Grasas y Aceites 76 (2)
April-June 2025, 2301
ISSN-L: 0017-3495, eISSN: 1988-4214
https://doi.org/10.3989/gya.0320251.2301

Machine learning analysis of fatty acid composition in *washingtonia filifera* and *sabal palmetto* palm kernels for potential industrial applications

T. Bozkurt[a,✉]

[a]Tekfen Agricultural Research Production and Marketing Inc., Adana, Türkiye
✉Corresponding author: 01tanerbozkurt@gmail.com

Submitted: 19 March 2025; Accepted: 09 June 2025; Published: 03 February 2026

SUMMARY: The fruits of many palm species, including *Washingtonia filifera* (*W. filifera*) and *Sabal palmetto* (*S. palmetto*), are rich in secondary metabolites and widely used for food and industrial purposes. This study aimed to analyze the fatty acid profiles of the kernels of these two underutilized palm species and to develop predictive models using machine learning to assess their potential for diverse industrial applications. Oleic acid was the most abundant fatty acid, constituting 37.13% in *W. filifera* and 33.29% in *S. palmetto*. Lauric acid followed at 25.80% in *W. filifera* and 25.57% in *S. palmetto*. Linoleic and myristic acids were also prevalent, with varying ranks between the two species. Total fat, total Saturated Fatty Acids (ΣSFA), total Monounsaturated Fatty Acids (ΣMUFA), and total Polyunsaturated Fatty Acids (ΣPUFA) were analyzed using machine learning (ML) models. The performances of Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP) models were evaluated using metrics like RMSE, R^2 score, and MAE. SVM achieved the highest R^2 scores (0.98-0.99), demonstrating its effectiveness in accurately predicting fatty acid profiles, which is crucial for assessing their suitability for various industrial uses, including food, biodiesel and potential applications in the cosmetic industry.

KEY-WORDS: Arecaceae*; Artificial intelligence; Kernel oil biodiesel; Machine Learning; Palm oil.*

RESUMEN: ***Aprendizaje automático del análisis de la composición de ácidos grasos en semillas de palma* washingtonia filifera *y* sabal palmetto *para posibles aplicaciones industriales.*** Los frutos de numerosas especies de palma, como *Washingtonia filifera* (*W. filifera*) y *Sabal palmetto* (*S. palmetto*), son ricos en metabolitos secundarios y se utilizan ampliamente con fines alimentarios e industriales. Este estudio tuvo como objetivo analizar los perfiles de ácidos grasos de los granos de estas dos especies de palma subutilizadas y desarrollar modelos predictivos mediante aprendizaje automático para evaluar su potencial para diversas aplicaciones industriales. El ácido oleico fue el ácido graso más abundante, representando el 37,13% en *W. filifera* y el 33,29% en *S. palmetto*. Le siguió el ácido láurico con el 25,80% en *W. filifera* y el 25,57% en *S. palmetto*. Los ácidos linoleico y mirístico también fueron predominantes, con rangos variables entre las dos especies. Se analizó la grasa total, los ácidos grasos saturados (ΣSFA), los ácidos grasos monoinsaturados (ΣMUFA) y los ácidos grasos poliinsaturados (ΣPUFA) mediante modelos de aprendizaje automático (AA). Se evaluó el rendimiento de los modelos de máquinas de vectores de soporte (SVM), bosques aleatorios (RF), potenciación de gradiente extremo (XGBoost) y perceptrón multicapa (MLP) mediante métricas como RMSE, R^2 y MAE. SVM obtuvo los R^2 más altos (0,98-0,99), lo que demuestra su eficacia para predecir con precisión los perfiles de ácidos grasos, lo cual es crucial para evaluar su idoneidad para diversos usos industriales, como la alimentación, el biodiésel y posibles aplicaciones en la industria cosmética.

PALABRAS CLAVE: *Aceite de palma; Aprendizaje automático;* Arecaceae*; Biodiésel de aceite de almendra; Inteligencia artificial.*

Citation/Cómo citar este artículo: Bozkurt T. 2025. Machine Learning Analysis of Fatty Acid Composition in *Washingtonia filifera* and *Sabal* palmetto Palm Kernels for Potential Industrial Applications. *Grasas Aceites* **76** (2), 2301. https://doi.org/10.3989/gya.0320251.2301

1. INTRODUCTION

The palm tree is a member of the Palmae or Arecaceae family, which is divided into five subfamilies, encompassing approximately 189 genera and an estimated 3,000 species (Rivas *et al.*, 2012). These plants are commonly found in tropical and subtropical ecosystems. In addition to being among the oldest monocotyledonous flowering plants, palm trees have a rich and widespread fossil record (Eiserhardt *et al.*, 2011). Documentation of palm fossils began in Europe in the early 19th century and also gained momentum with the increase in geological research in the United States in the late 19th century (Dransfield *et al.*, 2008).

In addition to having a very ancient history, palm trees are known to provide numerous economic and other benefits to humans (Rivas *et. al*., 2012). Palm trees are utilized in a wide range of areas, including materials for clothing, weaving, fuel, medicine, and food sources, as well as materials for the construction of temporary or permanent buildings, and for ornamental purposes (Dransfield *et al.*, 2008; Ergün, 2021). As food products, they provide fruits, seeds, and beverages derived from their fruit and sap, fibers, oils, and waxes (Rivas *et. al*., 2012).

Among the various products derived from palm trees, palm oil is the most significant. Palm oil is an edible oil obtained by extracting the orange-red mesocarp from the fruits of the oil palm tree known as *Elaeis guineensis*. The oil content ranges from 45 to 55%, and the palm tree can grow to an average height of 20-30 meters with an economic lifespan of 25-30 years (Edem, 2022). Palm oil makes substantial contributions to both local and global economies. According to USDA data, the global palm oil production in 2022 was 77.5 million MT, with Indonesia and Malaysia being the primary producers. The production data was collected from 28 countries. According to the 2022 export data, Indonesia and Malaysia together accounted for an average of 89% of the market share, while India (18%) and China (14%) were the largest importers.

The three most commonly grown palm species in the world are the date (*Phoenix dactylifera* L.), the coconut (*Cocos nucifera* L.) and the African oil palm (*Elaeis guineensis* L.) (Hemmati *et al.*, 2020). According to FAO (2022) data, the harvested area for the African oil palm in 2021 was 28.9 million hectares, with a production of approximately 416.4 million tons. For the coconut, the harvested area was 11.3 million hectares, with an average production of 63.6 million tons reported for the same year. For the date palm, the harvested area was 1.3 million hectares, with a production of about 9.6 million tons.

Palm trees, which bring significant economic benefits to many countries, are often considered purely as ornamental plants in others. A prime example of this is the Washingtonia genus. The Washingtonia genus includes two species: *W. filifera* and *Washingtonia robusta*. These two species are relatively similar. *W. robusta*, also known as the Mexican Fan Palm, can grow up to 15-20 meters tall with a trunk diameter of about 0.6-1.2 meters. A plant can produce approximately 50 leaves annually. The leaves are divided into 40-60 segments and are attached to the tree with an outwardly curved petiole that is 1.2-1.5 meters long. *W. robusta*, also known as the Mexican fan palm, is a palm species with a thick, reddish trunk and broad, dark green leaves that grow rapidly. Towards the end of spring, it blooms small, fleshy flowers on long, branched clusters. These flowers then turn into small, black-brown fruits filled with a sweet and sugary content. Each fruit contains a single seed and the fruits of this species are high in sugar and oil (Gomaa, 2019).

W. filifera is known by various names, such as desert fan palm, California fan palm, and Washington palm (Nehdi, 2011). The California fan palm can reach 9 to 21 meters and even up to 23 meters. The trunk diameter is 1.5 meters, and the leaves are 0.9-1.8 meters wide. The leaves are divided at the midrib, equipped with curved teeth, and attached to the tree with a petiole that is 1.5 meters long. Each tree can produce up to 15 flower clusters and 18,000 fruits per cluster. The ripe fruit is black, spherical or elongated, and is surrounded by a hemispherical brown seed, averaging 10-13 mm long (Dewir *et al.*, 2020).

S. palmetto, also known as the cabbage palm, can grow from 9 meters to 27 meters, even up to 30 meters. The trunk diameter ranges from 15 to 45 cm (McPherson and Williams, 1996). Each leaf segment is divided more than halfway along its length. The split segments are rigid or have tan-colored fiber strips at the tips. The leaves are salt-tolerant, but the roots are not. The branched flower clusters produced in late spring usually extend beyond the canopy leaves and contain thousands of small, creamy-white, fragrant flowers that attract bees. By late summer, it produces black fruits about ¼ inch in diameter (Broschat, 2013).

The fruits and seeds these palms produce are valuable, containing essential fatty acids like other plant seeds. Limited research has reported the oil content and significance of these seeds. For example, Nehdi (2011) found that *W. filifera* seeds contain 57.39% total unsaturated fatty acid content and reported that the seeds contain 16.3% oil. In light of these values, it would be correct to say that these fruits have critical potential in the food and oil industries.

On the other hand, the interest in clean and renewable resources is increasing day by day due to the depletion of fossil fuel reserves and the negative environmental effects of climate change. Biodiesel, defined as a renewable energy source, has ecologically critical value due to its biodegradability, carbon neutrality and lack of any toxicity. Biodiesel can be produced from edible oils such as sunflower, coconut, and palm oil, but the production cost is relatively high compared to diesel (Gopinath *et al.*, 2017). Therefore, different non-edible palm species and waste palm seeds present significant potential for biodiesel production.

Machine learning (ML) based on data science is widely used today to transform complex chemical and biological data into comprehensible outputs. ML technologies, which are becoming widespread today, are a superior alternative to traditional modeling approaches which are mostly uncertain, nonlinear, complex and multivariate (Aghbashlo *et al.*, 2021). Recent studies have also demonstrated the effectiveness of ML in analyzing fatty acid profiles and predicting their properties for various applications. For instance, Zhang *et al*. (2023) utilized machine learning models to predict the key properties of biodiesel based on its fatty acid composition, highlighting the capability of ML in establishing crucial relationships in lipid chemistry. This aligns with the potential of ML in extracting valuable insights from the complex fatty acid data in our study. This study used four ML algorithms with different metrics to analyze each data point in the Python programming language.

However, research specifically focusing on the detailed fatty acid profiles of *W. filifera* and *S. palmetto* seeds and their potential for diverse industrial applications, coupled with comprehensive machine learning analysis, remains limited. To address this gap, the present study aims to:

- Determine the detailed fatty acid profiles of the seeds of *W. filifera* and *S. palmetto*, two palm species commonly used for landscaping in Turkey whose fruits are largely unutilized.
- Develop and evaluate the performance of four different machine learning models (SVM, RF, XGBoost, and MLP) in predicting the total saturated, monounsaturated, and polyunsaturated fatty acid contents in these seeds.
- Provide insights into the potential of these palm seed oils for various industrial applications, including food, bioenergy (biodiesel), and potential cosmetic uses, based on their fatty acid composition and the predictive capabilities of the machine learning models.

The novelty of this research lies in its comprehensive analysis of the fatty acid profiles of these underutilized palm seeds using a comparative machine learning approach to predict their key properties relevant to industrial applications. The scope of the study encompasses the detailed gas chromatography analysis of the seed oils and the application of four distinct machine learning algorithms for predictive modeling. The potential outcomes of this research include providing valuable data for the valorization of these palm seeds, guiding future research into their industrial applications, and demonstrating the efficacy of machine learning in analyzing and predicting the properties of non-conventional oil sources.

2. MATERIAL AND METHODS

2.1. Plant material

Mature fruits of *S. palmetto* and *W. filifera* were collected from the Ali Nihat Gökyiğit Botanical Garden at Çukurova University. The seeds were manually extracted from these fruits and used as the primary plant material for the study.

2.2. Oil extraction

Oil was extracted from 50 g of dried and ground seeds using the Soxhlet extraction method with n-hexane as the solvent, following the procedure outlined by Akbari *et al.* (2012). The fatty acid composition of the extracted seed oils was analyzed using gas chromatography (GC; Perkin Elmer, Shelton, USA). A capillary column (30 m × 0.25 mm ID, 0.25 µm film thickness) coupled with a flame ionization detector (FID) was employed for chromatographic separation. The oven temperature was programmed to increase from 60 to 290°C at a rate of 6°C per minute, with helium serving as the carrier gas. The results were expressed as percentages, accompanied by their respective standard deviations, as described by Bentrad and Gaceb-Terrak (2020). The raw GC data, containing peak areas for each fatty acid, was normalized to calculate the percentage of each fatty acid relative to the total fatty acids identified.

2.3. Data preprocessing

Prior to training the machine learning models, the obtained fatty acid composition data underwent several preprocessing steps. These steps included: (1) Handling of Missing Values: The dataset was checked for any missing values; no missing values were found in this study. (2) Data Scaling: To prevent features with larger ranges from dominating the model performance, the fatty acid data (percentage values) were scaled to a range between 0 and 1 using the Min-Max scaling method. The formula for Min-Max scaling is:

$$X_{\mathrm{scaled}} = \frac{X - X_{\mathrm{min}}}{X_{\mathrm{max}} - X_{\mathrm{min}}}$$

where X is the original value, X_{min} is the minimum value of the feature, and X_{max} is the maximum value of the feature. (3) Data Splitting: The dataset was randomly split into training (70%) and testing (30%) sets to evaluate the generalization ability of the trained models (as illustrated in Figure 1). The training set was used to train the models, while the testing set was used to assess the performance of the trained models on unseen data.

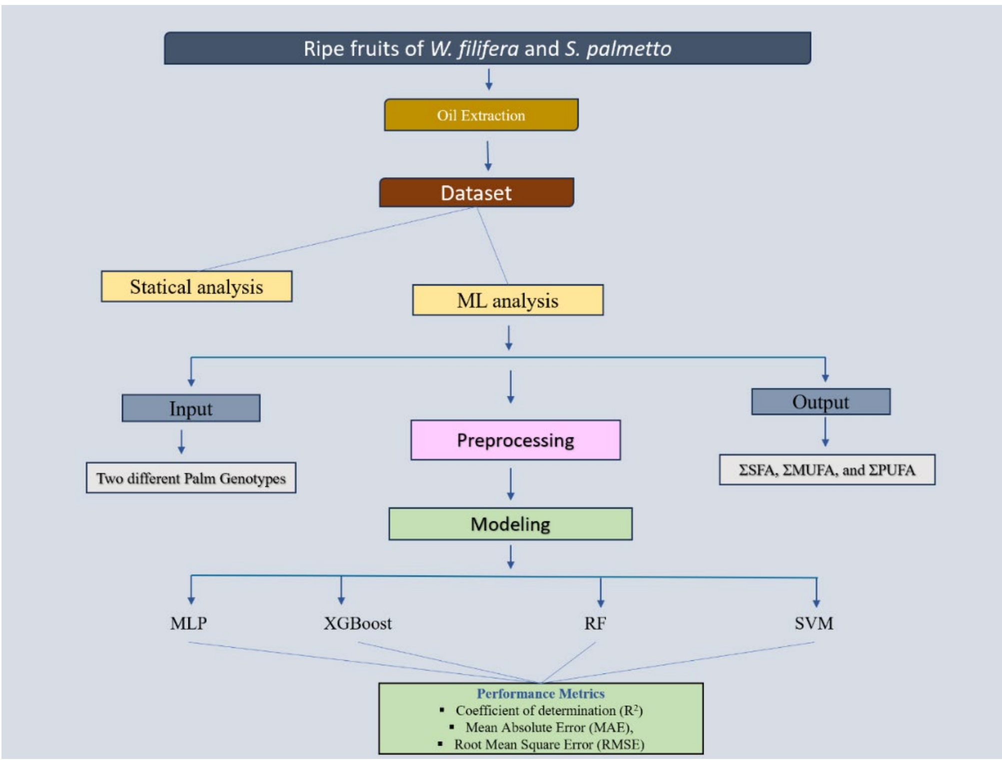

FIGURE 1. Schematic Diagram of methodology.

2.4. Machine learning modeling

The dataset used for training the machine learning models consisted of the fatty acid profiles obtained from the gas chromatography analysis of the seed oils of the two palm species, as described in Section 2.2. This tabular dataset included the percentage values of different fatty acids for each seed sample. The target variables for the machine learning models were the total saturated fatty acids (ΣSFA), total monounsaturated fatty acids (ΣMUFA), and total polyunsaturated fatty acids (ΣPUFA).

Four machine learning models—Support Vector Machine (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and Multilayer Perceptron (MLP)—were employed for regression analysis to predict the total fatty acid content (Figure 1). These models were selected due to their diverse algorithmic approaches and their potential to perform well on small to medium-sized datasets. SVM is effective in modeling linear and non-linear relationships and performs well in high-dimensional spaces. RF, a tree-based ensemble learning method, reduces overfitting and provides insights into feature importance. XGBoost, an optimized gradient boosting algorithm, is known for its high accuracy and efficiency. MLP, a flexible artificial neural network architecture, can model complex non-linear relationships.

The performance of each machine learning model was optimized using hyperparameter tuning with k-fold cross-validation (k=5) on the training data. Different combinations of hyperparameters were evaluated based on the Root Mean Square Error (RMSE), Coefficient of Determination (R^2), and Mean Absolute Error (MAE) metrics. The hyperparameters tuned for each model included: SVM (kernel type: linear, RBF, sigmoid; regularization parameter C), RF (number of trees, maximum depth of trees), XGBoost (learning rate, number of trees, maximum depth), and MLP (number of hidden layers, number of neurons per layer, activation function). The formulas for these metrics are presented in Table 1.

TABLE 1. Formulas of Metrics.

	Formula
1	$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \tilde{Y})^2}$
2	$RMSE = \sqrt{\frac{\left(\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\right)}{n}}$
3	$MAE = \frac{1}{n}\sum_{i=1}^{n}\lvert Y_i - \hat{Y}_i\rvert$

Y_i = actual value, $\hat{Y}_i$ = Predicted value, $\tilde{Y}$ = mean of the actual values and n = sample count.

The mathematical representations of some of the employed algorithms are as follows:

$$\hat{\boldsymbol{y}}_{\boldsymbol{i}} = \sum_{k=1}^{K} \boldsymbol{f}_{\boldsymbol{k}}(\boldsymbol{X}_{\boldsymbol{i}}) \text{ ,} \boldsymbol{f}_{\boldsymbol{k}} \in \boldsymbol{F} \qquad (1)$$

(Equation 1 represents the ensemble of $\boldsymbol{K}$ regression trees in the XGBoost model, where $\hat{\boldsymbol{y}}_{\boldsymbol{i}}$ is the predicted value for sample *i*, Xi is the feature vector, and $\boldsymbol{F}$ is the space of regression trees.)

$$\boldsymbol{\mathcal{L}}(\emptyset) = \sum_{i}^{n} \boldsymbol{l}(\boldsymbol{y}_{\boldsymbol{i}}, \hat{\boldsymbol{y}}_{\boldsymbol{i}}) + \sum_{k=1}^{K} \boldsymbol{\Omega}(\boldsymbol{f}_{\boldsymbol{k}}) \qquad (2)$$

(Equation 2 shows the objective function of XGBoost, which includes a loss function $\boldsymbol{l}$ and a regularization term $\boldsymbol{\Omega}$.)

$$\boldsymbol{E} = \frac{\boldsymbol{1}}{\boldsymbol{n}}\sum_{i=1}^{n}(\boldsymbol{t}_{\boldsymbol{i}} - \boldsymbol{y}_{\boldsymbol{i}})^{\boldsymbol{2}} \qquad (3)$$

(Equation 3 represents the Mean Squared Error (MSE) function, which was used as the loss function during the training of the Multilayer Perceptron (MLP) model. Here, $\boldsymbol{n}$ is the number of inputs, $\boldsymbol{t_i}$ is the desired output, and $\boldsymbol{y_i}$ is the actual output.)

$$\boldsymbol{f(x) = \langle w, x \rangle + b \ = 0} \qquad (4)$$

(Equation 4 represents the decision boundary of a linear Support Vector Machine (SVM), where $\boldsymbol{w}$ is the weight vector, $\boldsymbol{x}$ is the input vector, and $\boldsymbol{b}$ is the bias term.)

$$\boldsymbol{\hat{y} = \frac{1}{B} \sum_{i=1}^{B} T_i(x)} \qquad (5)$$

(Equation 5 shows the prediction of a Random Forest (RF) for a regression problem, where $\boldsymbol{B}$ is the number of trees and $\boldsymbol{T_i(x)}$ is the prediction of the $\boldsymbol{i}$ th tree.)

2.5. Statistical analysis

The experimental design followed a completely randomized design with three biological replicates per palm species. Data analysis was performed using the SAS-JMP statistical software (Version 16.0, SAS Institute Inc., Cary, NC, USA). Variance analysis (ANOVA) was conducted to determine the statistical significance of differences in fatty acid composition between the two palm species. The least significant difference (LSD) test was applied for post-hoc comparisons of means at a significance level of $p < 0.05$. Percentage data were subjected to arcsine transformation prior to ANOVA to ensure normality and homogeneity of variances. The results of the ANOVA and LSD tests provided a statistical basis for comparing the fatty acid profiles of *S. palmetto* and *W. filifera* seeds and helped to contextualize the findings of the machine learning models by identifying statistically significant differences in the fatty acid compositions that the models aimed to predict.

RESULTS AND DISCUSSION

This study aimed to determine the oil content and fatty acid composition of seeds from W. filifera and S. palmetto palm species and to model the total fat and major fatty acid group content using machine learning algorithms.

The statistical evaluation of the fatty acid analysis data revealed significant differences in the proportions of various fatty acids between the two palm species ($p \leq 0.05$). Across both species, the most abundant fatty acid was oleic acid, with an average proportion of 35.21%. This was followed by lauric acid (25.68%), myristic acid (13.19%), linoleic acid (10.80%), palmitic acid (8.63%), and stearic acid (2.40%). Minor fatty acids, including pentadecanoic acid (0.015%), linolelaidic acid (0.02%), and tricosanoic acid (0.025%), were detected in very low amounts, while heneicosanoic acid and erucic acid were not detected in either species. A comparison of the total fatty acid content indicated a slightly higher value in *W. filifera* (4.33%) compared to *S. palmetto* (4.23%). The individual fatty acid compositions within each palm species were also examined (Table 2). Statistical significance levels in fatty acid compositions presented in the table are indicated with one asterisk (*) for $p < 0.05$, two asterisks (**) for $p < 0.01$ and three asterisks (***) for $p < 0.001$. Significant differences between species were determined by the LSD (Least Significant Difference) test calculated after the angle transformation applied to the percentage data, and different letters (a–^) indicate significant differences at $p \leq 0.05$ level. The highest proportion was observed for oleic acid in *W. filifera* (37.13%), followed by oleic acid in *S. palmetto* (33.29%). Lauric acid was also abundant in both species, with values of 25.80% in *W. filifera* and 25.57% in *S. palmetto*. Myristic acid showed higher levels in *S. palmetto* (15.11%) compared to *W. filifera* (11.27%). Linoleic acid was more prevalent in *W. filifera* (11.68%) than in *S. palmetto* (9.92%). Some fatty acids were not detected (ND) in a specific species, such as linolelaidic acid in *W. filifera*, heneicosanoic acid and erucic acid in both species. Pentadecanoic acid and tricosanoic acid were found in very low amounts in *W. filifera*.

Oleic acid is predominantly present in animal-derived fats, including tallow and lard, but it is also abundant in certain vegetable oils such as olive oil, sunflower oil, and canola oil (Gillingham *et al*., 2011).

In research conducted by Sawaya *et al.* (1984), oleic acid was reported as the primary fatty acid in the seed oils of two date palm varieties (Ruzeiz and Sifri), accounting for 44.25% of the total fatty acid content. Other significant fatty acids identified included lauric acid (17.35%), myristic acid (11.45%), palmitic acid (10.30%), and linoleic acid (8.45%). Similarly, Zang *et al.* (2017) analyzed the seed oil of *Balanites aegyptiaca* (desert date) and found that the major fatty acids were palmitic acid (16:0) at 20.51%, oleic acid (18:1) at 28.32%, linoleic acid (18:2) at 19.20%, and stearic acid (18:0) at 15.97%. They emphasized that this oil is a good candidate for edible oil due to its high unsaturated fatty acid content (approximately 47.42%).

TABLE 2. Palm fatty acid compositions (%)

Compound name	*Washingtonia filifera*	*Sabal palmetto*
Caproic acid (C6:0)	0.11 u (1.90)	0.03 \ (0.99)
Caprylic acid (C8:0)	1.49 m (7.01)	0.45 o (3.84)
Capric acid (C10:0)	1.07 n (5.95)	0.43 p (3.75)
Lauric acid (C12:0)	25.80 c (30.52)	25.57 d (30.38)
Tridecanoic acid (C13:0)	0.04 [(1.14)	0.05 z (1.28)
Myristic acid (C14:0)	11.27 g (19.61)	15.11 e (22.87)
Pentadecanoic acid (C15:0)	0.01 ^ (0.57)	0.02] (0.81)
Palmitic acid (C16:0)	7.80 j (16.22)	9.46 i (17.91)
Heptadecanoic acid (C17:0)	0.06 y (1.40)	0.06 y (1.40)
Stearic acid (C18:0)	2.62 k (9.31)	2.19 l (8.51)
Arachidic acid (C20:0)	0.08 w (1.62)	0.09 v (1.78)
Heneicosanoic acid (C21:0)	ND	ND
Behenic acid (C22:0)	0.06 y (1.40)	0.06 y (1.40)
Tricosanoic acid (C23:0)	0.02] (0.81)	0.03 \ (0.99)
Lignoceric acid (C24:0)	0.07 x (1.51)	0.06 y (1.40)
ΣSFA		
Palmitoleic acid (C16:1) ω7	0.03 \ (0.99)	0.11 t (1.95)
Elaidic acid (C18:1n9t)	0.02] (0.81)	0.24 q (2.80)
Oleic acid (C18:1n9c) ω9	37.13 a (37.54)	33.29 b (35.23)
cis-11-Eicosenoic acid (C20:1)	0.22 r (2.68)	0.14 s (2.14)
Erucic acid (C22:1n9)	ND	ND
ΣMUFA		
Linolelaidic acid (Cl8:2n6t)	ND	0.04 [(1.13)
Linoleic acid (Cl8:2n6c) ω6	11.68 f (19.98)	9.92 h (18.35)
a-Linolenic acid (C18:3n3) ω3	0.07 x (1.51)	0.03 \ (0.99)
ΣPUFA		
Total Identified FA (%)	99.98	99.96

P < 0.05*, *p* < 0.01**, P < 0.001***, LSD value was calculated according to the angle conversion value. Different letters (a–^) indicate significant differences according to the LSD test (*p* ≤ 0.05). LSDcompound: 0,029***, LSDspecies: 0,008***, LSDcompound x species: 0,041***. Percentages were subjected to angle transformation and these values were presented in parentheses.

In another study by Floris (2021), the GC-FID analysis of n-hexane extracts from *W. filifera* seeds collected from two different regions (Sousse and Gabès) revealed that the highest fatty acid was lauric acid (C12:0) at 36.11 and 33.50%, followed by oleic acid at 25.09 and 25.47%. The third highest fatty acid was myristic acid

(C14:0) at 12.26 and 10.40%. In contrast, our study identified oleic acid as the highest fatty acid. The primary reasons for this discrepancy may include differences in climate, soil, nutrient availability, and potential stress factors affecting the plants in different regions.

In this study on palm seeds, the second highest fatty acid identified was lauric acid (C12:0), an edible oil found in coconut meat and reported to constitute 45-53% of the total fatty acid composition (Dayrit, 2015). According to the results of another study on oil palm (*Elaeis guineensis*), lauric acid was the predominant fatty acid at 45-55%, along with myristic acid (C14:0) at 14-18% and oleic acid (C18:1) at 12-19% (Codex Alimentarius Commission, 1999).

Ubgogu *et al*. (2006) conducted one of the initial studies on the antimicrobial effects of lauric acid in palm seed oil. They tested the antimicrobial effects of various palm seed oils on two bacteria and *Candida albicans*, finding that the kernel oil with the highest lauric acid content exhibited the best antimicrobial activity. Based on this result, the researchers suggested that lauric acid was the antimicrobial agent in palm kernel oil.

Rupilius and Ahmad (2007) emphasized the value of palm kernel oil, noting its significance for the oleochemical industry due to its content of short (C8-C10) and medium-chain (C12-C14) fatty acids. Approximately 60% of palm kernel oil and 5% of palm oil are used to produce oleochemicals (Ahmad, 2007). Palm kernel oil is primarily obtained from the seeds of the Elaeis guineensis tree (Kareem, 2017). Various applications have been reported to affect the efficiency and yield of biodiesel production. Abigor *et al.* (2000) reported that the conversion rate of palm kernel oil to biodiesel ranged from 15 to 72% depending on the method used, while Alamu *et al*. (2008) achieved biodiesel yields ranging from 29.5 to 96% with different methods. In this study, unlike *E. guineensis*, the fatty acid profiles of the seeds from two palm species, which are not typically utilized for their fruits or seeds, were determined to evaluate their potential usability from food to biodiesel.

3.1. Machine learning analysis

The predictive capabilities of four ML models (RF, XGBoost, MLP, and SVM) were evaluated for Total Saturated Fatty Acid (ΣSFA), Monounsaturated Fatty Acid (ΣMUFA), and Polyunsaturated Fatty Acid (ΣPUFA) contents, which are important for assessing the nutritional quality and potential applications of palm seed oils. Model performance was assessed using R^2, RMSE, and MAE to determine the most accurate model for predicting the fatty acid composition of these palm seeds (Table 3).

TABLE 3. Analysis results based on evaluation metrics with ML models

ML Models	Assessment matrices	*Washingtonia filifera*				*Sabal palmetto*			
		Total Fat	ΣSFA	ΣMUFA	ΣPUFA	Total Fat	ΣSFA	ΣMUFA	ΣPUFA
MLP	RMSE	0.13	1.11	3.21	0.19	0.13	1.42	0.42	0.25
	R^2	0.96	0.98	0.67	0.99	0.99	0.95	0.99	0.97
	MAE	0.11	0.88	2.86	0.16	0.11	1.09	0.32	0.20
RF	RMSE	0.10	1.73	0.78	.38	0.31	1.99	1.02	0.91
	R^2	0.97	0.95	0.98	0.95	0.92	0.90	0.93	0.57
	MAE	0.09	1.27	0.57	0.30	0.22	1.51	0.87	0.52
XGboost	RMSE	0.20	1.14	0.57	0.20	.26	1.29	0.19	0.65
	R^2	0.90	0.98	0.99	0.99	0.95	0.96	0.99	0.78
	MAE	0.16	0.90	0.48	0.16	0.13	0.95	0.16	0.36
SVM	RMSE	0.06	0.01	0.05	0.09	0.06	0.09	0.08	0.14
	R^2	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98
	MAE	0.05	0.08	0.05	0.08	0.05	0.07	0.06	0.12

The Multilayer Perceptron (MLP) model demonstrated high R^2 values for ΣPUFA (0.99), ΣSFA (0.98), and Total Fat (0.96) in *W. filifera*, while showing strong performance for ΣMUFA (0.99) in *S. palmetto*. The Random Forest (RF) model exhibited good predictive power for ΣMUFA in both species (0.98 for *W. filifera* and 0.93 for *S. palmetto*). The Extreme Gradient Boosting (XGBoost) model achieved high R^2 values for ΣMUFA and ΣPUFA (0.99) in *W. filifera* and ΣMUFA (0.99) in *S. palmetto*. Overall, the Support Vector Machine (SVM) model consistently outperformed the other algorithms, achieving R^2 values of 0.99 across all predicted variables for both *W. filifera* and *S. palmetto*, along with the lowest RMSE and MAE values. This indicates that SVM is the most accurate and reliable model for predicting the fatty acid composition of the studied palm seeds. While MLP and XGBoost showed promising results for specific components, RF generally exhibited lower predictive performance, particularly for *S. palmetto*.

The statistical analysis (Table 3) revealed a significantly higher proportion of oleic acid (a component of ΣMUFA) in *W. filifera* compared to *S. palmetto* ($p < 0.05$). This difference may have contributed to the high accuracy of the SVM model in predicting ΣMUFA content for *W. filifera* ($R^2 = 0.99$). Similarly, the relatively higher levels of myristic acid (a component of ΣSFA) observed in *S. palmetto* ($p < 0.05$) could be linked to the SVM model's strong performance in predicting ΣSFA for this species ($R^2 = 0.99$). Despite the higher linoleic acid (a component of ΣPUFA) content in *W. filifera* ($p < 0.05$), the SVM model demonstrated high prediction accuracy for ΣPUFA in both species ($R^2 = 0.99$ and 0.98), suggesting its robustness in modeling this polyunsaturated fatty acid group.

The high prediction accuracy achieved with the SVM model in our study aligns with the potential of ML techniques in analyzing and predicting complex biochemical data, as highlighted by Jin *et al*. (2022) in the context of bioenergy production. Our findings, utilizing multiple performance metrics and comparing different ML algorithms, provide a robust assessment of the predictive capabilities for fatty acid composition in these non-traditional oilseed sources. Further optimization of ML parameters could potentially enhance the prediction accuracy even further.

4. CONCLUSIONS

This study aimed to determine the fatty acid composition of the mature seeds of *W. filifera* and *S. palmetto*. It was determined that Oleic acid was the most dominant fatty acid in both species and Lauric acid was the second dominant fatty acid. Linoleic acid and Myristic Acid were observed to follow these. It should be noted that they could be utilized for critical applications such as biodiesel production, besides food products, including cosmetics, and pharmaceutical industries. The underutilization of these seeds and fruits, especially for landscaping purposes, remains evident in contemporary practices. By revealing the fatty acid profiles of seeds from these two palm species, this study shed light on the potential use of these overlooked resources in various fields. Total fat, ΣSFA, ΣMUFA, and ΣPUFA amounts were analyzed using four different ML models, and their performances were deeply analyzed using evaluation metrics such as R^2, RMSE, and MAE. It was determined that the SVM model emerged as the most effective and reliable model for predicting the total fatty acid components of *W. filifera* and *S. palmetto* species. Although MLP, Random Forest, and XGBoost models exhibited good performance in many cases, they generally lagged behind SVM. These findings clearly demonstrate the superiority of SVM, particularly in terms of accuracy and error rates, in predicting fatty acid compositions in this context. To enhance future research in this area, we suggest exploring the fatty acid profiles of other underutilized palm species, which could provide additional insight into their potential applications in biodiesel production and other industrial uses. Expanding the dataset to include a broader range of species and environmental conditions could improve the predictive accuracy of ML models, particularly in understanding the influence of external factors on fatty acid composition. Further, integrating advanced ML techniques, such as deep learning models, may enhance the prediction capabilities and provide more robust data interpretation for both food and non-food applications of palm kernel oil.

DECLARATION OF COMPETING INTEREST

The authors of this article declare that they have no financial, professional or personal conflicts of interest that could have inappropriately influenced this work.

AUTHORSHIP CONTRIBUTION STATEMENT

T. Bozkurt, participated in the design of experiments, conduct of the study, processing and interpretation of data, performing machine learning studies, writing the article, and conducting all other work.

REFERENCES

Abigor RD, Uadia PO, Foglia TA, Haas MJ, Jones KC, Okpefa E, Bafor ME. 2000. Lipase-catalysed production of biodiesel fuel from some Nigerian lauric oils. *Biochem. Soc. Trans.* **28** (6), 979–981. https://doi.org/10.1042/bst0280979

Aghbashlo M, Peng W, Tabatabaei M, Kalogirou SA, Soltanian S, Hosseinzadeh-Bandbafha H, Lam SS. 2021. Machine learning technology in biodiesel research – a review. *Progress in Energy Combust. Sci.* **85**, 100904. https://doi.org/10.1016/j.pecs.2021.100904

Akbari M, Razavizadeh R, Mohebbi GH, Barmak A. 2012. Oil characteristics and fatty acid profile of seeds from three varieties of date palm (*Phoenix dactylifera*) cultivars in Bushehr-Iran. *Afr. J. Biotechnol.* **11**, 12088-93. https://doi.org/10.5897/AJB11.4230.

Alamu OJ, Waheed MA, Jekayinfa SO. 2008. Effect of ethanol–palm kernel oil ratio on alkali-catalyzed biodiesel yield. *Fuel* **87**, 1529-1533.

Altobi MAS, Bevan G, Wallace P, Harrison D, Ramachandran KP. 2019. Fault diagnosis of a centrifugal pump using MLP-GABP and SVM with CWT. *Eng. Sci. Technol., Int. J.* **22**, 854-61. https://doi.org/10.1016/j.jestch.2019.01.005.

Bentrad N, Gaceb-Terrak R. 2020. Chemical identification of some toxic residues bioaccumulated in date palm seeds (*Phoenix dactylifera*). *Lett. Appl. NanoBioSci.* **9**, 1263-74. https://doi.org/10.33263/LIANBS93.12631274.

Broschat TK. 2013. *Sabal palmetto*: Sabal or cabbage palm. The University of Florida George A. Smathers Libraries 6.

Chen T, Guestrin C. 2016. Xgboost: A scalable tree boosting system. In: 2016 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California, USA. pp. 785-94. http://dx.doi.org/10.1145/2939672.2939785.

Codex Alimentarius Commission. 1999. Reports of the sixteenth session of the Codex Committee on Fats and Oils, Alinorm 99/17.

Dayrit FM. 2015. The properties of lauric acid and their significance in coconut oil. *J. Am. Oil Chem. Soc.* **92**, 1-15. https://doi.org/10.1007/s11746-014-2562-7

Dewir YH, El-Mahrouk ME, Seliem MK, Murthy HN. 2020. Bioactive compounds of California fan palm *Washingtonia filifera* (Linden ex André H. Wendl. ex de Bary). In: Murthy, H. & Bapat, V. (eds.), Bioactive Compounds in Underutilized Fruits and Nuts. Reference Series in Phytochemistry. Springer, Cham. pp. 63-74. https://doi.org/10.1007/978-3-030-30182-8_7.

Dransfield J, Uhl NW, Asmussen CB, Baker WJ, Harley MM, Lewis CE. 2008. Genera palmarum-the evolution and classification of the palms. Royal Botanic Gardens, Kew, London, UK. pp. 732.

Edem DO. 2002. Palm oil: Biochemical, physiological, nutritional, hematological and toxicological aspects – a review. *Plant Foods Hum. Nutr.* **57**, 319-41. https://doi.org/10.1023/A:1021828132707

Eiserhardt WL, Svenning JC, Kissling WD, Balslev H. 2011. Geographical ecology of the palms (Arecaceae): determinants of diversity and distributions across spatial scales. *Ann. Bot.* **108**, 1391-1416.

Ergün Z. 2021. Seed oil content and fatty acid profiles of endemic Phoenix theophrasti Greuter, Phoenixroebelenii O'Brien, Phoenix caneriensis Hort. Ex Chabaud, and *Phoenix dactylifera* L. grown in the same locality in Turkey. *Turk. J. Agric. For.* **45** (5), 557-564.

FAO. 2024. Available from: https://www.fao.org/faostat/en/#home: 01.05.2024.

Floris S. 2021. Biological activities and phenolic composition of *Washingtonia filifera* seeds. (Doctoral thesis, University of Cagliari, Italy).

Gillingham LG, Harris-Janz S, Jones PJ. 2011. Dietary monounsaturated fatty acids are protective against metabolic syndrome and cardiovascular disease risk factors. *Lipids* **46**, 209-28. https://doi.org/10.1007/s11745-010-3493-6.

Gomaa RA. 2019. Physico-chemical characteristics of *Washingtonia robusta* fruit oil. *Suez Canal Univ. J. Food Sci.* **6**, 19-25.
Gopinath S, Kumar PSM, Arafath KY, Thiruvengadaravi KV, Sivanesan S, Baskaralingam P. 2017. Efficient mesoporous SO_4^{2-}/Zr-KIT-6 solid acid catalyst for green diesel production from esterification of oleic acid. *Fuel* **203**, 488-500. https://doi.org/10.1016/j.fuel.2017.04.128.
Hemmati C, Nikooei M, Al-Sadi AM. 2020. Four decades of research on phytoplasma diseases of palms – a review. *Int. J. Agric. Biol.* **24**, 631-44. https://doi.org/10.17957/IJAB/15.1473.
Jin H, Kim YG, Jin Z, Rushchitc AA, Al-Shati AS. 2022. Optimization and analysis of bioenergy production using machine learning modeling: Multi-layer perceptron, Gaussian processes regression, K-nearest neighbors, and Artificial neural network models. *Energy Rep.* **8**, 13979-13996.
McPherson K, Williams K. 1996. Establishment growth of cabbage palm, *Sabal palmetto* (Arecaceae). *Am. J. Bot.* **83**, 1566-1570.
Nehdi IA. 2011. Characteristics and composition of *Washingtonia filifera* (Linden ex André) H. Wendl. seed and seed oil. *Food Chem.* **126**, 197-202.
Pink CM. 2016. Forensic ancestry assessment using cranial nonmetric traits traditionally applied to biological distance studies. In *Biol. Distance Anal.* (pp. 213-230). Academic Press.
Rivas M, Barbieri RL, Maia LCD. 2012. Plant breeding and *in situ* utilization of palm trees. *Ciênc. Rural* **42**, 261-269.
Rupilius W, Ahmad S. 2007. Palm oil and palm kernel oil as raw materials for basic oleochemicals and biodiesel. *Eur. J. Lipid Sci. Technol.* **109**, 433-439.
Sawaya WN, Khalil JK, Safi WJ. 1984. Chemical composition and nutritional quality of date seeds. *J. Food Sci.* **49**, 617-619.
Ubgogu OC., Onyeagba RA, Chigbu OA. 2006. Lauric acid content and inhibitory effect of palm kernel oil on two bacterial isolates and *Candida albicans*. *Afr. J. Biotechnol.* **5** (11).
Zang CU, Jock AA, Garba IH, Chindo IY. 2017. Physicochemical and phytochemical characterization of seed kernel oil from desert date (*Balanites aegyptiaca*). *J. Chem. Eng. Bioanal. Chem.* **2**, 49-62. https://doi.org/10.25177/JCEBC.2.1.1.
Zhang L, Wang J, Li S. 2023. Machine learning-based prediction of biodiesel fuel properties using fatty acid methyl ester composition. *Renew. Ener.* **210**, 15-24