

Impact of Internal and External Factors on Terpenoid Profiles in *Cannabis sativa* Leaves: Metabolomic Approach

Pattarawadee Sumthong Nakmee^{1,*}, Jannong Tanyasit¹, Panor Ruaysoongnoen²,
Boonorm Chomtee³, Chalothon Chootong⁴, Soontree Khuntong⁵,
Robert Verpoorte⁶ and Chatchai Kasemtaweekhok⁴

¹Department of Resources and Environment, Faculty of Science at Sriracha, Kasetsart University, Sri Racha Campus, Chonburi 20230, Thailand

²Department of Basic Science and Physical Education, Faculty of Science at Sriracha, Kasetsart University, Sri Racha Campus, Chonburi 20230, Thailand

³Department of Statistics, Faculty of Science, Kasetsart University, Bangkok 10900, Thailand

⁴Department of Computer Science and Information Technology, Faculty of Science at Sriracha, Kasetsart University, Sri Racha Campus, Chonburi 20230, Thailand

⁵Program in Safety Engineering and Environmental Management, Faculty of Engineering at Sri Racha, Kasetsart University, Bangkok 10900, Thailand

⁶Natural Products Laboratory, Institute Biology Leiden, Leiden University, Leiden, The Netherlands

(*Corresponding author's e-mail: pattarawadee.sum@ku.th)

Received: 4 November 2025, Revised: 11 December 2025, Accepted: 18 December 2025, Published: 20 March 2026

Abstract

Leaf maturity levels, cultivation plans, and growing seasons affected monoterpene and sesquiterpene production in *Cannabis sativa* leaves. Twenty-two terpenoid compounds were investigated from fresh leaves using headspace GC-MS. Among them, the major monoterpenes were α -pinene, β -myrcene, limonene, and β -ocimene, and the major sesquiterpenes were β -(E)-caryophyllene, (Z,E)- α -farnesene, β -bisabolene, (E)- α -bisabolene, aromadendrene, α -humulene, and α -bisabolol. All compounds showed significant correlations. However, all 3 factors influenced variation in 9 compounds, excluding α -pinene and β -pinene. β -(E)-caryophyllene, α -humulene, and β -ocimene were affected more by growing season than by cultivation plan or leaf maturity and were particularly abundant during the cool-dry season. Cultivation method (evaporative greenhouse, net greenhouse, or outdoor) significantly affected key terpenoids, including β -myrcene and limonene. Leaf maturity level significantly affected the production of aromadendrene and α -humulene. Machine learning algorithms were used to predict terpenoid levels based on internal and external factors. Model performance was evaluated using RMSE, R^2 , and correlation. The decision tree (DT) achieved the lowest prediction error for β -(E)-caryophyllene and was therefore the best model for this compound, whereas AdaBoost (ADA) performed best for α -humulene and 4 other major compounds.

Keywords: *Cannabis sativa*, Terpenoid, Caryophyllene, Secondary metabolite, Metabolomic, Machine learning, Smart Farming, Yield prediction, Decision tree, AdaBoost

Introduction

Cannabis sativa L. is an annual herb in the Cannabaceae family. Throughout history, *C. sativa* has been cultivated for fiber, medicine, food, supplements, and body care products. This plant is rich in trichomes,

hair-like glandular structures that cover the surfaces of leaves, petioles, stems, and flowers [1]. In traditional medicine, cannabis leaves are used for the treatment of sleep disorders, epilepsy, hemiplegia, paralysis,

hemorrhoids, and liver cancer. Leaves are also applied to wounds, cuts, skin diseases, and sores, either ground or crushed and applied as juice, paste, or powder [2]. Flavonoids and terpenoids in cannabis leaves can be used for anti-inflammatory, antibacterial, and antifungal purposes, to promote wound healing, and to treat diverse skin conditions [3].

C. sativa contains a unique group of natural compounds known as cannabinoids, each with diverse biological activities. For instance, CBD has been reported to have therapeutic effects in Parkinson's disease and epilepsy and to exert anti-inflammatory effects [4-9]. Beyond cannabinoids, other compound groups such as terpenes have been identified in *C. sativa* and exhibit various biological activities. Terpenes derived from *C. sativa* show pharmacological properties, including antimicrobial, anti-inflammatory, and antioxidant activities [10,11]. Specific terpenes such as β -caryophyllene and α -humulene demonstrate anti-inflammatory effects. Caryophyllene also displays gastroprotective effects and acts as an analgesic, antidepressant, anticancer, antifungal, antibacterial, and anti-inflammatory agent [12,13]. Cannabis terpenoids demonstrate anti-inflammatory effects *in vitro* and *in vivo*, and they provide short-term immunosuppression, which may be useful for managing acute inflammation [14]. The composition of secondary metabolites such as cannabinoids, terpenoids, and flavonoids varies across different parts of the *C. sativa* plant, including inflorescences, leaves, stems, and roots [15,16]. For example, inflorescences are used in TCM for acute pain, mania, insomnia, coughing, and wound healing; leaves are indicated for malaria, panting, roundworm infestation, scorpion stings, hair loss, hair graying, anti-inflammatory, fungicides, insecticides and food seasoning [17,18]; stem bark is used for strangury and physical injury; and roots are used for strangury, spotting, vaginal discharge, difficult births, retention of the placenta, and physical injury by oral or dermal administration [19].

Both primary and secondary metabolites in plants are affected by internal and external factors, including genetic variation and environmental conditions such as light intensity, photoperiod, temperature, drought, and moisture [20]. Several studies have emphasized the significant impact of temperature and moisture on the synthesis of plant secondary metabolites [21-27]. For

example, dry and windy conditions during the cool-dry season can elevate cannabinoid content in *C. sativa*, whereas it tends to lower during the hot-dry season. Additionally, total cannabinoid content depends on the plant developmental stage [28]. Light quality also influences the cannabinoid and terpenoid profiles in *C. sativa* inflorescences [29], and factors such as tropical day length, temperature, and genotype affect the cannabinoid profile and growth of hemp [30]. The light spectrum further affects the cannabinoid profile of medical cannabis inflorescences [31]. UV radiation affects terpenoid and phenolic accumulation in leaf [32]. High temperatures, especially during the grain-filling phase, can limit cannabis seed oil accumulation [33]. Moreover, significant variations in the metabolomic profile of cannabis occur across environments, with outdoor-grown inflorescences containing more cannabinoids than indoor-grown ones. Terpene profiles also differ significantly, with outdoor samples showing higher concentrations of sesquiterpenes such as β -caryophyllene, α -humulene, α -guaiene, and germacrene B relative to indoor samples [34]. Furthermore, the stage of plant development affects the production of secondary metabolites, including terpenoids [24]. As terpenoids may contribute to the biological activities of cannabis, it is important to investigate the factors that influence variation in these compounds.

Given the thousands of metabolites present in plants and the multiple factors involved in the biosynthesis and accumulation of secondary metabolites, ML and AI are promising approaches for the management and prediction of agricultural products in smart farming. ML has been used to identify conditions for decision making to improve yields. Resource efficiency, prediction, and smart farming management are important parameters for agricultural productivity [35-37]. ML algorithms combined with multidata sources enable quality evaluation and production control of medicinal plants [38,39]. ML can be used to identify conditions for predicting levels of terpenoids, cannabinoids, phenolic compounds, and flavonoids [40-42]. Leveraging ML within a metabolomic framework to predict the production of desired secondary metabolites for foods, medicines, cosmetics, and crop-protecting plants is an innovative approach that enables efficient production of high-

quality materials and enhances product consistency and value.

This study aims to explore the internal and external factors that influence terpenoid profiles in *C. sativa* leaves and to design an ML model for smart farming management. Terpenoids in *C. sativa* leaves can serve as a value-added resource for pharmaceutical and cosmeceutical product innovation.

Materials and methods

Plant materials

C. sativa (var. Hang Krarok) was cultivated in greenhouses at Kasetsart University, Chalermprakiat Campus, Chiang Khrua, Mueang, Sakon Nakhon, Thailand (17.283767, 104.107488) at 175 m elevation, under license number 6/2562 issued by the Narcotics Control Division, Food and Drug Administration, Ministry of Public Health, Thailand. Plant voucher specimen number BK086341. Leaf samples were collected before sunrise in 3 growing seasons (hot-dry, rainy, cool-dry) using 3 cultivation plans (evaporative greenhouse, net greenhouse, outdoor) and 3 leaf maturity levels (first, fifth, and seventh leaves from the shoot apex). Fresh leaves (1 g) were ground for each GC vial, tightly capped, and stored at 4 ± 1 °C prior to analysis.

Environmental data

Temperature, rainfall, relative humidity (RH), solar radiation (SR), and photosynthetically active radiation (PAR) for Mueang Sakon Nakhon (the cultivation region for *C. sativa*) were obtained from the Thai Meteorological Department station. Data were recorded daily for 50 days during each season, namely the hot-dry (March - April), rainy (May - June), and cool-dry (November - December) seasons. The net greenhouse provided 50% shading, whereas the evaporative greenhouse also provided 50% shading but reduced the internal temperature by 2 - 5 °C and increased relative humidity by 70% - 80% compared with outdoor conditions.

Headspace-GC/MS analysis

C. sativa leaf samples were analyzed under cannabis production and possession license numbers 21/2563 and 43/2563, respectively, issued by the Narcotics Control Division, Food and Drug

Administration, Ministry of Public Health, Thailand. Headspace (HS) analysis was performed using a TurboMatrix 40 Headspace Sampler (PerkinElmer, USA). Each 10 mL GC vial contained 1 g of ground *C. sativa* leaves and was tightly capped. Vials were placed in the HS autosampler tray, and the HS oven was programmed at 125 °C [43] to extract volatile compounds prior to injection into the GC-MS. The transfer line and needle temperatures were set to 130 °C.

GC-MS analysis of volatiles from *C. sativa* leaves was carried out using a Clarus 680 GC coupled to a Clarus SQ8C MS (PerkinElmer, USA) with a capillary column containing 5% phenylmethylpolysiloxane as the stationary phase (30 m \times 320 μ m i.d. \times 0.25 μ m film). The oven temperature was held at 60 °C, then increased at 7 °C min⁻¹ to 200 °C. The injector temperature was 200 °C. Helium was used as the carrier gas at 1 mL min⁻¹. Electron ionization mass spectra were acquired at 70 eV over m/z 45 - 550. The electron multiplier voltage was 1,400 V. The ion source and quadrupole temperatures were 200 °C. Linear temperature-programmed retention index (LTPRI) values were calculated from retention times. Compound identification was based on Kováts retention indices relative to C8-C22 *n*-paraffin standards and by comparison of mass spectra with NIST (2019) libraries and an in-house library of reference compounds analyzed under identical conditions. The percentage relative abundance was calculated as the peak area of each compound divided by the total peak area of each sample.

Statistical analysis

Data were analyzed using a factorial experiment with 3 replications, comprising 3 factors: Leaf maturity (first, fifth, seventh leaves from the shoot apex), cultivation plan (evaporative greenhouse, net greenhouse, outdoor), and growing season (hot-dry, rainy, cool-dry). Main effects and interactions on percentage relative abundances were assessed by analysis of variance, followed by Tukey's post hoc test. Results were considered significant at $p < 0.05$.

Prediction Model from ML algorithms

Machine learning focused on predicting 5 specific components derived from principal component analysis. Model development comprised feature selection, model

training and testing, and model selection. The dataset was constructed from experimental features (internal and external factors) and the percentage relative abundances of terpenoids. It contained 10 features, including growing season, cultivation plan, and leaf maturity level, along with other environmental variables. The target variables were the contents of terpenoids found in *C. sativa* leaves. Prediction performance using all features was evaluated.

Each model was tested using a 5-fold cross-validation procedure. Performance was assessed using 4 metrics: Root mean square error (RMSE), mean square error (MSE), coefficient of determination (R^2), and correlation. RMSE and MSE were computed using Eqs. (1) and (2), where lower values indicate better performance. R^2 indicates the proportion of variability explained by the model, with 1 indicating complete explanation and 0 indicating none. Correlation measures the strength and direction of the relationship between actual and predicted values and ranges from -1 to 1 , as shown in Eq. (4).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

$$MSE = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

$$Correlation = \frac{n \sum_{i=1}^n y_i \hat{y}_i - (\sum_{i=1}^n y_i)(\sum_{i=1}^n \hat{y}_i)}{\sqrt{(n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2)(n \sum_{i=1}^n \hat{y}_i^2 - (\sum_{i=1}^n \hat{y}_i)^2)}} \quad (4)$$

where n is the number of compounds in this experiment, y_i is the actual values of compound i , \bar{y} is the mean value, and \hat{y}_i is the predicted value of compound i .

Results and discussion

Environmental data

Rainfall data was mainly used for the classification of these 3 growing seasons. The data in **Table 1** indicated that SR and PAR were highest during the first growing season of the year (hot-dry season), rainfall and relative humidity were highest during the second growing season (rainy season), and the minimum temperature was lowest during the third growing season (cool-dry season).

Table 1 Average environmental data (mean \pm SD) for *C. sativa* growth in the hot-dry, rainy, and cool-dry seasons.

Season	Max. Temp. (°C)	Mean Temp. (°C)	Min. Temp. (°C)	Rainfall (mm)	RH (%)	SR (MJ/M ³ /day)	PAR (mol/m ² /day)
Hot-dry (Mar-Apr)	33.1 \pm 0.4	28.0 \pm 0.4	23.0 \pm 2.7	2.4 \pm 0.7	70.9 \pm 1.4	19.5 \pm 0.6	28.1 \pm 0.9
Rainy (May-Jun)	32.7 \pm 0.2	28.9 \pm 0.1	25.1 \pm 0.9	5.0 \pm 1.1	83.6 \pm 0.8	17.3 \pm 0.5	15.3 \pm 1.5
Cool-dry (Nov-Dec)	31.6 \pm 0.2	26.2 \pm 0.2	20.8 \pm 2.5	0.5 \pm 0.3	71.8 \pm 0.8	8.5 \pm 1.1	13.0 \pm 1.7

Headspace-GC/MS data

The chemical compositions of volatile compounds from *C. sativa* leaves collected at 3 leaf maturity levels, under 3 cultivation plans, and across 3 seasons are reported. Monoterpene and sesquiterpene hydrocarbons were present in all samples. The cool-dry season presented 21 terpenoids, followed by the hot-dry season (16) and the rainy season (12). The heatmap with hierarchical clustering of compounds based on their seasonal occurrence patterns is shown in **Figure 1**. Three major chemotype groups were observed across the seasons. Low-retention-index (low-RI) monoterpenes clustered together and occurred

predominantly in the cool-dry season. Mid-RI compounds remained relatively stable across all seasons. High-RI oxygenated sesquiterpenes formed a separate cluster and exhibited their highest abundance in the hot-dry and cool-dry seasons. Overall, the clustering indicates clear seasonal shifts, with lighter volatiles favored under cooler conditions and heavier compounds favored during hot-dry periods. Previous work [44] reported that heat stress induces monoterpene emissions more strongly than sesquiterpene emissions in certain plant species. The GC-MS chromatogram and MS spectrum of β -(E)-caryophyllene, the major compound in *C. sativa* leaves, are shown (**Figure 2**).

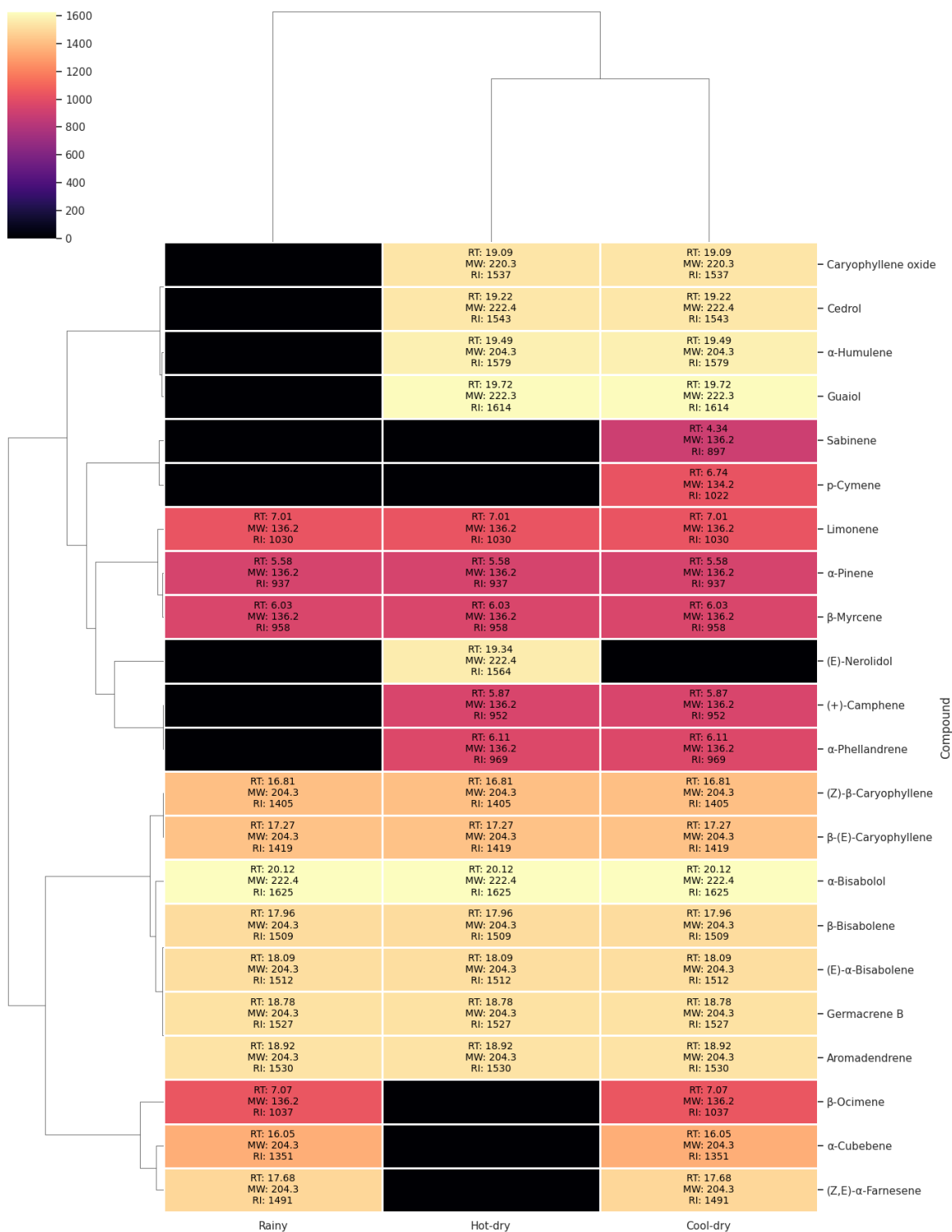


Figure 1 Hierarchical clustering of compounds illustrating co-occurrence patterns across seasons.

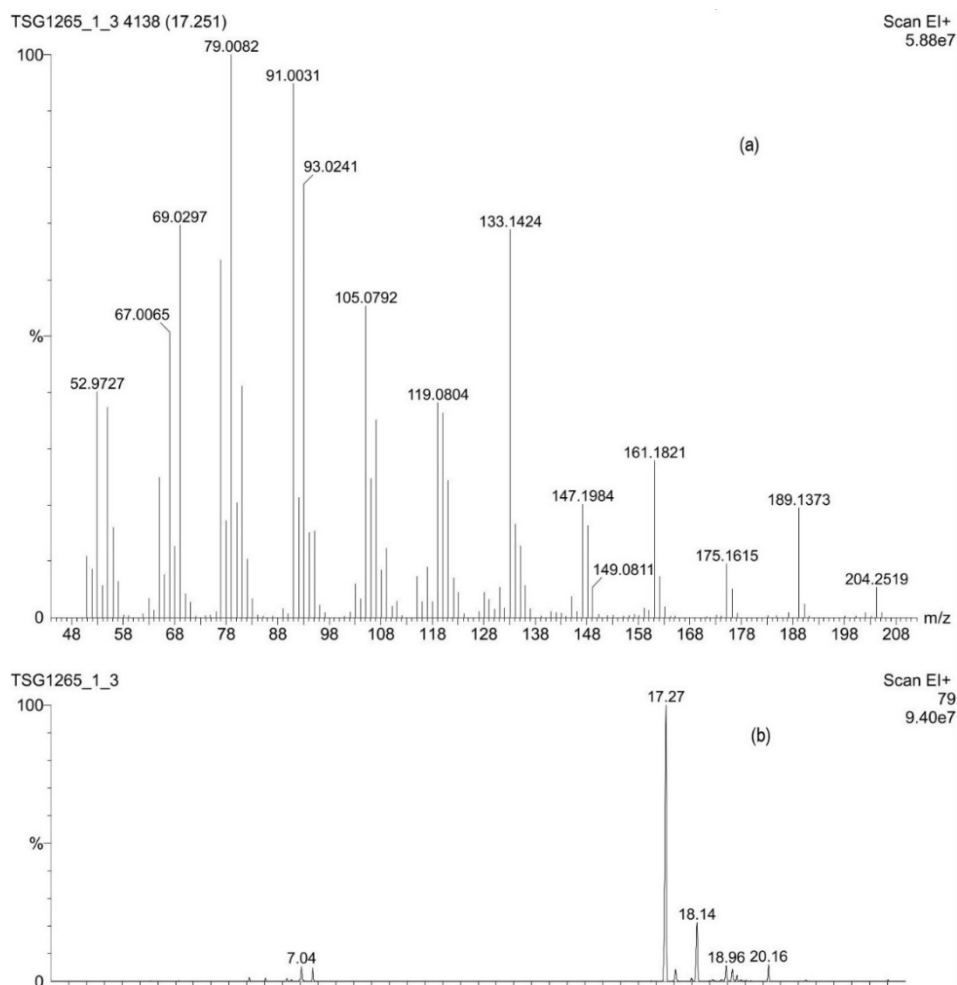


Figure 2 MS spectrum (a) and GC/MS chromatogram (b) of β -(E)-caryophyllene, a major terpenoid in *C. sativa* leaves.

Internal and external factors, including leaf maturity, cultivation plan, and season, affected variation in the terpenoid profile of *C. sativa* leaves. Variation in terpenoids was greater in the cool-dry season (November - December) than in the hot-dry (March - April) and the rainy season (May–June), when rainfall and humidity were lower and minimum temperature was lower. This indicates that temperature, moisture, and rainfall, which differed significantly among seasons, affected the production of terpenoids and other secondary metabolites [21-27]. Vapor pressure deficit (VPD) during the dry season, characterized by dry air and low relative humidity, influenced both stomatal conductance and the volatilization rates of mono- and sesquiterpenes. High VPD induced stomatal closure to reduce excessive water loss, thereby decreasing the evaporation of volatile compounds. Although volatilization is primarily driven by temperature, stomatal closure can still restrict the diffusion of volatile compounds, particularly oxygenated terpenoids [45,46].

Although SR and PAR directly support photosynthesis, the highest SR ($19.5 \pm 0.6 \text{ MJ m}^{-2} \text{ day}^{-1}$) and PAR ($28.1 \pm 0.9 \text{ mol m}^{-2} \text{ day}^{-1}$) observed in the hot-dry season corresponded to only moderate variation in terpenoids (**Table 1**). This may be due to direct sunlight exposure promoting volatilization of mono- and sesquiterpenes from leaves or to redistribution and accumulation in other plant organs [45].

The major terpenoids identified in *C. sativa* (var. Hang Krarok) leaves in every season were β -myrcene, limonene, (Z)- β -caryophyllene, β -(E)-caryophyllene, β -bisabolene, (E)- α -bisabolene, aromadendrene, and α -bisabolol (**Figure 1**). Several terpenoids, including myrcene and caryophyllene, were consistent with other *C. sativa* varieties [47-49]. However, α -pinene and β -pinene, which are common in other varieties, were not dominant here [50-52]. Conversely, camphene, cubebene, germacrene B, aromadendrene, cedrol, nerolidol, and guaiol were detected in ‘Hang Krarok’ but have not been reported for some other varieties, such as

an Iranian variety and the ‘Annapurna’ cultivar from Israel [53,54], although some of these compounds have been reported in *C. sativa* inflorescences [55,56]. These findings highlight a distinct terpenoid profile for ‘Hang Krarok’ relative to other varieties.

Terpenoid biosynthesis provides precursors for cannabinoid biosynthesis [32,49,57]. The mevalonate (MVA) pathway, originating from acetyl-CoA, produces C5 isoprenoid units, isopentenyl pyrophosphate (IPP) and its isomer dimethylallyl pyrophosphate (DMAPP). The methylerythritol phosphate (MEP) pathway, originating from pyruvate and glyceraldehyde 3-phosphate, also produces IPP and DMAPP [58,59]. For monoterpene biosynthesis, DMAPP condenses with IPP to form geranyl pyrophosphate (GPP, C10), which is then cyclized to yield acyclic or cyclic monoterpenes such as myrcene and limonene. For sesquiterpene biosynthesis, DMAPP sequentially condenses with 2 IPP units to form farnesyl pyrophosphate (FPP, C15), which is then cyclized to produce carbocation

intermediates that lead to compounds such as caryophyllene and humulene. The biosynthetic pathways of mono- and sesquiterpenes are illustrated (Figure 3). The MVA pathway primarily produced volatile sesquiterpenes, whereas the MEP pathway contributed precursors for volatile hemiterpenes, monoterpenes, and diterpenes. Terpenoid formation is regulated by terpene synthases (TPSs), the key enzymes producing C5 - C20 terpenoids. Many TPSs contain light-responsive regulatory elements, and light strongly influences terpenoid emissions by modulating photosynthetic carbon flux and the availability of MEP-derived precursors. In addition to light, temperature, water availability, drought, and air humidity can all modulate the biosynthesis, transport, and release of plant volatile compounds. However, further work is needed to elucidate the specific mechanisms by which temperature and water-related factors regulate these processes [60].

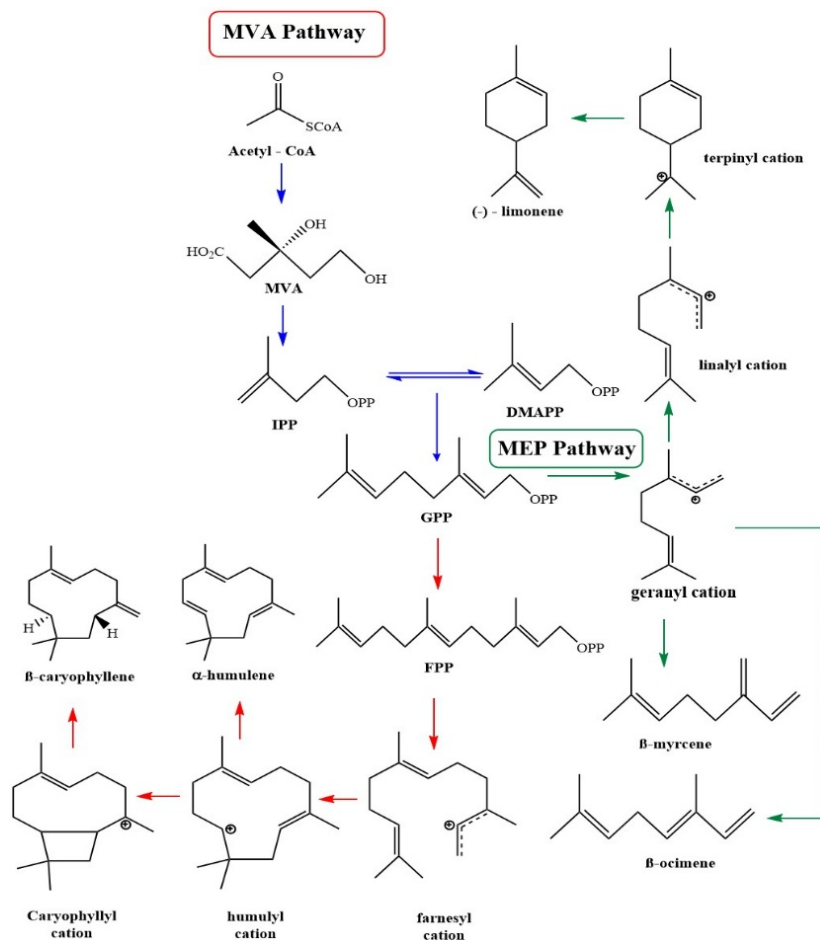


Figure 3 Biosynthesis pathway of mono- and sesquiterpenoids found in *Cannabis sativa*.

Interaction effects

Terpenoid constituents varied with leaf maturity, cultivation plan, and growing season (Figures 4 - 6). Interaction effects were significant for 9 of the eleven selected terpenoids, namely α -pinene, β -myrcene, limonene, β -ocimene, β -(E)-caryophyllene, α -farnesene, β -bisabolene, (E)- α -bisabolene, aromadendrene, α -humulene, and α -bisabolol, whereas α -pinene and β -pinene did not respond to all treatments. F-values and P-values from the ANOVA are reported in Table S1.

For β -myrcene, each factor had a significant effect, and all factors interacted significantly with its percentage relative abundance. The fifth leaf from the shoot apex showed the highest percentage relative abundance (1.921), followed by the seventh (1.478) and first (1.197) leaves (Figure 4). Thus, the fifth leaf was the recommended maturity stage for harvesting β -myrcene-rich leaves when compared with the first and seventh leaves. β -Myrcene, a common monoterpene in other *C. sativa* varieties [48,50], was higher in the evaporative greenhouse (2.179) than in the net greenhouse (1.473) and outdoor cultivation (0.944) (Figure 5). This could be attributed to the greenhouse covering blocking UV radiation, thereby increasing UV stress and the elicitation of plant secondary metabolites [32].

Limonene was affected by all 3 factors. The first leaf from the shoot apex showed the highest percentage relative abundance (0.698), followed by the fifth (0.538) and seventh (0.443) leaves. The evaporative greenhouse showed the highest percent relative abundance of limonene (0.983), followed by the net greenhouse (0.462) and outdoor cultivation (0.235) (Figure 5).

Seasonally, limonene was highest during the hot-dry season (0.783), followed by the cool-dry (0.567) and rainy (0.330) seasons.

Cultivation plan and growing season affected β -ocimene accumulation. Outdoor cultivation resulted in higher β -ocimene than the evaporative and net greenhouses. The highest β -ocimene occurred during the cool-dry season, followed by the rainy and hot-dry seasons (Figure 6).

Caryophyllene, a major sesquiterpene in *C. sativa* with anti-inflammatory and antimicrobial activity [12,13], was higher in the first leaf, whereas the fifth and seventh leaves did not differ significantly (Figure 4). The net greenhouse showed the highest β -(E)-caryophyllene (62.805), followed by outdoor cultivation (57.819) and the evaporative greenhouse (50.765) (Figure 5). Growth during the cool-dry season presented the highest percent relative abundance of β -(E)-caryophyllene (64.349), followed by the rainy (58.165) and hot-dry (48.875) seasons (Figure 6).

Aromadendrene was affected by all 3 factors. The seventh leaf from the shoot apex showed the highest value (6.379), followed by the fifth (5.658) and first (4.681) leaves (Figure 4). The rainy season had a higher percentage relative abundance (6.093) than the hot-dry (5.513) and cool-dry (5.113) seasons.

α -Humulene was also affected by all 3 factors. The fifth (0.097) and seventh (0.110) leaves had higher values than the first leaf (0.074). The evaporative greenhouse showed the highest α -humulene (0.169), followed by the net greenhouse (0.061) and outdoor cultivation (0.051). Seasonally, α -humulene was highest during the cool-dry season (0.196), followed by the hot-dry (0.085) and rainy (0.000) seasons (Figure 6).

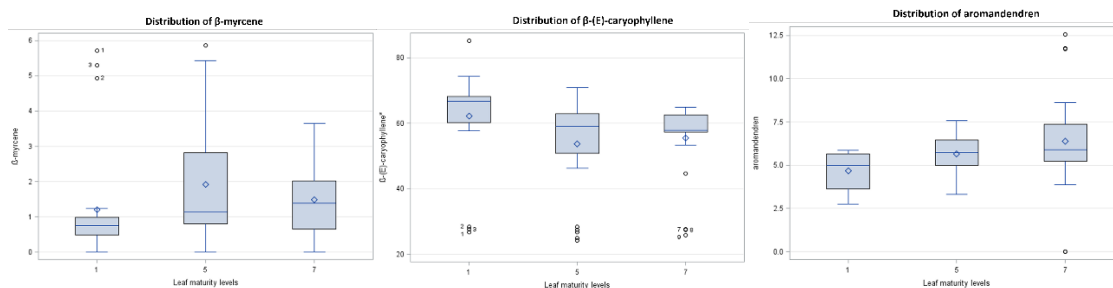


Figure 4 Data distribution of percentage relative abundance of β -myrcene, β -(E)-caryophyllene, and aromadendrene at different leaf maturity levels: First (1), fifth (5), and seventh (7) leaves from the shoot apex.

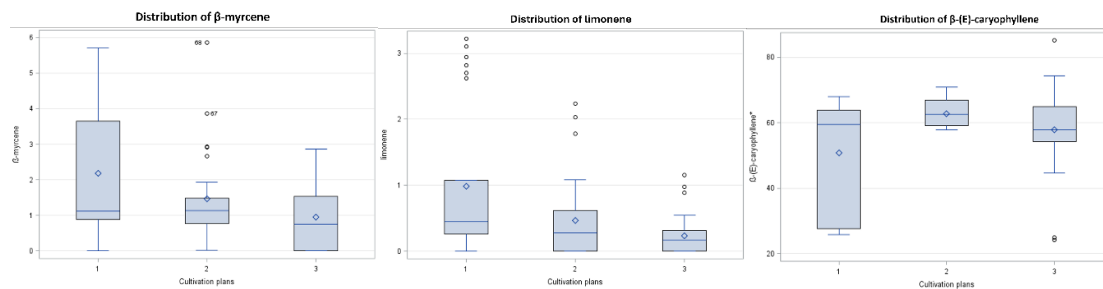


Figure 5 Data distribution of percentage relative abundance of β -myrcene, limonene, and β -(E)-caryophyllene across cultivation plans: Evaporative greenhouse (1), net greenhouse (2), and outdoor growth (3).

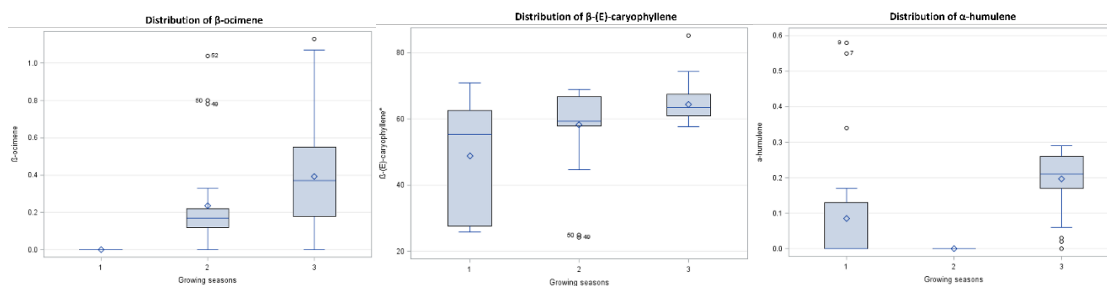


Figure 6 Data distribution of percentage relative abundance of β -ocimene, β -(E)-caryophyllene, and α -humulene across growing seasons: Hot-dry (1), rainy (2), and cool-dry (3).

This interaction analysis provides the first evidence that terpenoid variation in *C. sativa* leaves is influenced by the combined effects of leaf maturity, cultivation plan, and growing season. These factors should therefore be considered when cultivating and harvesting *C. sativa* for terpenoid extraction to support value-added applications. Previous work has indicated that certain terpenes form at later shoot developmental stages and that environmental factors affect secondary metabolic profiles in other species, such as *Tithonia diversifolia* [24].

Terpenoids are widely used in foods, medicines, flavors, and fragrances. Although *C. sativa* inflorescences have received the most attention, leaves should not be considered waste, but the leaves constitute the second most consumed part of this plant [18]. Leaf terpenoids can serve as sources for anti-inflammatory and antimicrobial products, for example β -caryophyllene and α -humulene [12-14], and as natural

volatile inputs for industry. Enhancing terpenoid yield in agriculture by targeting factors that influence their synthesis and accumulation in *C. sativa* leaves is therefore of practical interest.

Machine learning results

Model selection was based on RMSE, MSE, R^2 , and correlation. To evaluate the prediction performance of 4 ML models, MSE and RMSE were computed. RMSE is more interpretable than MSE because it is expressed in the same units as the target variable. RMSEs for all compound predictions from the 4 models are shown (Figure 7). AdaBoost (ADA) achieved the lowest RMSE for α -humulene and limonene, whereas the multilayer perceptron (MLP) outperformed the other methods for β -ocimene and β -myrcene. The decision tree (DT) achieved the lowest prediction error for β -(E)-caryophyllene.

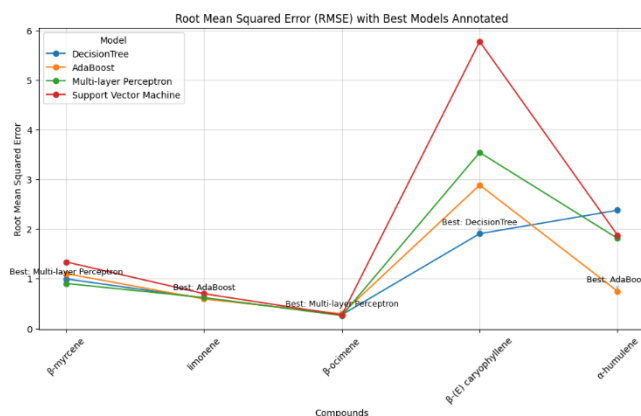
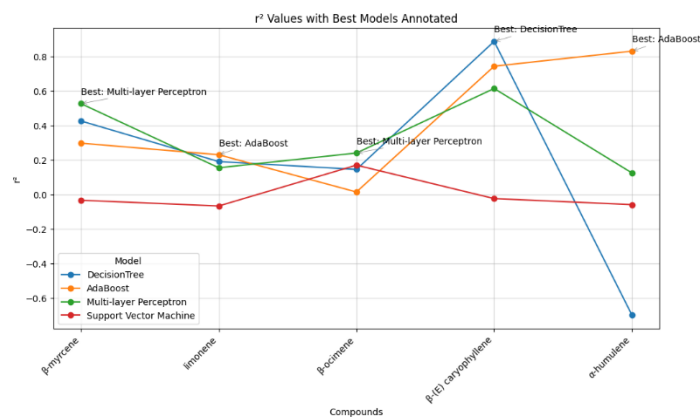
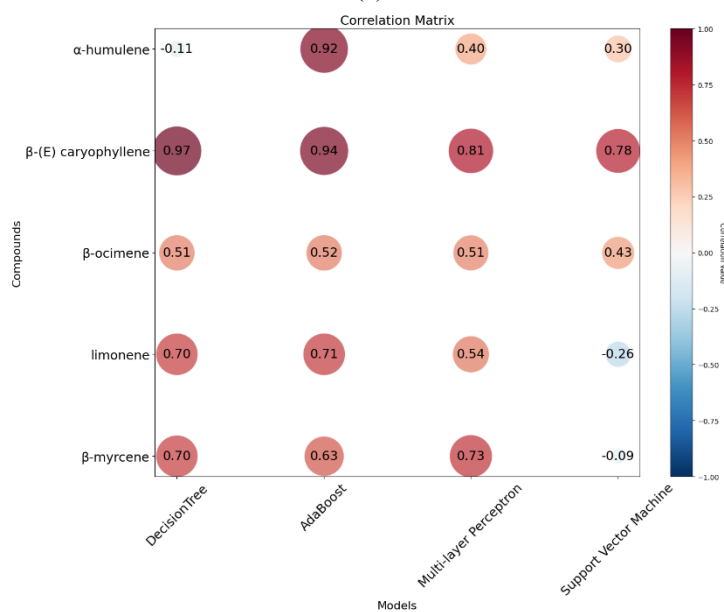


Figure 7 RMSE of 5 major terpenoid compounds from *C. sativa* leaves predicted by 4 models. The R^2 values showed the same pattern as the prediction errors (**Figure 8(a)**). ADA provided the highest correlation for almost all compounds except β -myrcene. DT presented the best correlation for β -(E)-caryophyllene (**Figure 8(b)**). MLP achieved the lowest error for β -myrcene and β -ocimene and the highest correlation for β -myrcene, whereas ADA was the best model for limonene and α -humulene.



(a)



(b)

Figure 8 R^2 (a) and correlation (b) of 5 major terpenoids from *C. sativa* leaves predicted by 4 models.

Table 2 Average RMSE, R², and correlation of 5 major terpenoids predicted by best models.

Compounds	Best models	RMSE	R ²	Correlation
β-Myrcene	MLP	0.905 ± 0.391	0.527 ± 0.193	0.726 ± 0.161
Limonene	ADA	0.594 ± 0.560	0.231 ± 0.369	0.706 ± 0.269
β-Ocimene	MLP	0.259 ± 0.071	0.241 ± 0.718	0.515 ± 0.385
β-(E)-Caryophyllene	DT	1.907 ± 0.548	0.888 ± 0.558	0.971 ± 0.252
α-Humulene	ADA	0.749 ± 0.075	0.831 ± 0.543	0.922 ± 0.560

The average RMSE, R², and correlation values for 5 major terpenoids predicted by the best-performing models are shown in **Table 2**. The MLP and ADA models achieved the lowest RMSE and highest R² for 2 major terpenoids. The average R² values for 5 major terpenoids predicted by each model were 0.391 ± 0.415 (DT), 0.550 ± 0.289 (ADA), 0.240 ± 0.312 (MLP), and 0.012 ± 0.083 (SVM). These results indicated that the ADA model provided the best predictions for 5 major terpenoids followed by DT model. Overall, ADA appears to be the most suitable model across the 5 compounds. DT and ADA are widely used in agricultural applications [61-63]. In smart farming management, diverse features, including internal plant factors and external environmental variables, can be incorporated to predict various groups of secondary metabolite profiles. Previous reports indicate that ML can enhance the production of secondary metabolites such as phenolics and flavonoids [41,42]. This approach offers a promising strategy for decision making on terpenoid profile, particularly monoterpenoids and sesquiterpenoids in *C. sativa* leaves, adding value to agricultural production and facilitating transfer of farming products to pharmaceutical and cosmeceutical industries. Despite relying on relative abundance as semi-quantitative data, this approach was adequate for distinguishing key terpenoids and comparing terpenoid profile. The integration of metabolomics and ML can inform adaptive cultivation approaches, from microclimate management in controlled systems to the optimization of harvest stage for targeted terpenoid profiles. Thus, smart farming technology is increasingly valued for managing processes, enhancing efficiency, and ensuring sustainability.

Conclusions

The terpenoid profile of *C. sativa* (var. Hang Krarok) leaves is influenced by internal and external factors, namely leaf maturity, cultivation plan, and growing season. Caryophyllene is the most abundant terpenoid. α-Humulene and β-ocimene are affected more by growing season than by cultivation plan or leaf maturity and are particularly abundant during the cool-dry season. β-Myrcene and limonene are higher under greenhouse conditions with evaporative cooling and net structures than under outdoor cultivation. Leaf maturity, particularly the fifth and seventh leaves from the shoot apex, significantly affects the production of aromadendrene and α-humulene. Overall, all 3 factors affect terpenoid profile in *C. sativa* leaves. β-(E)-Caryophyllene and α-humulene, both sesquiterpenes, are closely related in their biosynthesis and accumulation in *C. sativa* leaves relative to other sesquiterpenes. Likewise, the monoterpenes β-myrcene, limonene, and β-ocimene are closely linked within the biosynthetic pathway. Terpenes serve as key intermediates in cannabinoid biosynthesis and are valuable ingredients for natural pharmaceutical and cosmeceutical product development. For smart farming applications that aim to enhance and predict terpenoid profile using model performance metrics (RMSE, R² and correlation), the DT model is the best predictor for β-(E)-caryophyllene, achieving the lowest prediction error with the highest correlation. ADA performs best for α-humulene and limonene and shows competitive performance for the remaining targets, while the MLP performs best for β-myrcene and β-ocimene. These results support the application of ML in farm management for optimizing secondary metabolite production.

Acknowledgements

This research was supported by funding from the Kasetsart University Sri Racha Campus, Chonburi, Thailand, under the 2024 fiscal year budget.

Declaration of Generative AI in Scientific Writing

No content generation or data interpretation was performed by AI. The authors take full responsibility for the content and conclusions of this work.

CRedit Author Statement

Pattarawadee Sumthong Nakmee: Contributed to conceptualization, methodology and wrote the manuscript. **Jamnong Tanyasit** and **Panor Ruaysoongnoen:** Investigated GC/MS analysis. **Boonorm Chomtee:** Performed statistical analysis. **Chatchai Kasemtawechok:** Implemented algorithms. **Chalothon Chootong:** Prepared visualizations. **Soontree Khuntong:** Validated experiments. **Robert Verpoorte:** Supervised the manuscript. All authors approved the final version.

References

- [1] F Fulvio, Y Pieracci, R Ascrizzi, L Bassolino, G Flamini and R Paris. Insights into terpenes profiling and transcriptional analyses during flowering of different *Cannabis sativa* L. chemotypes. *Phytochemistry* 2025; **229**, 114294.
- [2] M Balant, A Gras, M Ruz, J Vallès, D Vitales and T Garnatje. Traditional uses of Cannabis: An analysis of the CANNUSE database. *Journal of Ethnopharmacology* 2021; **279**, 114362.
- [3] CM Andre, JF Hausman and G Guerriero. *Cannabis sativa*: The Plant of the thousand and one molecules. *Frontiers in Plant Science* 2016; **7**, 19.
- [4] MA Leehey, Y Liu, F Hart, C Epstein, M Cook, S Sillau, J Klawitter, H Newman, C Sempio, L Forman, L Seeberger, O Klepitskaya, Z Baud and J Bainbridge. Safety and tolerability of cannabidiol in Parkinson disease: An open label, dose-escalation study. *Cannabis Cannabinoid Research* 2020; **5(4)**, 326-336.
- [5] F Patricio, AAM Andrade, AP Martinez and ID Limon. Cannabidiol as a therapeutic target: Evidence of its neuroprotective and neuromodulatory function in Parkinson's disease. *Frontiers in Pharmacology* 2020; **11**, 595635.
- [6] A Arzimanoglou, U Brandl, JH Cross, A Gil-Nagel and L Lagae. Epilepsy and cannabidiol: A guide to treatment. *Epileptic Disorders* 2020; **22(1)**, 1-14.
- [7] BK O'Connell, D Gloss and O Devinsky. Cannabinoids in treatment-resistant epilepsy: A review. *Epilepsy and Behavior* 2017; **70**, 341-348.
- [8] LK Blevins, AP Bach, RB Crawford, J Zhou, JE Henriquez, MD Rizzo, S Sermet, DMIO Khan, H Turner, ALS Howard and NE Kaminski. Evaluation of the anti-inflammatory effects of selected cannabinoids and terpenes from *Cannabis sativa* employing human primary leukocytes. *Food and Chemical Toxicology* 2022; **170**, 113458.
- [9] J Peng, M Fan, C An, F Ni, W Huang and J Luo. A narrative review of molecular mechanism and therapeutic effect of cannabidiol (CBD). *Basic and Clinical Pharmacology and Toxicology* 2022; **130(4)**, 439-456.
- [10] W Shebawy, J Saliba, WH Faour, J Ismail, ME Hage, CF Daher, RI Taleb, B Nehmeh, C Dagher, E Chrabieh and M Mroueh. *In vivo* and *in vitro* anti-inflammatory activity evaluation of Lebanese *Cannabis sativa* L. ssp. *Indica* (Lam.). *Journal of Ethnopharmacology* 2021; **270**, 113743.
- [11] S Palmieri, F Fanti, E Oliva, E Viteritti, M Sergi, A Pepe and D Compagnone. Chemical characterization and evaluation of antioxidant activity from different cultivars of *Cannabis sativa* L. of Abruzzo's region. *Natural Product Research* 2023; **37(15)**, 2591-2595.
- [12] F Francomano, A Caruso, A Barbarossa, A Fazio, C La Torre, J Ceramella, R Mallamaci, C Saturnino, D Iacopetta and MS Sinicropi. β -Caryophyllene: A sesquiterpene with countless biological properties. *Applied Science* 2019; **9(24)**, 5420.
- [13] MMH Viveiros, MG Silva, JGM da Costa, AG de Oliveira, C Rubio, CR Padovani, CA Rainho and SA Schellini. Anti-inflammatory effects of α -humulene and β -caryophyllene on pterygium fibroblasts. *International Journal of Ophthalmology* 2022; **15(12)**, 1903-1907.
- [14] R Gallily, Z Yekhtin and LO Hanuš. The anti-inflammatory properties of terpenoids from

- Cannabis*. *Cannabis Cannabinoid Research* 2018; **3(1)**, 282-290.
- [15] O Aizpurua-Olaizola, U Soydaner, E Ozturk, D Schibano, Y Simsir, P Navarro, N Etxebarria and A Usobiaga. Evolution of the cannabinoid and terpene content during the growth of *Cannabis sativa* plants from different chemotypes. *Journal of Natural Products* 2016; **79(2)**, 324-331.
- [16] JT Fishedick, A Hazekamp, T Erkelens, YH Choi and R Verpoorte. Metabolic fingerprinting of *Cannabis sativa* L., cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. *Phytochemistry* 2010; **71**, 2058-2073.
- [17] EB Russo and J Marcu. Chapter three - Cannabis pharmacology: The usual suspects and a few promising leads. *Advances in Pharmacology* 2017; **80**, 67-134.
- [18] E Fordjour, CF Manful, AA Sey, R Javed, TH Pham, R Thomas and M Cheema. *Cannabis*: A multifaceted plant with endless potentials. *Frontiers in Pharmacology* 2023; **14**, 1200269.
- [19] D Jin, K Dai, Z Xie and J Chen. Secondary metabolites profiled in cannabis inflorescences, leaves, stem barks, and roots for medicinal purposes. *Scientific Reports* 2020; **10**, 3309.
- [20] J Petit, EMJ Salentijn, MJ Paulo, C Thouminot, BJ van Dinter, G Magagnini, HJ Gusovius, K Tang, S Amaducci, S Wang, B Uhrlaub, J Mussg and LM Trindade. Genetic variability of morphological, flowering and biomass quality traits in Hemp (*Cannabis sativa* L.). *Frontiers in Plant Science* 2020; **20(11)**, 102.
- [21] Q Duan, A Kleiber, K Jansen, LVJ Frohn, B Kammerer, G Han, I Zimmer, H Rennenberg, JP Schnitzler, I Ensminger, A Gessler and J Kreuzwieser. Effects of elevated growth temperature and enhanced atmospheric vapour pressure deficit on needle and root terpenoid contents of 2 Douglas fir provenances. *Environmental and Experimental Botany* 2019; **166**, 103819.
- [22] B Kharel, L Rusalepp, B Bhattarai, A Kaasik, P Kupper, R Lutter, P Mänd, G Rohula-Okunev, K Rosenvald and A Tullus. Effects of air humidity and soil moisture on secondary metabolites in the leaves and roots of *Betula pendula* of different competitive status. *Oecologia* 2023; **202**, 193-210.
- [23] R Kumar, M Kaundal, S Sharma, M Thakur, N Kumar, T Kaur, D Vyas and S Kumar. Effect of elevated [CO₂] and temperature on growth, physiology and essential oil composition of *Salvia sclarea* L. in the western Himalayas. *Journal of Applied Research on Medicinal and Aromatic Plants* 2017; **6**, 22-30.
- [24] Y Li, D Kong, Y Fu, MR Sussman and H Wu. The effect of developmental and environmental factors on secondary metabolites in medicinal plants. *Plant Physiology and Biochemistry* 2020; **148**, 80-89.
- [25] M Mahajan, R Kuiry and PK Pal. Understanding the consequence of environmental stress for accumulation of secondary metabolites in medicinal and aromatic plants. *Journal of Applied Research on Medicinal and Aromatic Plants* 2020; **18**, 100255.
- [26] G Prinsloo and N Nogemane. The effects of season and water availability on chemical composition, secondary metabolites and biological activity in plants. *Phytochemistry Reviews* 2018; **17**, 889-902.
- [27] MM Qaderi, AB Martel and CA Strugnell. Environmental factors regulate plant secondary metabolites. *Plants* 2023; **12(3)**, 447.
- [28] D Pasifico, F Miselli and A Carboni. Time course of cannabinoid accumulation and chemotype development during the growth of *Cannabis sativa* L. *Euphytica* 2008; **160**, 231-240.
- [29] V Morello, VD Brousseau, N Wu, BS Wu, S MacPherson and M Lefsrud. Light quality impacts vertical growth rate, phytochemical yield and cannabinoid production efficiency in *Cannabis sativa*. *Plants* 2022; **11(21)**, 2982.
- [30] LD Prato, M Timmins, O Ansari, KX Ruthrof, GESJ Hardy, J Howieson and G O'Hara. Semi-quantitative analysis of cannabinoids in hemp (*Cannabis sativa* L.) using gas chromatography coupled to mass spectroscopy. *Industrial Crops and Products* 2022; **178**, 114605.
- [31] N Danziger and N Bernstein. Light matters: Effect of light spectra on cannabinoid profile and plant development of medical cannabis (*Cannabis sativa* L.). *Industrial Crops and Products* 2021; **164**, 113351.

- [32] VD Brousseau, BS Wu, S MacPherson, V Merollo and M Lefsrud. Cannabinoids and terpenes: How production of photo-protectants can be manipulated to enhance *Cannabis sativa* L. Phytochemistry. *Frontiers in Plant Science* 2021; **12**, 620021.
- [33] M Baldini. Suitability assessment of different hemp (*Cannabis sativa* L.) varieties to the cultivation environment. *Industrial Crops and Products* 2020; **143**, 111860.
- [34] F Zandkarimi, J Cecatur, J Casali, T Gordon, C Skibola and C Nuckolls. Comparison of the cannabinoid and terpene profiles in commercial cannabis from natural and artificial cultivation. *Molecule* 2023; **28(2)**, 833.
- [35] I Attri, LK Awasthi and TP Sharma. Machine learning in agriculture: A review of crop management applications. *Multimedia Tools and Applications* 2024; **83**, 12875-12915.
- [36] S Mourtzinis, PD Esker, JE Specht and SP Conley. Advancing agricultural research using machine learning algorithms. *Scientific Reports* 2021; **11**, 17879.
- [37] FS Prity, MDM Hasan, SH Saif, MDM Hossain, SH Bhuiyan, MDA Islam and MDTH Lavlu. Enhancing agricultural productivity: A machine learning approach to crop recommendations. *Human-Centric Intelligent Systems* 2024; **6(12)**, 394-403.
- [38] Y Zhang and Y Wang. Recent trends of machine learning applied to multi-source data of medicinal plants. *Journal of Pharmaceutical and Biomedical Analysis* 2023; **13**, 1388-1407.
- [39] DV Nazarenko, IA Rodin and OA Shpigun. The use of machine learning in the analytical control of the preparations of medicinal plants. *Inorganic Materials* 2019; **55(14)**, 1428-1438.
- [40] SK Abeysekera, A Robinson, MPL Ooia, YC Kuang, M Manley-Harris, W Holmes, E Hirst, J Nowak, M Caddie, G Steinhorn and S Demidenko. Sparse reproducible machine learning for near infrared hyperspectral imaging: Estimating the tetrahydrocannabinolic acid concentration in *Cannabis sativa* L. *Industrial Crops and Products* 2023; **192**, 116137.
- [41] P García-Pelez, E Lozano-Milo, PP Gallego and M Landín. Combining medicinal plant in vitro culture with machine learning technologies for maximizing the production of phenolic compounds. *Antioxidants* 2020; **9(3)**, 210.
- [42] K Kusumiyati and Y Asikin. Machine learning-based prediction of total phenolic and flavonoid in horticultural products. *Open Agriculture* 2023; **8**, 20220163.
- [43] Z Triaux, H Petitjean, E Marchioni, D Steyer and C Marcic. Comparison of headspace, hydrodistillation and pressurized liquid extraction of terpenes and terpenoids from food matrices-qualitative and quantitative analysis. *Journal of Analytical Chemistry* 2021; **76(3)**, 284-295.
- [44] S Nagalingam, R Seco, S Kim and A Guenther. Heat stress strongly induces monoterpene emissions in some plants with specialized terpenoid storage structures. *Agricultural and Forest Meteorology* 2023; **333**, 109400.
- [45] SD Tetali. Terpenes and isoprenoids: A wealth of compounds for global use. *Planta* 2019; **249**, 1-8.
- [46] ME Bergman, XQ Huang, S Baudino, JC Caissard and N Dudareva. Plant volatile organic compounds: Emission and perception in a changing world. *Current Opinion in Plant Biology* 2025; **85**, 102706.
- [47] DS Wishart, MH Giesbrecht, G Inchehborouni, X Cao, AC Guo, MA LeVatte, C Torres-Calzada, V Gautam, M Johnson, J Liigand, F Wang, S Zahraei, S Bhumireddy, Y Wang, J Zheng, R Mandal and JRB Dyck. Chemical composition of commercial cannabis. *Journal of Agricultural and Food Chemistry* 2024; **72**, 14099-14113.
- [48] EM Mudge, PN Brown and SJ Murch. The Terroir of Cannabis: Terpene metabolomics as a tool to understand *Cannabis sativa* selections. *Planta Medica* 2019; **85(9-10)**, 781-796.
- [49] JK Booth, MMS Yuen, S Jancsik, LL Madilao, JE Page and J Bohlmann. Terpene synthases and terpene variation in *Cannabis sativa*. *Plant Physiology* 2020; **184(1)**, 130-147.
- [50] JK Booth and J Bohlmann. Terpene in *Cannabis sativa* from plant genome to humans. *Plant Science* 2019; **284**, 67-72.
- [51] JT Fishedick. Identification of terpenoid chemotypes among high (-)-trans-D9-tetrahydrocannabinol-producing *Cannabis sativa*

- L. cultivars. *Cannabis Cannabinoid Research* 2017; **2(1)**, 34-47.
- [52] S Gaggiotti, S Palmieri, FD Pelle, M Sergi, A Cichelli, M Mascini and D Compagnone. Piezoelectric peptide-hpDNA based electronic nose for the detection of terpenes; Evaluation of the aroma profile in different *Cannabis sativa* L.(hemp) samples. *Sensors and Actuators, B: Chemical* 2020; **308**, 127697.
- [53] H Mansouri, Z Asrar and R Amarowicz. The response of terpenoids to exogenous gibberellic acid in *Cannabis sativa* L. at vegetative stage. *Acta Physiologiae Plantarum* 2011; **33**, 1085-1091.
- [54] A Saloner and N Bernstein. Nitrogen supply affects cannabinoid and terpenoid profile in medical cannabis (*Cannabis sativa* L.). *Industrial Crops and Products* 2021; **167**, 113516.
- [55] J Raeber, M Poetzsch, A Schmidli, S Favrod and C Steuer. Simultaneous quantification of terpenes and cannabinoids by reversed-phase LC-APCI-MS/MS in *Cannabis sativa* L. samples combined with a subsequent chemometric analysis. *Analytical and Bioanalytical Chemistry* 2024; **416(18)**, 4193-4206.
- [56] MM Radwan, S Chandra, S Gul and MA ElSohly. Cannabinoids, phenolics, terpenes and alkaloids of *Cannabis*. *Molecules* 2021; **26**, 2774.
- [57] MN Tahir, F Shahbazi, S Rondeau-Gagné and JF Trant. The biosynthesis of the cannabinoids. *Journal of Cannabis Research* 2021; **3**, 7.
- [58] MP Dewick. *Medicinal natural products*. 2nd ed. John Wiley and Ltd Sons, Ltd., West Sussex, 2001, p. 167-202.
- [59] BM Lange and N Srividya. Cannabis monoterpene synthases: Evaluating structure-function relationships. *Phytochemistry Reviews* 2023; **22**, 449-465.
- [60] R Escobar-Bravo, PA Lin, JM Waterman and M Erb. Dynamic environmental interactions shaped by vegetative plant volatiles. *Natural Product Report* 2023; **40**, 840-865.
- [61] AK Apat, J Mishra, KS Raju and N Padhy. The robust and efficient machine learning model for smart farming decisions and allied intelligent agriculture decisions. *Journal of Integrated Science and Technology* 2022; **10(2)**, 139-155.
- [62] A Kaur, GS Randhawa, F Abbas, M Ali, TJ Esau, AA Farooque and R Singh. Artificial intelligence driven smart farming for accurate detection of potato diseases: A systematic review. *IEEE Access* 2024; **12**, 193902-193922.
- [63] M Mohy-eddine, A Guezzaz, S Benkirane and M Azrou. Malicious detection model with artificial neural network in IoT-based smart farming security. *Cluster Computing* 2024; **27**, 7307-7322.

Supplementary Material

Table S1 F-values and *p*-values for the main effects and interaction effects in the factorial ANOVA.

Factors	α -pinene	β -myrcene	limonene	β -ocimene	β -(E)-caryophyllene	(Z,E)- α -farnesene	β -bisabolene	(E)- α -bisabolene	aromaden drene	α -humulene	α -bisabolol
S	4.98* (01.0103)	6.00* (0.0044)	57.74* (< 0.0001)	24.15* (< 0.0001)	93.64* (< 0.0001)	327.83* (< 0.0001)	36.48* (< 0.0001)	1.63ns	8.35* (0.0007)	169.41* (< 0.0001)	0.45 ns
P	4.26* (0.0191)	27.73* (< 0.0001)	165.23* (< 0.0001)	4.14* (0.0213)	56.49* (< 0.0001)	12.98* (< 0.0001)	32.71* (< 0.0001)	14.61* (< 0.0001)	3.05 ns	75.38* (< 0.0001)	9.21* (0.0004)
M	22.46* (< 0.0001)	9.61* (0.0003)	18.56* (< 0.0001)	0.82 ns	31.62* (< 0.0001)	40.24* (< 0.0001)	7.45* (0.0014)	0.77ns	24.96* (< 0.0001)	5.71* (0.0056)	1.83 ns
S*P	8.76* (< 0.0001)	26.85* (< 0.0001)	274.70* (< 0.0001)	1.64 ns	74.04* (< 0.0001)	11.94* (< 0.0001)	76.44* (< 0.0001)	29.88* (< 0.0001)	12.01* (< 0.0001)	34.18* (< 0.0001)	105.22* (< 0.0001)
S*M	2.36 ns	20.40* (< 0.0001)	36.26* (< 0.0001)	0.59 ns	8.87* (< 0.0001)	39.84* (< 0.0001)	7.59* (< 0.0001)	4.93* (0.0019)	13.26* (< 0.0001)	15.04* (< 0.0001)	5.26* (0.0012)
P*M	1.08 ns	8.34* (< 0.0001)	18.03* (< 0.0001)	1.56 ns	6.47* (0.0003)	18.48* (< 0.0001)	7.48* (< 0.0001)	2.57* (0.0484)	15.38* (< 0.0001)	15.22* (< 0.0001)	15.15* (< 0.0001)
S*P*M	0.81 ns	10.62* (< 0.0001)	38.48* (< 0.0001)	1.23 ns	4.49* (0.0003)	18.34* (< 0.0001)	7.01* (< 0.0001)	4.66* (0.0002)	14.12* (< 0.0001)	10.35* (< 0.0001)	19.44* (< 0.0001)

Remarks: S = growing seasons, P = cultivation plans, M = leaf maturity levels, "*" is significant at $\alpha = .05$, "ns" is not significant at $\alpha = .05$, (*p*-value).