentrations (0.0–1.0 mg/L), suggesting that these conditions favor activation of the flavonoid biosynthetic pathway. In contrast, higher concentrations of either BAP (≥ 1.5 mg/L) or 2,4-D (≥ 1.5 mg/L) markedly reduced rutin content, implying an inhibitory effect of excessive hormone levels on secondary metabolism. Overall, the RF model effectively captured these trends, with predicted values closely aligning with experimental observations.

Figure. 6 further illustrates the predictive performance of the RF model. Observed versus predicted rutin concentrations cluster around the ideal 1:1 line, underscoring the model's ability to faithfully reproduce the relationship between measured and estimated values. Regression analysis yielded a slope of 0.79 and an intercept of 0.25, with a high coefficient of determination ($R^2 = 0.821$) and a highly significant p -value, confirming the statistical robustness and reliability of the model. These results demonstrate that the model accurately captures overall trends, even within the complexity of a biological system. Ensemble tree-based algorithms are notable for their flexibility, capacity to model high-order nonlinear interactions, ability to handle heteroscedasticity, and robustness against overfitting. These qualities render the model a powerful tool for predicting and understanding the biosynthesis of secondary metabolites, such as rutin, in plant tissue cultures.

Experimental validation of rutin accumulation in callus cultures

The predictions of rutin accumulation obtained by the RF model, based on the BAP/2, 4-D combinations presented in section 3.3, show overall agreement with the experimental data, particularly for the condition BAP = 2.0 + 2,4-D = 2.0, highlighting the robustness of the model in anticipating metabolic

responses to different hormonal treatments. The slight overestimation observed for certain combinations ($\text{BAP} = 1.5 + 2.4\text{-D} = 2.0$ and $\text{BAP} = 2.0 + 2.4\text{-D} = 1.0$) does not alter the overall consistency of the predictions. These results demonstrate that the model is a reliable tool for estimating rutin accumulation in *C. spinosa* callus cultures and provide effective support for optimizing culture conditions to maximize secondary metabolite production.

Discussion

Callus induction is a central step in plant tissue culture, representing a cellular reprogramming process in which differentiated cells revert to an undifferentiated and proliferative state (Lee et al. 2024). This process is primarily regulated by the exogenous application of PGRs, which provide the plasticity required for subsequent morphogenetic events, such as organogenesis and somatic embryogenesis (Pasternak and Steinmacher 2024). Callus also serves as a reservoir for secondary metabolite production, including rutin. Therefore, investigating the effects of hormones, cytokinins, and auxins on callus induction and rutin accumulation is crucial for effective utilization of this biological resource.

In this study, we evaluated the effects of varying concentrations of BAP, KIN, NAA, and 2,4-D on FWG and rutin accumulation in *C. spinosa* using RF, GB, XGBoost, and PL2. ML algorithms, including RF and XGBoost, have been studied in plant tissue culture, yet comparative analyses remain limited.

These findings highlight that classic statistical methods, based on linear relationships or limited curvature, struggle to capture the complex interactions, threshold effects, and saturation phenomena characteristic of plant hormonal responses (Sadat-Hosseini et al. 2022). Modeling callus formation *in vitro*, where multiple hormonal factors interact in a nonlinear and potentially synergistic manner, requires more flexible approaches (Fallah Ziarani et al. 2022). Machine learning models, particularly tree-based ensemble algorithms such as RF and XGBoost, have proved particularly well suited to this challenge due to their ability to model high-order interactions, handle heteroscedasticity, and remain robust to overfitting (Munasinghe et al. 2020; Ali and Aasim 2024; Kaushik et al. 2025). The demonstrated accuracy of these models highlights the value of integrating machine learning into plant tissue culture research, not only to improve predictive ability but also to improve our understanding of the complex mechanisms driving hormonal responses.

To evaluate the relative contributions of individual growth regulators to callus development, we focused the SHAP analysis on the RF model, selected for its superior performance in terms of predictive accuracy. This approach allows for a quantitative assessment of the effect of each factor (BAP, KIN, NAA, and 2,4-D) on FWG predictions for *C. spinosa* callus tissue, while taking into account potential interactions between variables.

Recent studies comparing machine learning models for modeling and predicting saffron (*Crocus sativus*) callogenesis demonstrated that XGBoost and GB achieved superior accuracy ($R^2 > 0.95$) compared to artificial neural networks and RF regression in modeling *in vitro* culture systems (Sarabandi et al. 2024). In our study, we compare the predictive performance of RF, GB, XGBoost, and PL2 for FWG in *C. spinosa*.

Tree-based Learning models consistently outperformed polynomial regression, with RF achieving the highest predictive accuracy ($R^2 > 0.86$) for callogenesis. Although this represents the first direct comparison of these models in *C. spinosa in vitro* culture, other studies have similarly reported the superior accuracy of RF (Özcan et al. 2023; Bozkurt et al. 2024; Sarmah et al. 2025). Overall, a model's predictive capability is largely determined by the size and complexity of the training dataset (Hesami and Jones 2020). In this context, the RF model has consistently demonstrated a remarkable ability to accurately predict complex callus-related biological outcomes, achieving a high coefficient of determination ($R^2 = 0.97$) for both callus growth parameters and secondary metabolite production in *Nilgirianthus ciliatus* (Jeevan Ram et al. 2025). These results confirm RF as a robust and reliable tool for predicting callus growth and metabolite accumulation, providing confidence in its application for optimizing *in vitro* culture conditions.

Sensitivity analyses were applied to interpret the RF model and quantify the contribution of each variable to FWG predictions. SHAP values as an interpretative tool for identifying and quantifying variable importance have been demonstrated in multiple studies (Ekanayake et al. 2022; Antonini et al. 2024). SHAP provides a mathematically rigorous, objective, and consistent method to decompose model predictions into contributions from each variable (Lundberg and Lee 2017). This approach enables precise local explanation of individual predictions while offering a global perspective on variable importance, making it robust and broadly applicable to machine learning interpretation (Lundberg et al. 2020). Using SHAP, key variables such as 2,4-D, NAA, sucrose concentration, and culture duration have been successfully identified (Sarabandi et al. 2024). In our study, SHAP analysis confirmed that 2,4-D was the most influential factor, KIN exhibited inhibitory effects, and BAP acted synergistically with 2,4-D to enhance callus formation. Identification of these factors is particularly relevant when secondary metabolite production is the primary objective of callus culture.

To our knowledge, this study provides the first systematic assessment of the impact of various PGRs on both callus induction and rutin production in *C. spinosa*. To directly link callus growth with metabolite production, we employed a stacked RF model, using callogenesis outputs as inputs to predict rutin accumulation. This approach effectively modeled the nonlinear relationship between cell proliferation and rutin biosynthesis, yielding high predictive performance ($R^2 > 0.82$) and minimal deviation between observed and predicted values, confirming method robustness.

For instance, Yin et al. (2014) reported that MS medium supplemented with 1.5 mg/L 2,4-D and 3.0 mg/L BAP induces callogenesis and that callus may serve as a source of volatile organic compounds. Similarly, Duran and Issah (2022) demonstrated that BAP at 1 mg/L combined with 2 mg/L NAA produced a callogenesis rate of approximately 120 mg after 10 days. While these studies highlighted the impact of the strigolactone GR24 on callus production and phenolic content in *C. spinosa*, our results showed that BAP + NAA achieved over 1000 mg FWG over 30 days, although it was less effective than BAP + 2,4-D. This confirms that NAA is less favorable for callogenesis. This combination also resulted in substantial rutin accumulation, with a modest increase in the presence of 0.1 μ M GR24, indicating that *C. spinosa* callus can serve as a rutin reservoir (Duran and Issah 2022).

Direct comparison between studies is limited due to differences in explant source, methodology, and objectives. In the present work, maximum callus formation was observed with 2,4-D, consistent with reports in *C. spinosa* (Sobhy et al. 2025), and other species where auxin application significantly promotes callogenesis (Slimani et al. 2021; Teoh et al. 2023; Ranade et al. 2023; Mahood et al. 2024; Abdelazeez et al. 2025).

Regarding rutin accumulation, low concentrations of BAP and 2,4-D (≤ 1 mg/L) promoted significant accumulation, peaking at 0.5 mg/L for each hormone, paralleling biomass increases. Similar trends were reported by Kianersi et al. (2020), with further enhancements achieved using elicitors such as methyl jasmonate. (Goda et al. 2017) also reported that 2,4-D-induced callus exhibited significantly higher total flavonoid content. These findings highlight the importance of optimizing hormone concentrations and combining them with elicitors for large-scale production of high-quality bioactive compounds (Sagharyan et al. 2020; Ozyigit et al. 2023; Pérez-Mejía et al. 2024).

The reliability of RF predictions was validated experimentally. Predicted maximal FWG results closely matched observed values, demonstrating the model's accuracy. For rutin accumulation, satisfactory performance was observed, particularly for BAP + 2,4-D at 2 mg/L, with minor overestimations in some combinations. These results confirm RF as a robust and effective machine learning algorithm for modeling and optimizing callus culture in this study.

Future studies using larger datasets and additional variables could further improve model performance. Investigations on *C. spinosa* could explore 2,4-D with other PGRs or extend to cell suspension cultures and bioreactor-based production systems.

Conclusion

This research provides the first comprehensive evaluation of the combined effects of BAP, KIN, NAA, and 2,4-D on callogenesis and rutin biosynthesis in *C. spinosa* using an integrated machine learning and explainable-AI framework. Ensemble tree-based models, particularly Random Forest, accurately predicted callus biomass and metabolite accumulation, outperforming classical statistical approaches. SHAP analysis revealed 2,4-D as the primary determinant of callus proliferation, while BAP acted synergistically to enhance growth, and KIN and NAA contributed weak or inhibitory effects. Experimental validation confirmed the reliability of the predicted optimal hormonal combinations, demonstrating strong agreement between predicted and observed FWG and rutin levels. Moderate concentrations of BAP and 2,4-D favored rutin accumulation, underscoring the importance of fine hormonal tuning for metabolic optimization.

Overall, this study highlights the relevance of ML-driven modeling for deciphering complex hormonal interactions and accelerating protocol optimization in plant tissue culture. The established models constitute a robust foundation for scaling up callus-based production systems, including suspension cultures and bioreactors, aimed at sustainable, high-yield biosynthesis of rutin and other valuable metabolites in *C. spinosa*. Future investigations integrating larger datasets, molecular markers, and

elicitation strategies will further strengthen predictive performance and advance the biotechnological exploitation of this species.

Declarations

CRedit authorship contribution statement

Marouane Mohaddab: Writing – review & editing, Writing – original draft, Conceptualization, Methodology, Data curation, Formal analysis. **Younes EL Goumi:** Supervision, Writing – review and editing, Methodology. **Mohammed Elakrouch:** Formal analysis, Software, Methodology **Soufiane Hasni:** Formal analysis, Software, Data curation. **Clément Burgeon:** Writing – review & editing, Software, Data curation. **Manon Genva:** Writing – review & editing, Software. **Malika Fakiri:** Writing – review & editing, Validation, Project administration **Marie-Laure Fauconnier:** Writing – review & editing, Validation, Project administration

Ethics declaration

Not applicable.

Declaration of interests

The authors declare that there is no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Aasim M, Ali SA, Altaf MT et al (2023) Artificial neural network and decision tree facilitated prediction and validation of cytokinin-auxin induced *in vitro* organogenesis of sorghum (*Sorghum bicolor* L). Plant Cell Tiss Organ Cult 153:611–624. <https://doi.org/10.1007/s11240-023-02498-3>
2. Abdelazeez WMA, Aboueldis GR, Suliman AA, Mohammed DM (2025) Production of secondary metabolites in callus cultures of *Scutellaria baicalensis* L. and assessment of their anti-inflammatory and antioxidant efficacy in ulcerative colitis rats. Plant Cell Tiss Organ Cult 160:80. <https://doi.org/10.1007/s11240-025-02996-6>
3. Alcalde MA, Müller M, Munné-Bosch S et al (2022) Using machine learning to link the influence of transferred *Agrobacterium rhizogenes* genes to the hormone profile and morphological traits in *Centella asiatica* hairy roots. Front Plant Sci 13:1001023
4. Ali SA, Aasim M (2024) Response surface methodology and artificial intelligence modeling for *in vitro* regeneration of Brazilian micro sword (*Lilaeopsis brasiliensis*). Plant Cell Tiss Organ Cult

- 157:10. <https://doi.org/10.1007/s11240-024-02734-4>
5. Al-Safadi B, Elias R (2011) Improvement of caper (*Capparis spinosa* L.) propagation using *in vitro* culture and gamma irradiation. *Scientia Horti* 127:290–297
 6. Annaz H, Sane Y, Bitchagno GTM et al (2022) Caper (*Capparis spinosa* L.): an updated review on its phytochemistry, nutritional value, traditional uses, and therapeutic potential. *Front Pharmacol.* 2022; 13: 1–22
 7. Antonini AS, Tanzola J, Asiain L et al (2024) Machine learning model interpretability using SHAP values: Application to Igneous Rock Classification task. *Appl Comput Geosci* 23:100178. <https://doi.org/10.1016/j.acags.2024.100178>
 8. Baker RE, Peña J-M, Jayamohan J, Jérusalem A (2018) Mechanistic models versus machine learning, a fight worth fighting for the biological community? *Biol Lett* 14:20170660. <https://doi.org/10.1098/rsbl.2017.0660>
 9. Belkessam M, Genva M, Kouki A et al (2025) Phytochemical Profile and Cosmeceutical Potential of Leaf Extracts of Two Species of the *Anacardiaceae* Family from the Mediterranean Scrubland: *Pistacia lentiscus* L. and *Pistacia atlantica* Desf. *Processes* 13:3712. <https://doi.org/10.3390/pr13113712>
 10. Bozkurt T, İnan S, Dündar İ et al (2024) Optimizing the *In Vitro* Propagation of Tea Plants: A Comparative Analysis of Machine Learning Models. *Horticulturae* 10:721. <https://doi.org/10.3390/horticulturae10070721>
 11. Bravo-Vázquez LA, Angulo–Bejarano PI, Bandyopadhyay A et al (2023) Regulatory roles of noncoding RNAs in callus induction and plant cell dedifferentiation. *Plant Cell Rep* 42:689–705. <https://doi.org/10.1007/s00299-023-02992-0>
 12. Caglar G, Caglar S, Ergin O, Yarim M (2005) The influence of growth regulators on shoot proliferation and rooting of *in vitro* propagated caper. *J Environ Biol* 26:479–485
 13. Carra A, Del Signore MB, Sottile F et al (2012) Potential use of new diphenylurea derivatives in micropropagation of *Capparis spinosa* L. *Plant Growth Regul* 66:229–237. <https://doi.org/10.1007/s10725-011-9645-3>
 14. Dar SA, Nawchoo IA, Tyub S, Kamili AN (2021) Effect of plant growth regulators on *in vitro* induction and maintenance of callus from leaf and root explants of *Atropa acuminata* Royle ex Lindl. *Biotechnol Rep* 32:e00688
 15. Demirel F, Uğur R, Popescu GC et al (2023) Usage of Machine Learning Algorithms for Establishing an Effective Protocol for the *In Vitro* Micropropagation Ability of Black Chokeberry (*Aronia melanocarpa* (Michx.) Elliott). *Horticulturae* 9:1112. <https://doi.org/10.3390/horticulturae9101112>
 16. Duran RE, Issah H (2022) The impact of strigolactone GR24 on *Capparis spinosa* L. callus production and phenolic compound content. *Plant Cell Tiss Organ Cult* 149:197–204
 17. Efferth T (2019) Biotechnology applications of plant callus cultures. *Engineering* 5:50–59
 18. Ekanayake IU, Meddage DPP, Rathnayake U (2022) A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive

- explanations (SHAP). *Case Stud Constr Mate* 16:e01059.
<https://doi.org/10.1016/j.cscm.2022.e01059>
19. El Amri N, Chabir R, Kachkoul R, Bour A (2025) Phenolic Composition and Phytochemical Analysis of Moroccan Raw and Processed *Capparis spinosa* Buds and Berries. *Chem Biodivers*.
<https://doi.org/10.1002/cbdv.202500443>
 20. Eren B, Türkoğlu A, Haliloğlu K et al (2023) Investigation of the influence of polyamines on mature embryo culture and DNA methylation of wheat (*Triticum aestivum* L.) using the machine learning algorithm method. *Plants* 12:3261
 21. Esmaeilpour HR, Boostani R, Shoeibi A et al (2020) The potential effects of Caper (*Capparis spinosa* L.) in the treatment of diabetic neuropathy. *Trad Med Res* 5:442–448
 22. Fahmideh L, Sheikhi M, Benakashani F, Solouki M (2019) Callus induction and organogenesis from various explants of plant *Capparis spinosa* L. under *In vitro* conditions. *J Plant Prod Res* 26:75–88.
<https://doi.org/10.22069/jopp.2019.14326.2284>
 23. Fallah Ziarani M, Tohidfar M, Navvabi M (2022) Modeling and optimizing *in vitro* percentage and speed callus induction of carrot via Multilayer Perceptron-Single point discrete GA and radial basis function. *BMC Biotechnol* 22. <https://doi.org/10.1186/s12896-022-00764-4>
 24. Fotiadou E, Cheilari A, Graikou K, Chinou I (2025) Comparative Study by HPTLC of Selected *Capparis spinosa* Samples (Buds and Leaves) from the Cycladic Islands in Greece. *Foods* 14:1827
 25. Gianguzzi V, Inglese P, Barone E, Sottile F (2019) *In Vitro* Regeneration of *Capparis spinosa* L. by Using a Temporary Immersion System. *Plants* 8:177. <https://doi.org/10.3390/plants8060177>
 26. Goda SM, Ahmed SA, El Sherif F et al (2017) Genetically stable plants with boosted flavonoids content after *in vitro* regeneration of the endangered *Capparis spinosa* L. *Glob. Drugs Ther* 2:1–7
 27. Grimalt M, Hernández F, Legua P et al (2018) Physicochemical composition and antioxidant activity of three Spanish caper (*Capparis spinosa* L.) fruit cultivars in three stages of development. *Scientia hortic* 240:509–515
 28. Grimalt M, Hernández F, Legua P et al (2022) Antioxidant activity and the physicochemical composition of young caper shoots (*Capparis spinosa* L.) of different Spanish cultivars. *Scientia Hortic* 293:110646
 29. Gristina AS, Fici S, Siragusa M et al (2014) Hybridization in *Capparis spinosa* L.: Molecular and morphological evidence from a Mediterranean island complex. *Flora* 209:733–741
 30. Hagaggi NShA, Abdul-Raouf UM, Radwan TAA (2024) Variation of antibacterial and antioxidant secondary metabolites and volatiles in leaf and callus extracts of Phulai (*Acacia Modesta* Wall). *BMC Plant Biol* 24:93. <https://doi.org/10.1186/s12870-024-04747-9>
 31. Hazrati S, Mousavi Z, Mollaei S et al (2025) Effect of Phenological Variation on the Phytochemical Composition and Antioxidant Activity of Different Organs of *Capparis spinosa* L. *Horticulturae* 11:702. <https://doi.org/10.3390/horticulturae11060702>
 32. Hesami M, Jones AMP (2021) Modeling and optimizing callus growth and development in *Cannabis sativa* using random forest and support vector machine in combination with a genetic algorithm.

- Appl Microbiol Biotechnol 105:5201–5212. <https://doi.org/10.1007/s00253-021-11375-y>
33. Hesami M, Jones AMP (2020) Application of artificial intelligence models and optimization algorithms in plant cell and tissue culture. Appl Microbiol Biotechnol 104:9449–9485. <https://doi.org/10.1007/s00253-020-10888-2>
34. Higazy AE, El-Mahrouk ME, El-Banna AN et al (2023) Production of Black Cumin via Somatic Embryogenesis, Chemical Profile of Active Compounds in Callus Cultures and Somatic Embryos at Different Auxin Supplementations. Agronomy 13:2633. <https://doi.org/10.3390/agronomy13102633>
35. Ikeuchi M, Sugimoto K, Iwase A (2013) Plant callus: mechanisms of induction and repression. Plant Cell 25:3159–3173
36. Jeevan Ram PS, Singh S, Ali SA et al (2025) Machine Learning-Driven Screening of Promising Mutants in *Nilgiranthus ciliatus* Callus Cultures for Enhanced Phytochemicals Production. Physiol Plant 177:e70536. <https://doi.org/10.1111/ppl.70536>
37. Karous O, Ben Haj Jilani I, Ghrabi-Gammar Z (2021) Ethnobotanical study on plant used by semi-nomad descendants' community in Ouled Dabbeeb—Southern. Tunisia Plants 10:642
38. Katırcı R, Aasim M, Deveci G, Mustafa Z (2024) Comparing quantum machine learning and classical machine learning for *in vitro* regeneration of cowpea (*Vigna unguiculata*). Plant Cell Tiss Organ Cult 159:32. <https://doi.org/10.1007/s11240-024-02880-9>
39. Kaushik P, Rani M, Khurana N et al (2025) Revolutionizing Plant Tissue Culture: Harnessing Artificial Intelligence for Precision Propagation and Optimization. NPJ 15:e040624230642. <https://doi.org/10.2174/0122103155302871240527094915>
40. Kianersi F, Abdollahi MR, Mirzaie-asl A et al (2020) Biosynthesis of rutin changes in *Capparis spinosa* due to altered expression of its pathway genes under elicitors' supplementation. Plant Cell Tiss Organ Cult 141:619–631
41. Kirkan B, Ceylan O, SARIKÜRKÜ C, Bektas T (2021) Phenolic profile, antioxidant and enzyme inhibitory activity of the ethyl acetate, methanol and water extracts of *Capparis spinosa* L. Int J Secondary Metabolite 8:337–351
42. Koçak M, Yılmaz MC, Kuzğun C et al (2025) Synthetic seed production in *Crataegus monogyna* L. and prediction of regeneration of synthetic seeds with machine learning algorithms. Plant Cell Tiss Organ Cult 161. <https://doi.org/10.1007/s11240-025-03049-8>
43. Koufan M, Belkoura I, Mazri MA (2022) *In Vitro* Propagation of Caper (*Capparis spinosa* L.): A Review. Horticulturae 8:737. <https://doi.org/10.3390/horticulturae8080737>
44. Labbafi M, Mehrafarin A, Badi H et al (2018) Improve germination of caper (*Capparis spinosa* L.) seeds by different induction treatments of seed dormancy breaking. Trakia J Sci 16:71–74
45. Lee KY, Nam D-H, Jeon Y et al (2024) Exploring the Production of Secondary Metabolites from a Halophyte *Tetragonia tetragonoides* through Callus Culture. Horticulturae 10:244. <https://doi.org/10.3390/horticulturae10030244>
46. Liu W, He Y, Xiang J et al (2011) The physiological response of suspension cell of *Capparis spinosa* L. to drought stress. J Med Plants Res 5:5899–5906

47. Lundberg SM, Erion G, Chen H et al (2020) From Local Explanations to Global Understanding with Explainable AI for Trees. *Nat Mach Intell* 2:56–67. <https://doi.org/10.1038/s42256-019-0138-9>
48. Lundberg SM, Lee S-I (2017) A Unified Approach to Interpreting Model Predictions
49. Ma Q, Grönes P, Robert S (2018) Auxin signaling: a big question to be addressed by small molecules. *J Exp Bot* 69:313–328. <https://doi.org/10.1093/jxb/erx375>
50. Mahood HE, Sarropoulou V, Tzatzani T-T et al (2024) Biomass and polyphenolic enhancement in date palm (*Phoenix dactylifera* L.) callus cultures through *Fusarium oxysporum* elicitation. *Plant Cell Tiss Organ Cult* 159:76. <https://doi.org/10.1007/s11240-024-02934-y>
51. Moghadamnia Y, Kani SNM, Ghasemi-Kasman M et al (2019) The anti-cancer effects of *Capparis spinosa* hydroalcoholic extract. *Avicenna J med biotechnol* 11:43
52. Mohaddab M, El Goumi Y, Gallo M et al (2022) Biotechnology and *In Vitro* Culture as an Alternative System for Secondary Metabolite Production. *Molecules* 27:8093. <https://doi.org/10.3390/molecules27228093>
53. Mohaddab M, Genva M, Malika F et al (2024) *Capparis spinosa*: A Rich Source of Phenolic Compounds-A Comprehensive Review of Its Phytochemistry, Health Benefits, and Biotechnological Applications. *Biocatal Agric Biotechnol* 103409
54. Movafeghi A, HABIBI G, ALI AM (2008) Plant regeneration of *Capparis spinosa* L. using hypocotyl explants
55. Munasinghe SP, Somaratne S, Weerakoon SR, Ranasinghe C (2020) Prediction of chemical composition for callus production in *Gyrinops walla* Gaertner through machine learning. *Inf Process Agric* 7:511–522
56. Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. *Physiol Plant* 15:473–497. <https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>
57. Noordijk B, Garcia Gomez ML, ten Tusscher KHWJ et al (2024) The rise of scientific machine learning: a perspective on combining mechanistic modelling with machine learning for systems biology. *Front Syst Biol* 4. <https://doi.org/10.3389/fsysb.2024.1407994>
58. Nowruzian A, Aalami A (2023) Effect of different treatments on dormancy breaking for seed germination enhancement and metabolite analysis of *Capparis spinosa* L. *J Med plants By-Products* 12:387–396
59. Özcan E, Atar HH, Ali SA, Aasim M (2023) Artificial neural network and decision tree–based models for prediction and validation of *in vitro* organogenesis of two hydrophytes—*Hemianthus callitrichoides* and *Riccia fluitans*. *In Vitro Cell Dev Biol - Plant* 59:547–562. <https://doi.org/10.1007/s11627-023-10367-z>
60. Ozyigit II, Dogan I, Hocaoglu-Ozyigit A et al (2023) Production of secondary metabolites using tissue culture-based biotechnological applications. *Front Plant Sci* 14:1132555
61. Pasternak TP, Steinmacher D (2024) Plant Growth Regulation in Cell and Tissue Culture. *Vitro Plants* 13:327. <https://doi.org/10.3390/plants13020327>

62. Peng S, Rajjou L (2024) Advancing plant biology through deep learning-powered natural language processing. *Plant Cell Rep* 43:208
63. Pérez-Mejía N, Villarreal ML, Sánchez-Carranza JN et al (2024) Phytochemical Profiles and Cytotoxic Activity of *Bursera fagaroides* (Kunth) Engl. Leaves and Its Callus Culture. *Plants* 13:1622. <https://doi.org/10.3390/plants13121622>
64. Petch J, Di S, Nelson W (2022) Opening the Black Box: The Promise and Limitations of Explainable Machine Learning in Cardiology. *Can J Cardiol* 38:204–213. <https://doi.org/10.1016/j.cjca.2021.09.004>
65. Rahimi VB, Rajabian A, Rajabi H et al (2020) The effects of hydro-ethanolic extract of *Capparis spinosa* (C. *spinosa*) on lipopolysaccharide (LPS)-induced inflammation and cognitive impairment: Evidence from *in vivo* and *in vitro* studies. *J Ethnopharmacol* 256:112706
66. Ranade R, Joshi N, Kudale S (2023) Comparative secondary metabolite expression in callus cultures and mother plant in *Barleria prionitis* L. *Plant Cell Tiss Organ Cult* 155:653–663. <https://doi.org/10.1007/s11240-023-02585-5>
67. Sadat-Hosseini M, Arab MM, Soltani M et al (2022) Predictive modeling of Persian walnut (*Juglans regia* L.) *in vitro* proliferation media using machine learning approaches: a comparative study of ANN, KNN and GEP models. *Plant Methods* 18:48. <https://doi.org/10.1186/s13007-022-00871-5>
68. Sagharyan M, Ganjeali A, Cheniany M, Kouhi SMM (2020) Optimization of callus induction with enhancing production of phenolic compounds production and antioxidants activity in callus cultures of *Nepeta binaloudensis* Jamzad (*Lamiaceae*). *Iran J Biotechnol* 18:e2621
69. Sarabandi M, Singh RK, Kalantari S et al (2024) Unveiling machine learning's impact on *in vitro* callogenesis optimization in *Crocus sativus* L. *S Afr J Bot* 168:1–8
70. Sarmah P, Borah T, Mudoi KD et al (2025) *In vitro* micropropagation of the critically endangered *Aquilaria malaccensis* Lam. and comparing the effect of PGRs' through machine learning models. *Plant Cell Tiss Organ Cult* 162:6. <https://doi.org/10.1007/s11240-025-03115-1>
71. Slimani C, El Goumi Y, Rais C et al (2021) Optimization of callogenesis/caulogenesis induction protocol in saffron plant (*Crocus sativus* L.) using response surface methodology. *Biointerface Res Appl Chem* 12:4731–4746
72. Sobhy FS, Ahmed S, Abdelrahim E et al (2025) Chemo Preventive Potential of Rutin-Rich Extract Obtained from *Capparis spinosa* Culture against HepG2 Cancer Cell Line Activity. *Egypt J Chem* 68:855–866
73. Sozzi GO, Peter KV, Babu KN, Divakaran M (2012) Capers and caperberries. *Handbook of herbs and spices*. Elsevier, pp 193–224
74. Sozzi GO, Vicente AR (2006) Capers and caperberries. *Handbook of Herbs and Spices*. Elsevier, pp 230–256
75. Svolacchia N, Sabatini S (2023) Cytokinins. *Curr Biol* 33:R10–R13. <https://doi.org/10.1016/j.cub.2022.11.022>

76. Tan WH, Tong CY, Chua MX, Derek CJC (2024) Modelling and optimizing secondary metabolites production in *Spirodela polyrhiza* using machine learning. J Clean Prod 478:143986. <https://doi.org/10.1016/j.jclepro.2024.143986>
77. Teoh SC, Subramaniam S, Chew BL (2023) The effects of 2, 4-dichlorophenoxyacetic acid on the induction of callus from cotyledon and hypocotyl explants of butterfly pea (*Clitoria ternatea*). Malays Appl Biol 52:61–72
78. Wang YT, Gan L, Liu W et al (2007) Research on the callus inducing and the cell growth and metabolism characteristics of *Capparis spinosa* L. College of Life Sci and Technolo, Huazhong Univ of Sci and Technolo. Wuhan China 7:1779–1783
79. Wijerathna-Yapa A, Hiti-Bandaralage J, Pathirana R (2025) Harnessing metabolites from plant cell tissue and organ culture for sustainable biotechnology. Plant Cell Tiss Organ Cult 162:55. <https://doi.org/10.1007/s11240-025-03180-6>
80. Yin Y, He Y, Liu W et al (2014) The durative use of suspension cells and callus for volatile oil by comparative with seeds and fruits in *Capparis spinosa* L. PLoS ONE 9:e113668
81. Yousefi E, Abedi M, Aghajanzadeh TA, Moreno DA (2025) Caper bush (*Capparis spinosa* L.) bioactive compounds and antioxidant capacity as affected by adaptation to harsh soils. Sci Rep 15:11893. <https://doi.org/10.1038/s41598-025-97298-4>
82. Zarbakhsh S, Shahsavar AR, Soltani M (2024) Optimizing PGRs for *in vitro* shoot proliferation of pomegranate with bayesian-tuned ensemble stacking regression and NSGA-II: a comparative evaluation of machine learning models. Plant Methods 20:1–24. <https://doi.org/10.1186/s13007-024-01211-5>
83. Zhang H, Ma ZF (2018) Phytochemical and pharmacological properties of *Capparis spinosa* as a medicinal plant. Nutrients 10:116

Figures

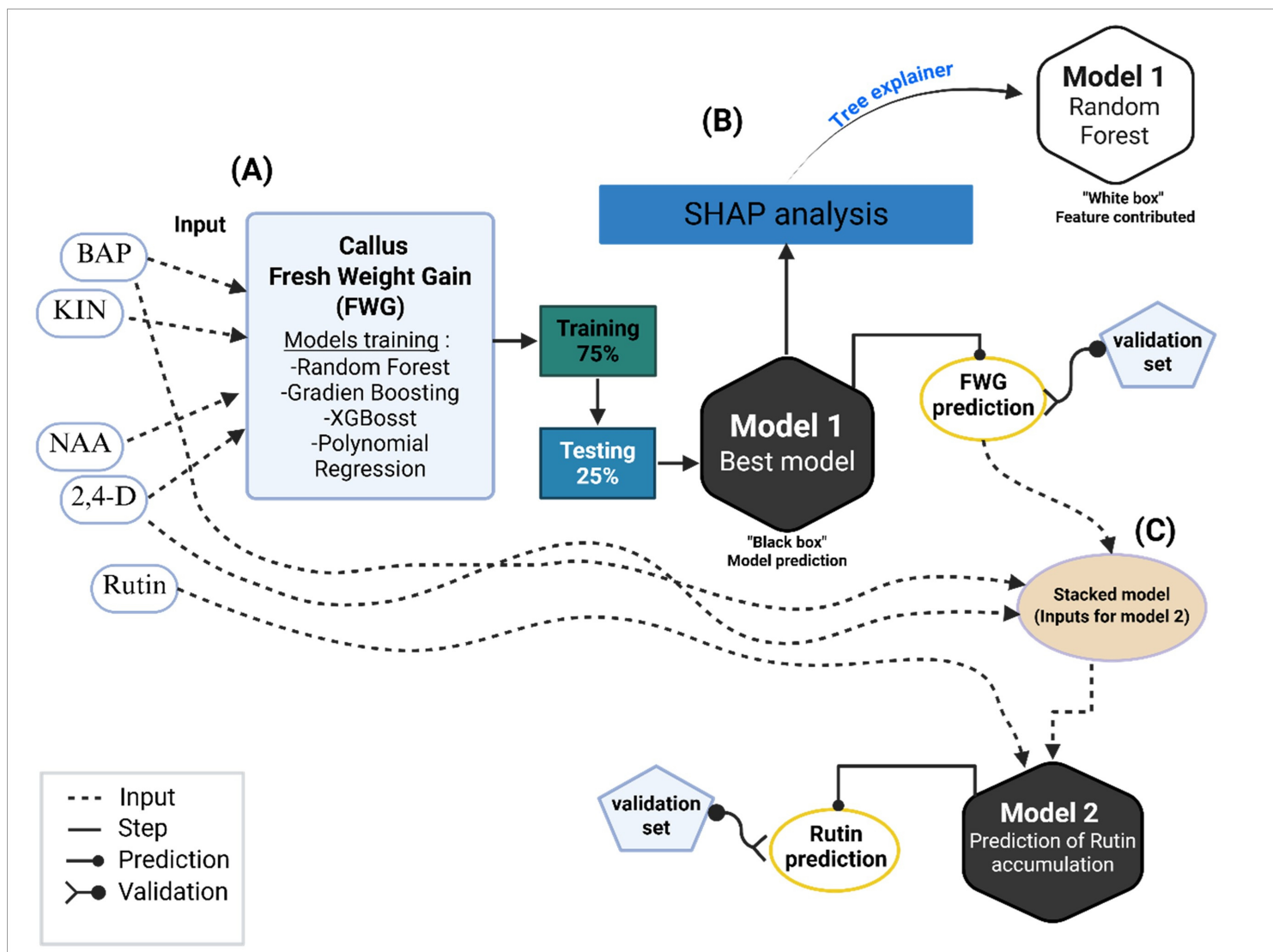


Figure 1

A schematic representation of the stepwise computational approach used in this study. (A) Modeling and prediction of callus Fresh Weight Gain (FWG) (output) based on 6-Benzylaminopurine (BAP), Kinetin (KIN), Naphthalene Acetic Acid (NAA), and 2,4-Dichlorophenoxyacetic Acid (2,4-D) as inputs using four models. (B) SHAP value analysis of the best-selected model to identify and quantify the key model features. (C) Stacking of the best model and using it as input for the production and accumulation of rutin (Model 2).

Predicted vs Real Values for Each Model

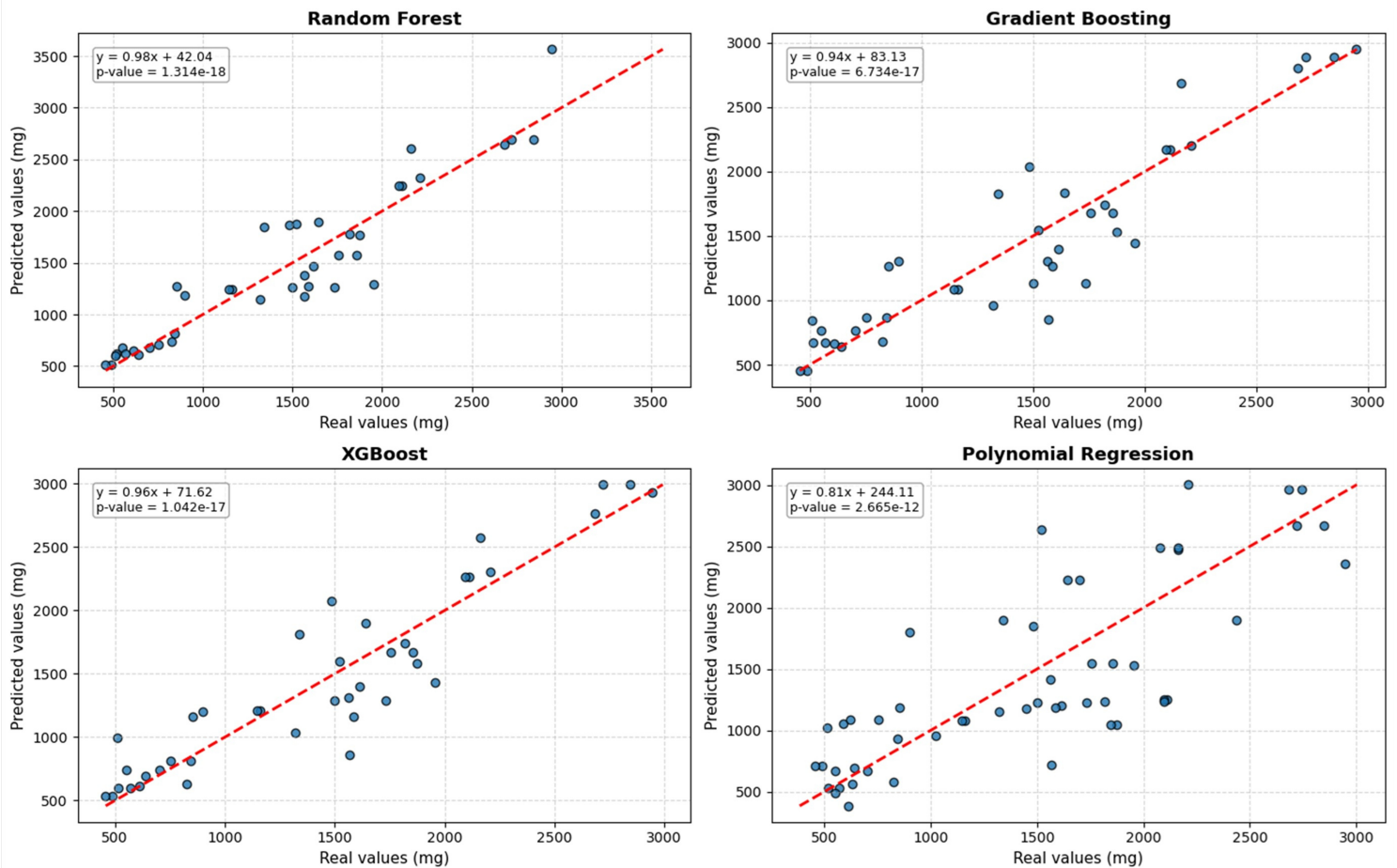


Figure 2

Scatter plots presenting the values of real vs predictions of callus fresh weight gain for four models: Random Forest, Gradient Boosting, XGBoost, and second-degree Polynomial regression.

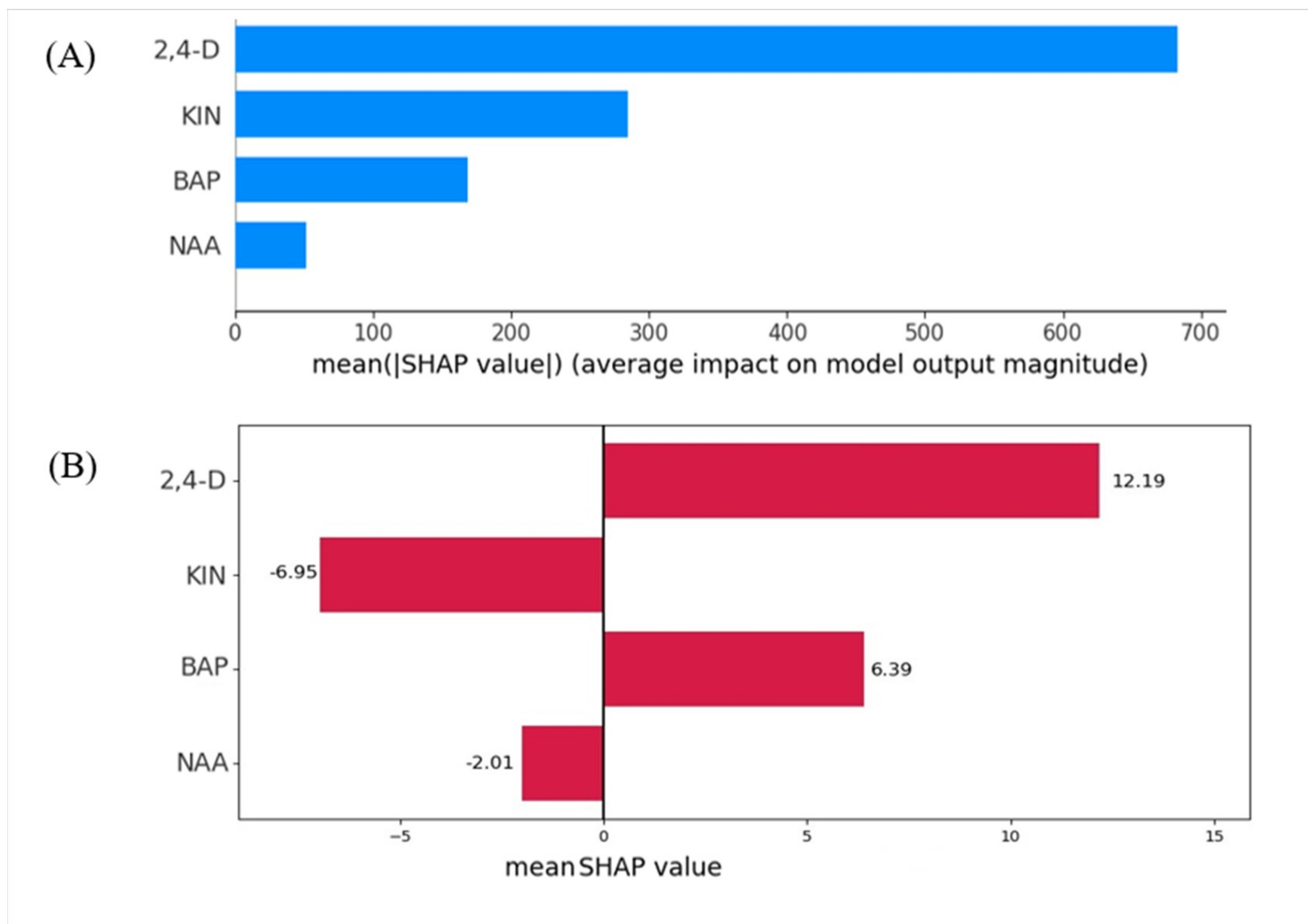


Figure 3

Global importance (A) and directional effects (B) of hormones on predicted callus FWG using SHAP values in Random Forest models. 6-benzylaminopurine (BAP), kinetin (KIN), naphthaleneacetic acid (NAA), and 2,4-dichlorophenoxyacetic acid (2,4-D).

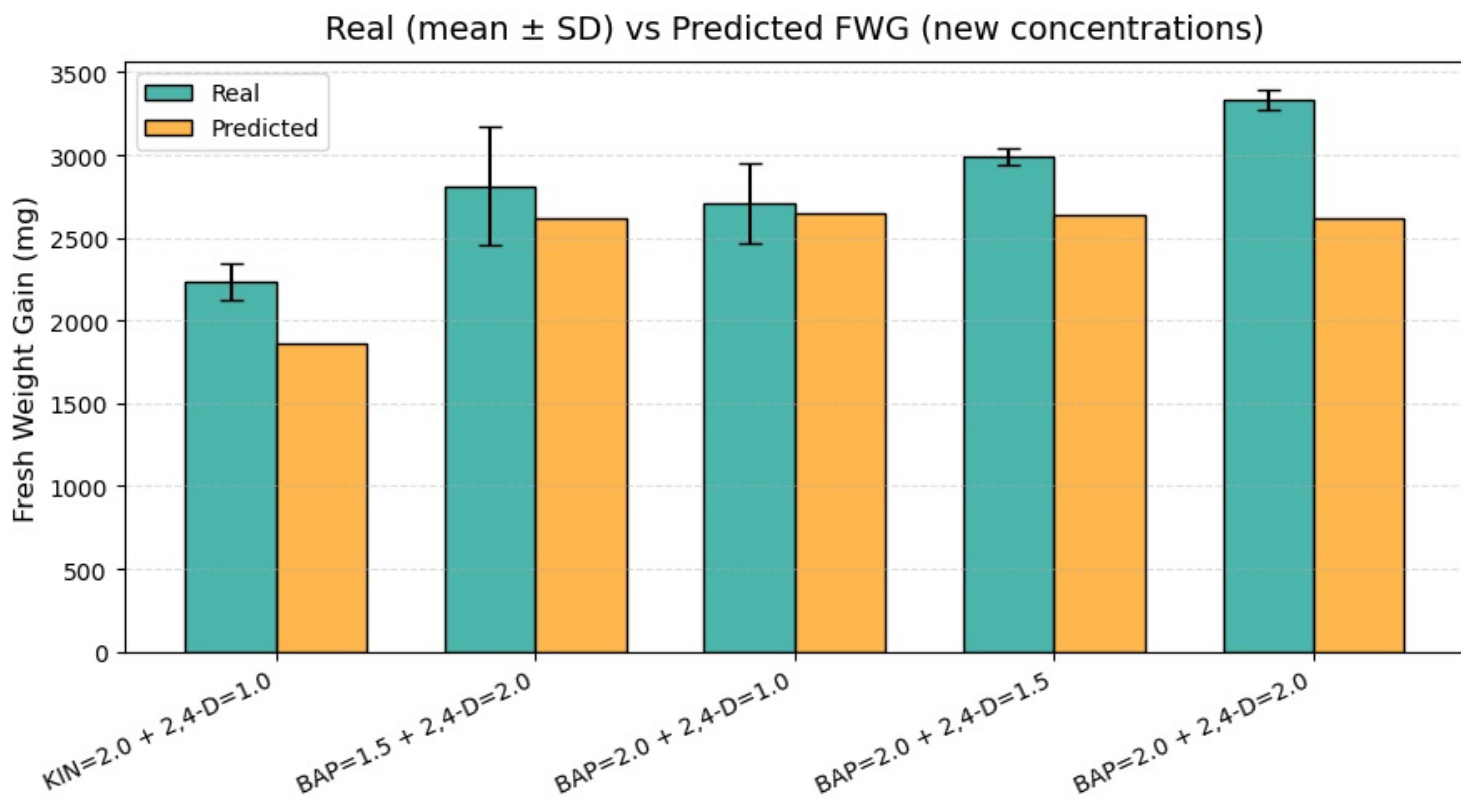


Figure 4

Real vs top 5 predicted combinations for maximization of callus fresh weight gain using random forest model.

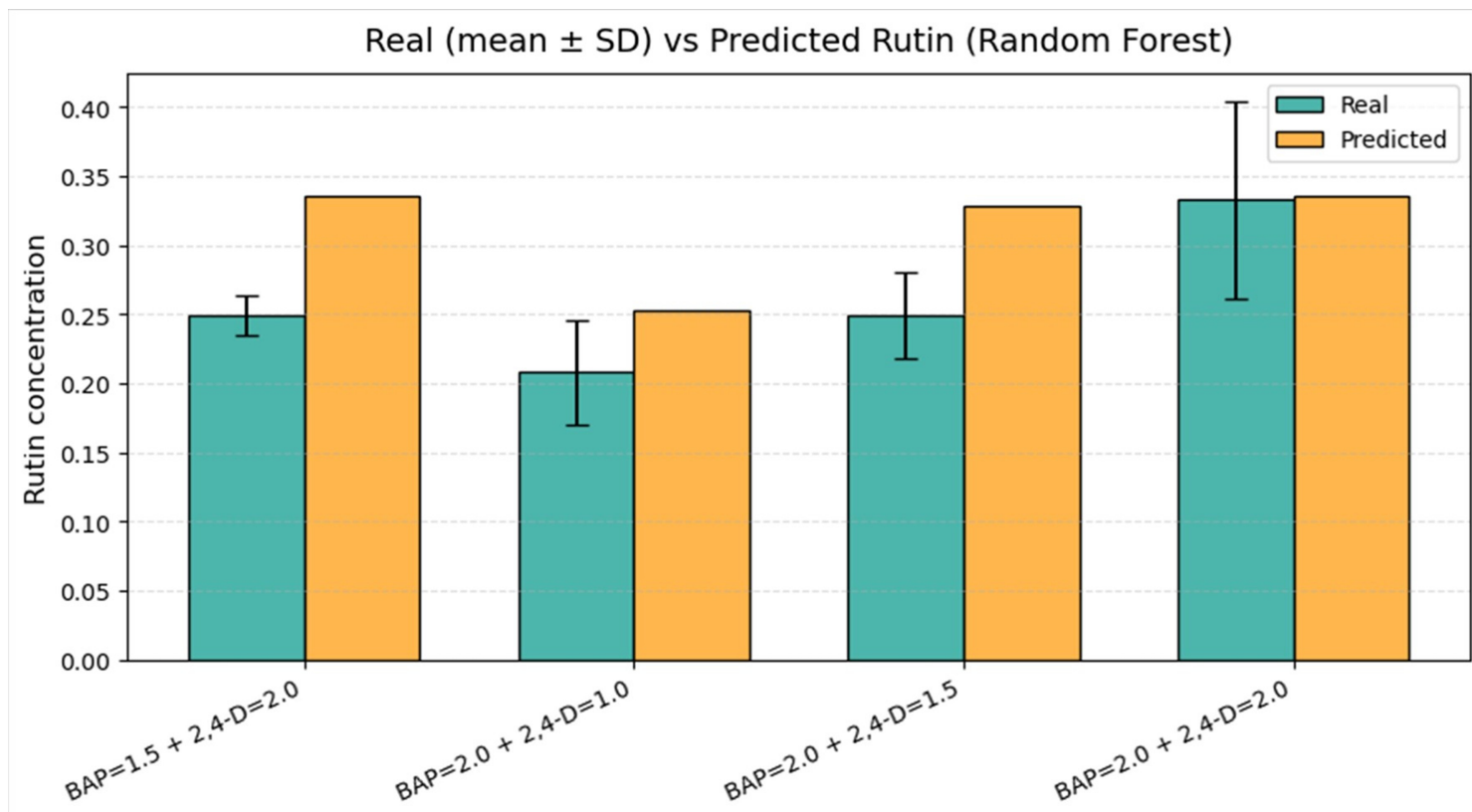


Figure 5

Comparison of real (green) and predicted (orange) rutin under different hormonal combinations using the random forest model.

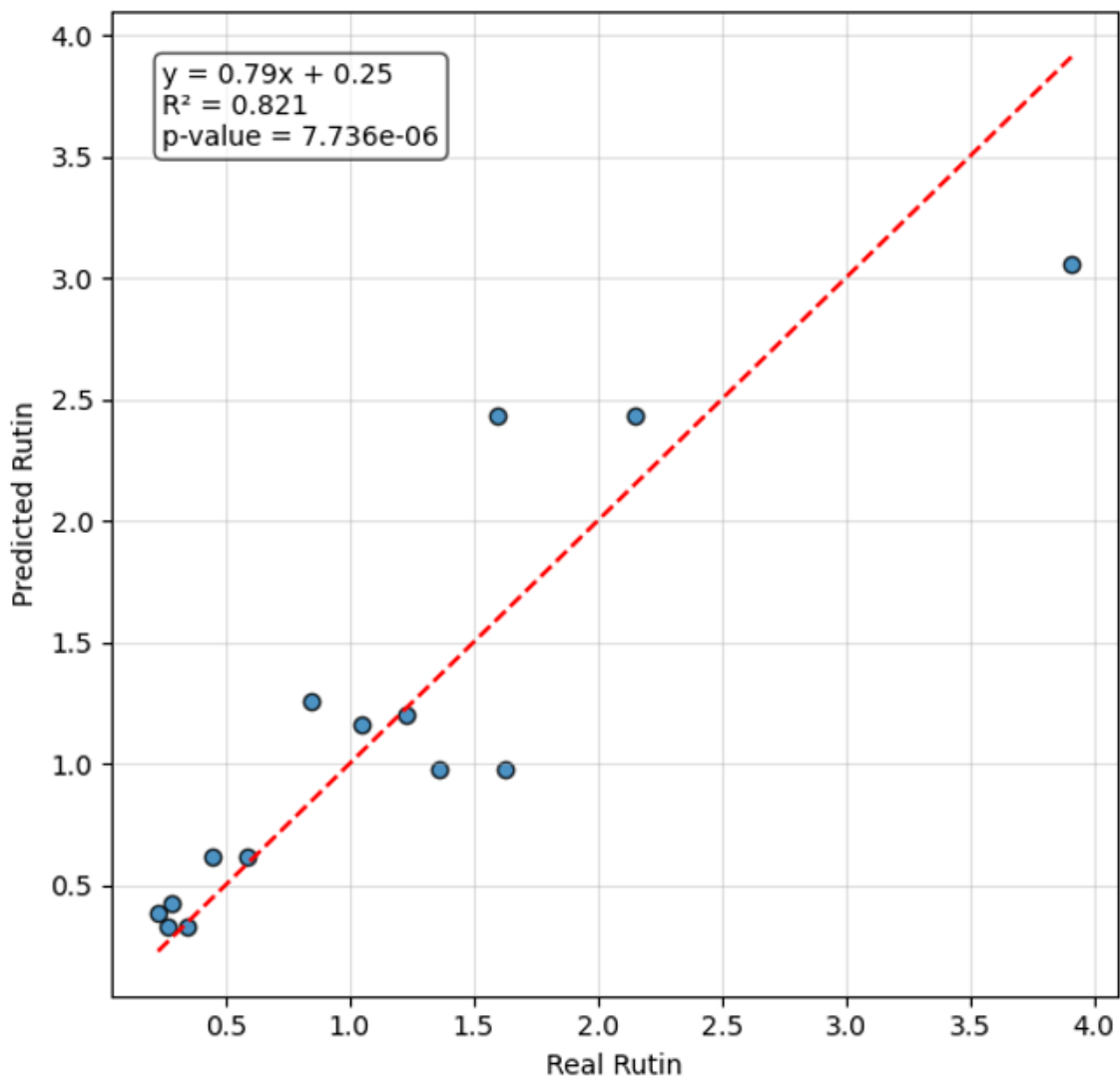


Figure 6

Scatter plots presenting the values of real vs predictions of rutin obtained by the random forest model.

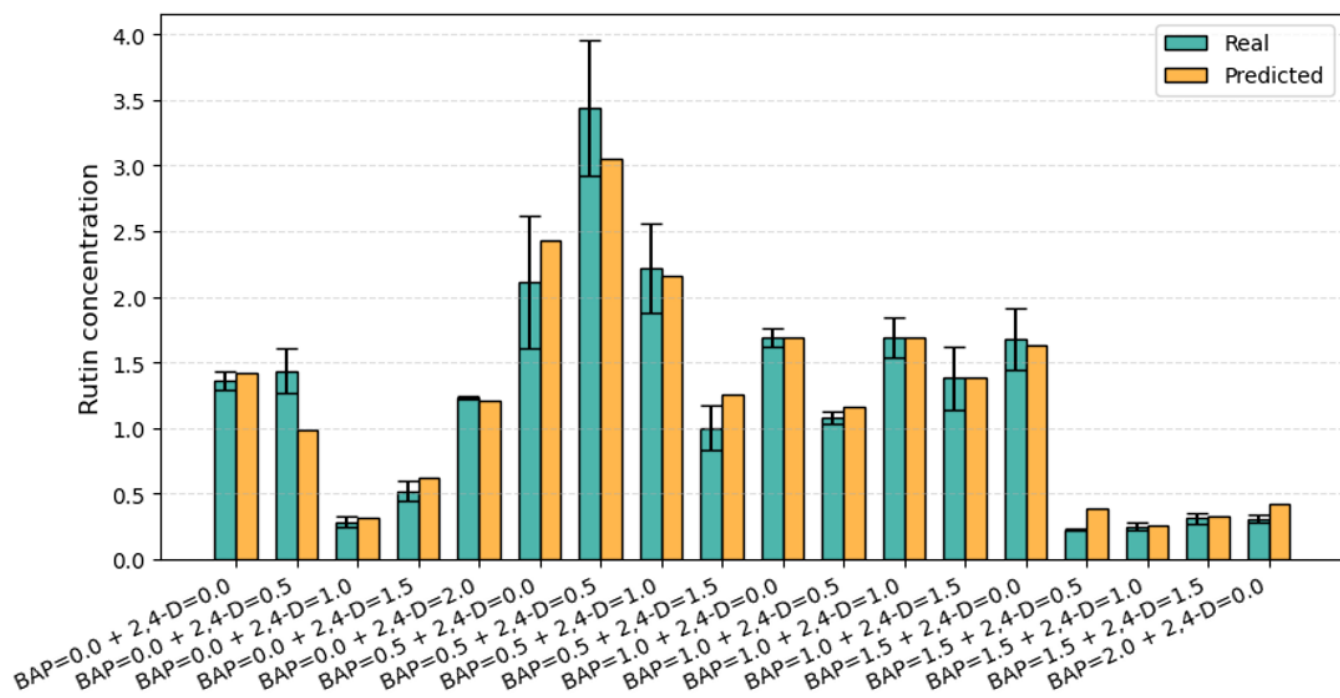


Figure 7

Real vs predicted accumulation of rutin using random forest model.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplemetarydata.docx](#)