Journal of
Food and Pharmaceutical Sciences

Research Article

Application of Clustering Analysis with Unsupervised Technique on Fish Oil Samples

Nadia Miftahul Jannah[1]*, Karina Primatyas Ningrum[2] and Firhani Anggriani Syafrie[3]

[1]Faculty of Pharmacy, Sultan Agung Islamic University, Semarang, Indonesia
[2]Clinical and Communities of Pharmacy, Al-Fatah Health College, Bengkulu, Indonesia
[3] Pharmacy, Mandala Waluya University, Kendari, Indonesia
*Corresponding author: Nadia Miftahul Jannah | Email: nadiamj@unissula.ac.id

Received: 13 May 2025; Revised: 5 June 2025; Accepted: 11 June 2025; Published: 30 June 2025

Abstract: Fish oil contains many fatty acids (FAs) important for human health. Each fish oil tends to have a different fatty acid content. This study aims to group the fish oil content profile in several samples, such as keting fish oil (KFO), catfish oil (CFO), and pomfret fish oil (PFO), based on GC-MS analysis data with unsupervised techniques. GC-MS is a method that can be used to identify the fatty acid content in fish oil. Fish oil is extracted using the dry rendering method combined with a hydraulic press to obtain the oil, then derivatised before being analysed on the GC-MS instrument. Due to the multivariate data from GC-MS, multivariate statistical techniques are required to effectively group the fish oil samples based on their Fatty Acid profiles. Principal Component Analysis (PCA) and Cluster Analysis (CA) chemometrics are unsupervised techniques that can group multivariate data by displaying plot scores and dendrograms of sample analysis results. The fatty acid content of keting fish, catfish, and pomfret fish oil has three dominant compounds in sequence, namely oleic acid, palmitic acid, and stearic acid, with different percentages. The grouping profile of fish oil was successfully determined by PCA, total variance explained by the first four components (PC4) of 99.4%, and CA, which produced three groups based on the fatty acid content of fish oil.

Keywords: fish oil, content profile, PCA, CA

1. INTRODUCTION

Fish has a very high fatty acid (FA) content [1], fish oil contains as much as 25% saturated fatty acid and 75% unsaturated fatty acid [2]. Fish oil contains FA, which is good for health, including EPA (eicosapentaenoic acid) and DHA (docosahexaenoic acid) [3], [4], [5]. The FA composition in fish is influenced by many factors, including the type and abundance of food, habitat, size, and age [1]. So, the different composition of FA can be used as a characteristic of the type of fish oils.

Gas Chromatography-Mass Spectrometry (GC-MS) is commonly used to separate volatile organic compounds in complex mixtures [6], [7], [8], combining two methods, namely GC to analyze the type of compound qualitatively and MS to analyze quantitatively the molecular weight structure of the analyte compound. The number of variables measured simultaneously using GC-MS will produce multivariate data [9], [10].

Chemometrics is often associated with analysing multivariate data obtained from a measurement. Multivariate data, or data from measurements of many variables, has correlations between each variable [7], [11]. The advantage of multivariate analysis is that more information will

be obtained because multivariate analysis considers many variables simultaneously, compared to only considering each variable individually [12]. In addition, it can reduce noise, be more selective, and detect false samples [10], [11], [13]. The applications of these techniques in volatiles lead the research in food control and quality, for example, it could be used to determine adulteration of food, determination of different origins, different applied treatments, etc [14].

The development of chemometric methods for clustering is divided into two types, namely unsupervised clustering, such as Principal Component Analysis (PCA), Cluster Analysis (CA), and supervised clustering, such as discriminant analysis [15]. Supervised grouping input and output are known, and aim to determine the function that most closely approximates the relationship between input and output. Unsupervised clustering does not require designing the desired results, making it suitable for identifying and grouping data [16]. PCA is a statistical procedure to summarize the information content in large data tables by means of a smaller set that can be more easily visualized and analyzed. While, CA is a data analysis technique that explores the naturally occurring groups within a data set, doesn't need to group data points into any predefined groups.. PCA will reduce the dimensionality of data and identify patterns or clusters base on their volatile content [7], [17], [18], while CA can be categorize samples based on compound similarities [7].

Food authentication is critically important to prevent adulteration and ensure product quality, nutritional value, and consumer safety [19]. Consequently, reliable analytical methods for authentication are essential. Previous research has demonstrated the successful application of GC-MS combined with PCA in classifying Wistar rat oil adulterated with various agents as part of halal product authentication efforts [20]. The GC-MS-PCA technique has also proven effective in distinguishing between processed and unprocessed Mangalitsa pork fat based on the relative concentrations of dominant fatty acids, thereby supporting quality assurance measures [21]. Moreover, GC-MS analysis coupled with PCA has been successfully employed to detect camellia oil (CAO) adulterated with 11 different oils, using fatty acid, squalene, and phytosterol profiles as discriminating variables [22]. In addition, FTIR-PCA has been developed to classify Gayo Arabica coffee bean and powder samples [23]. However, based on the existing literature, no studies have been reported utilizing unsupervised clustering techniques in the analysis of GC-MS data of fish oil samples for food quality assurance.

2. MATERIALS AND METHODS

2.1. Materials

Keting fish oil (KFO) is obtained using the cabinet dryer method for approximately 24 hours, followed by a pressing process using a hydraulic press tool (150 kN) for 5 minutes. BF_3 derivatization agent is used in the derivatization process into fatty acid methyl ester (FAME) form [20], [24]. The comparison oils, catfish oil (CFO) and pomfret oil (PFO), were obtained with the same treatment.

2.2 Analysis GC-MS

KFO and comparison oils were analyzed using GC-MS (Agilent GC/MSD 5977B). Measurements were made by injecting 1 μL of sample into the GC instrument set to the column oven temperature of 150°C; split injection mode with a split ratio of 1:10, with an injection temperature of 250°C; and helium gas as the mobile phase. The GC-MS system was started with an initial oven temperature of 15°C for 1 minute, then increased to 200°C. Mass spectrum detection was carried out

in electron ionization mode by scanning at 15 to 500 m/z. The total time required to analyze a single sample was 52 minutes. Identify each peak that appears by analyzing the results of the mass spectrum obtained, compared to the mass spectrum in the MS index library.

2.3 Clustering Analysis with Unsupervised Techniques

Clustering analysis with unsupervised techniques using Principal Component Analysis (PCA) and Cluster Analysis (CA) chemometrics processed with Minitab software ver 19. The fatty acid content of fish oil is used as a variable to group various fish oils.

3. RESULTS AND DISCUSSION

3.1. Analysis GC-MS

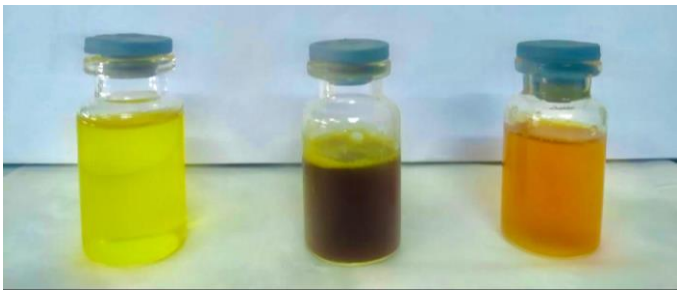
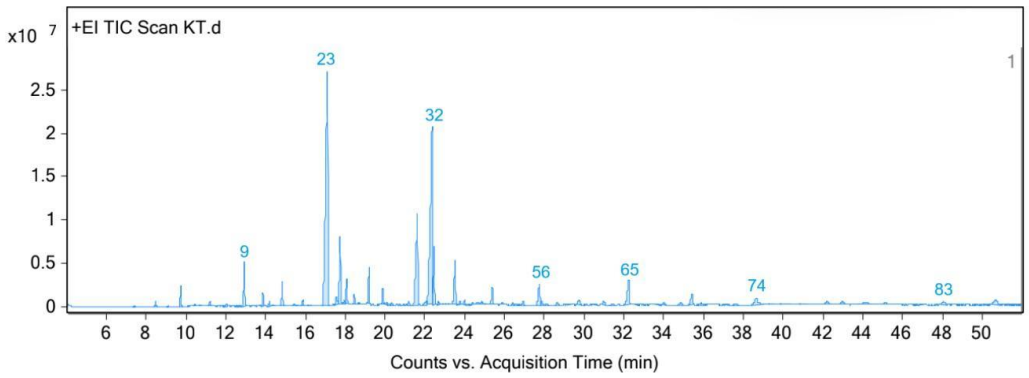


Figure 1. Fish oil extracted from (a) CFO, (b) KFO, and (c) PFO

KFO, CFO, and PFO extraction was carried out using the dry rendering method, using a cabinet dryer assisted by a hydraulic press, with the aim of providing a higher extract yield, and not using organic solvents so that it is safer for consumption. Fish oil extraction was carried out by inserting the dried sample into a porous cloth and then extracting it by pressing. The sample was extracted with a pressure of 150 kN for 5 minutes. The pressure in the extraction process aims to force the oil liquid out of the solid tissue into the gap on the side of the cylinder which is then collected in a closed container and stored [2]. for make FAME, BF_3 derivatization agent is used in the process derivatization [20], [24]. The top layer of the derivatization results is then put into a special tube, to be read on the GC-MS instrumentation system. Each peak in the mass spectrum of each fish oil sample was identified by comparing it using the Supelco 37 Component FAME Mix standard (Sigma-Aldrich) to obtain information on the detected sample FAME. Figure 2 shows the GC-MS chromatogram of each fish oil sample.



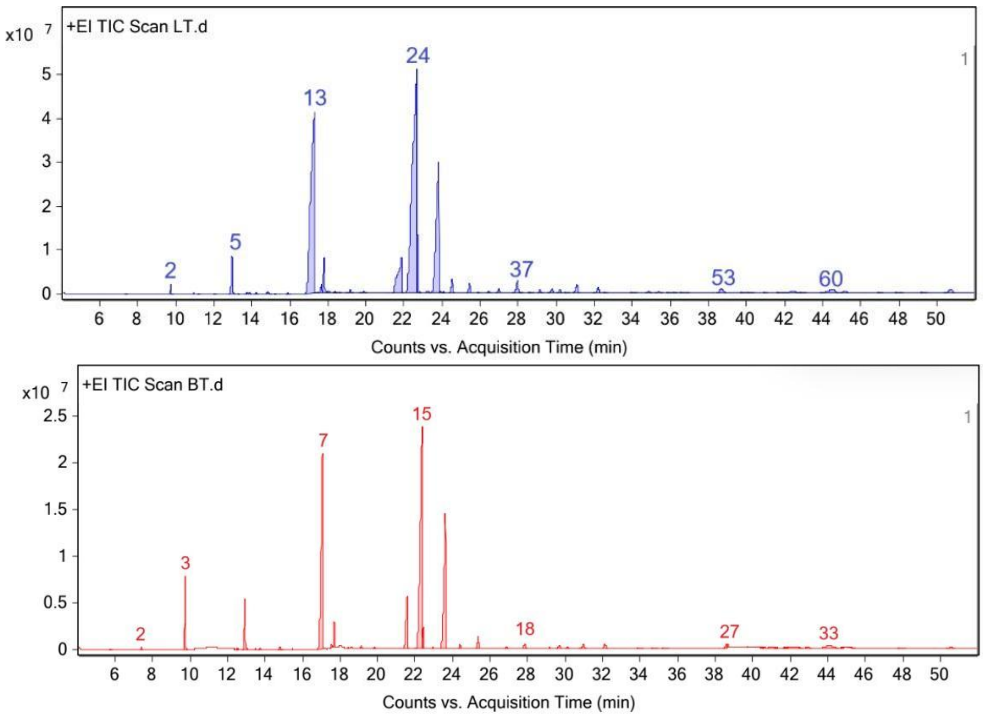


Figure 2. Chromatogram (a) KFO, (b) CFO, and (c) PFO

Based on Irnawati et al., (2024), there are differences in the chromatogram peaks identified in the three fish oil samples (KFO, CFO, and PFO) [1]. The figure shows that KFO contains no less than 27 types of organic compounds, which are fatty acids and their derivatives, CFO includes 26 kinds of fatty acids and their derivatives. At the same time, PFO has the least content of fatty acids and their derivatives compared to other fish oils, namely 21 types and their derivatives. All three samples shared three dominant compounds, namely 1) oleic acid, 2) palmitic acid, and 3) stearic acid, with different percentages. In KFO, the three dominant compounds are presented by peaks 41 (1), 21 (2), and peak 36 (3), CFO is displayed at peaks 29 (1), 13 (2), and 24 (3), while PFO at peaks 13 (1), 7 (2), and 12 (3). The relative abundance (%) and retention time (minutes) of major peaks tend to have different values , as shown in Table 1.

Table 1. This is a table. Tables should be placed in the main text near to the first time they are cited

Fatty Acids	KFO*		CFO*		PFO*	
	tR (minute)	% Relatif	tR (minute)	% Relatif	tR (minute)	% Relatif
Oleate	22.16	20.7	23.55	41.92	22.15	33.37
Palmitate	16.91	26.53	16.93	25.11	16.90	23.87
Stearate	21.47	9.31	21.48	5.42	21.45	7.17

*testing was done with 3x replication

3.2. Clustering Analysis with Unsupervised Techniques

In this study, unsupervised techniques, particularly PCA and CA, were employed to group multivariate data from GC-MS analysis of fish oil samples, specifically KFO, CFO, and PFO. Data processing is carried out at this stage by entering each type of organic compound detected in the sample (KFO, CFO, PFO) into Minitab ver 19 software (Windows). Clustering analysis with unsupervised techniques, PCA, is used to reduce variable data, where the number of variables in a matrix is reduce to produce new variables while maintaining the information owned by the data as

reported by Kartini et al., 2019 in their research through PCA was able to reduce and classify data on Plantago leaf characteristics from 7 different regions in Indonesia [25]. In addition, it was reported that PCA could display similarities between objects and relationships between variables. Several studies have reported that the use of PCA techniques has been commonly developed in the pharmaceutical world, such as Irnawati et al., (2024) were able to classify fatty acids in various marine fish oils using GC and chemometrics, Zilhadia et al., (2024) conducted an analysis of pork fat in cod liver oil emulsion, Mustafidah et al., (2021) conducted an authentication analysis of milkfish oil using a combination of FTIR spectroscopy and chemometrics [1], [26], [27].

The score plot and dendrogram are the analysis results obtained using the PCA technique. A score plot explains the similarity of a sample to another sample based on PC1 and PC2, which are marked by points that are close to each other [28]. The score plot represents a series of new coordinates along a very informative (relevant) direction; the score can be used in a two- or three-dimensional scatter diagram [29]. The closer the points displayed, the more similar the samples are. Conversely, the further the distance between points, the less similar the samples are based on the organic compound content in the sample. In comparison, the dendrogram functions to present a grouping picture based on the Euclidean distance measured based on the closest point of an object (Single linkage) [30]. The smaller the resulting value, the more similar the properties are.

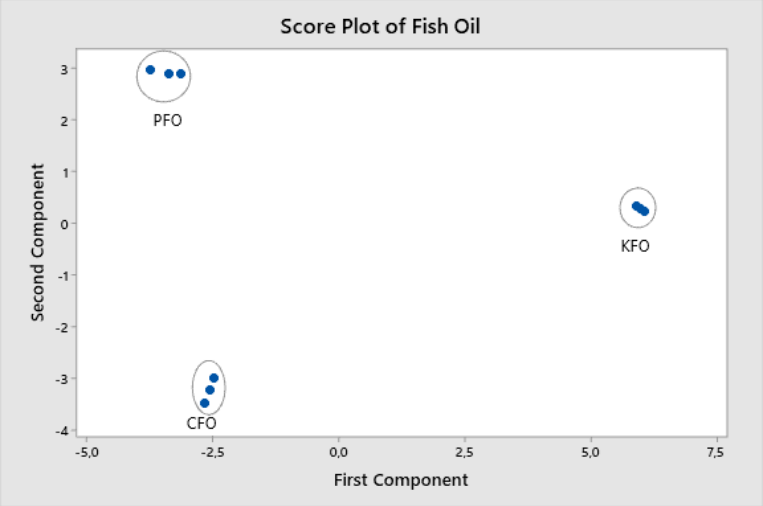


Figure 3. Score plot of fish oil

The score plot verifies what was identified by visually inspecting each chromatogram. The basis for grouping samples is generated through Eigenvalue, which is the total variance shared by 29 types of fatty acids in the sample (29 PCs), Proportion, which shows the proportion value of the total of each variance, and Cumulative, which shows the cumulative value. From the 29 PCs analyzed from the three fish oil samples, data reduction produced a new PC4 with a variance obtained reaching 99.4%. Thus, PCA can reduce data that originally had 29 variables (29 types of organic compounds) can be explained by four new variables (up to PC4), because up to PC4 it can extract information of 99.4%. In this study, the visualization of the score plot presented shows that the three fish oil samples are divided into three different quadrants, which shows that each sample has different characteristics, as seen from the fatty acid content in the sample.

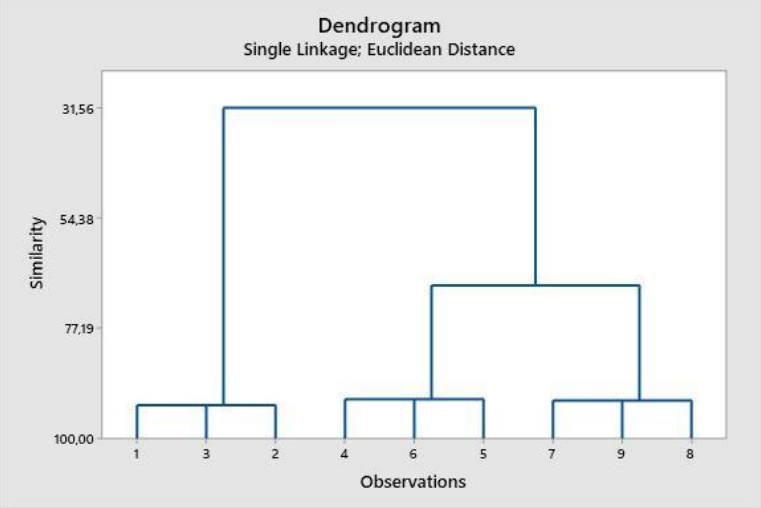


Figure 4. Dendrogram of fish oil

CA is an identification process for grouping samples based on similarity measures represented as a dendrogram. CA is based on the Euclidean distance (Ed); two or more samples with Ed equal to zero can be considered the same sample [31]. In this study, clustering can be used as a simple tool to categorise fish oil samples based on their fatty acid profiles [8]. In Figure 4, showing the dendrogram produced in this study, the data from the analysis are grouped into 2, namely the first group consisting of KFO1, KFO2, KFO3, and the second group consisting of CFO1, CFO2, CFO3, PFO1, PFO2, and PFO3. When viewed in Figure 4, this examination shows the main source of variability, showing differences between KFO with CFO and PFO.

4. CONCLUSION

The fatty acid content of keting fish, catfish, and pomfret fish oil has three dominant compounds in sequence: oleic acid, palmitic acid, and stearic acid, with different percentages. The grouping profile of fish oil was successfully determined by PCA, which produced four principal compound (PC4) was 99.4% and CA, which produced three groups based on the fatty acid content of fish oil.

Conflicts of interest: The authors declare no conflict of interest.

References

[1] I. Irnawati, A. Windarsih, L. O. M. H. Nadia, A. Rohman, and W. O. Fasrida, "Quantification and classification of fatty acids in marine fish oil from Southeast Sulawesi using gas chromatography and chemometrics," *J Appl Pharm Sci*, 2024, doi: 10.7324/JAPS.2024.159140.

[2] M. Pandiangan, J. Kaban, B. Wirjosentono, and J. Silalahi, "Analisis Kandungan Asam Lemak Omega 3 dan Omega 6 pada Minyak Ikan Mas (Cyprinus Carpio)," *st*, vol. 2, no. 1, pp. 37–44, Jan. 2019, doi: 10.32734/st.v2i1.309.

[3] N. I. Iberahim, Z. Hamzah, Y. J. Yin, and K. S. A. Sohaimi, "Extraction and Characterization of Omega-3 Fatty Acid from Catfish Using Enzymatic Hydrolysis Technique," *MATEC Web Conf.*, vol. 187, p. 01005, 2018, doi: 10.1051/matecconf/201818701005.

[4] C. Von Schacky, "Importance of EPA and DHA Blood Levels in Brain Structure and Function," *Nutrients*, vol. 13, no. 4, p. 1074, Mar. 2021, doi: 10.3390/nu13041074.

[5] W. Khalid *et al.*, "RETRACTED ARTICLE: Functional behavior of DHA and EPA in the formation of babies brain at different stages of age, and protect from different brain-related diseases," *International Journal of Food Properties*, vol. 25, no. 1, pp. 1021–1044, Dec. 2022, doi: 10.1080/10942912.2022.2070642.

[6] C.-Q. Li *et al.*, "Collaboration patterns and network in chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 191, pp. 21–29, Aug. 2019, doi: 10.1016/j.chemolab.2019.05.011.

[7] M. Tufariello, S. Pati, L. Palombi, F. Grieco, and I. Losito, "Use of Multivariate Statistics in the Processing of Data on Wine Volatile Compounds Obtained by HS-SPME-GC-MS," *Foods*, vol. 11, no. 7, p. 910, Mar. 2022, doi: 10.3390/foods11070910.

[8] R. Aparicio-Ruiz, D. L. García-González, M. T. Morales, A. Lobo-Prieto, and I. Romero, "Comparison of two analytical methods validated for the determination of volatile compounds in virgin olive oil: GC-FID vs GC-MS," *Talanta*, vol. 187, pp. 133–141, Sep. 2018, doi: 10.1016/j.talanta.2018.05.008.

[9] P. Tian *et al.*, "Analysis of volatile compound changes in fried shallot (Allium cepa L. var. aggregatum) oil at different frying temperatures by GC–MS, OAV, and multivariate analysis," *Food Chemistry*, vol. 345, p. 128748, May 2021, doi: 10.1016/j.foodchem.2020.128748.

[10] R. Bro, "Multivariate calibration," *Analytica Chimica Acta*, vol. 500, no. 1–2, pp. 185–194, Dec. 2003, doi: 10.1016/S0003-2670(03)00681-0.

[11] S. Grassi, M. Tarapoulouzi, A. D'Alessandro, S. Agriopoulou, L. Strani, and T. Varzakas, "How Chemometrics Can Fight Milk Adulteration," *Foods*, vol. 12, no. 1, p. 139, Dec. 2022, doi: 10.3390/foods12010139.

[12] S. Tortorella and S. Cinti, "How Can Chemometrics Support the Development of Point of Need Devices?," *Anal. Chem.*, vol. 93, no. 5, pp. 2713–2722, Feb. 2021, doi: 10.1021/acs.analchem.0c04151.

[13] J. L. Aleixandre-Tudo, L. Castello-Cogollos, J. L. Aleixandre, and R. Aleixandre-Benavent, "Chemometrics in food science and technology: A bibliometric study," *Chemometrics and Intelligent Laboratory Systems*, vol. 222, p. 104514, Mar. 2022, doi: 10.1016/j.chemolab.2022.104514.

[14] G. Lubes and M. Goodarzi, "Analysis of Volatile Compounds by Advanced Analytical Techniques and Multivariate Chemometrics," *Chem. Rev.*, vol. 117, no. 9, pp. 6399–6422, May 2017, doi: 10.1021/acs.chemrev.6b00698.

[15] F. Kherif and A. Latypova, "Principal component analysis," in *Machine Learning*, Elsevier, 2020, pp. 209–225. doi: 10.1016/B978-0-12-815739-8.00012-2.

[16] M. F. Dupont *et al.*, "Chemometrics for environmental monitoring: a review," *Anal. Methods*, vol. 12, no. 38, pp. 4597–4620, 2020, doi: 10.1039/D0AY01389G.

[17] M. Kamal and R. Karoui, "Analytical methods coupled with chemometric tools for determining the authenticity and detecting the adulteration of dairy products: A review," *Trends in Food Science & Technology*, vol. 46, no. 1, pp. 27–48, Nov. 2015, doi: 10.1016/j.tifs.2015.07.007.

[18] K. Tyagi, C. Rane, R. Sriram, and M. Manry, "Unsupervised learning," in *Artificial Intelligence and Machine Learning for EDGE Computing*, Elsevier, 2022, pp. 33–52. doi: 10.1016/B978-0-12-824054-0.00012-5.

[19] A. Rebiai *et al.*, "Current Application of Chemometrics Analysis in Authentication of Natural Products: A Review," *CCHTS*, vol. 25, no. 6, pp. 945–972, May 2022, doi: 10.2174/1386207324666210309102239.

[20] A. Guntarti, K. P. Ningrum, I. G. Gandjar, and N. Salamah, "AUTHENTICATION OF SPRAGUE DAWLEY RATS (RATTUS NORVEGICUS) FAT WITH GC-MS (GAS CHROMATOGRAPHY-MASS SPECTROMETRY) COMBINED WITH CHEMOMETRICS," *Int J App Pharm*, pp. 134–139, Mar. 2021, doi: 10.22159/jap.2021v13i2.40130.

[21] C. Petroman *et al.*, "Fatty Acid Profile of Lipid Fractions of Mangalitza (Sus scrofa domesticus) from Northern Romania: A GC-MS-PCA Approach," *Foods*, vol. 10, no. 2, p. 242, Jan. 2021, doi: 10.3390/foods10020242.

[22] T. Shi, G. Wu, Q. Jin, and X. Wang, "Detection of camellia oil adulteration using chemometrics based on fatty acids GC fingerprints and phytosterols GC–MS fingerprints," *Food Chemistry*, vol. 352, p. 129422, Aug. 2021, doi: 10.1016/j.foodchem.2021.129422.

[23] Z. Akbar *et al.*, "Identification of Gayo arabic coffee beans and powder using the FTIR-PCA combination method," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 1087, no. 1, p. 012059, Feb. 2021, doi: 10.1088/1757-899X/1087/1/012059.

[24] N. M. Jannah and K. P. Ningrum, "Studi perbandingan profil asam lemak minyak ikan lele dan minyak ikan keting melalui pemurnian dengan bentonit," *JPS*, pp. 360–366, Mar. 2025, doi: 10.36490/journal-jps.com.v8i1.378.

[25] K. Kartini, C. Avanti, C. Phechkrajang, and O. Vallisuta, "Antioxidant Activity, HPTLC Fingerprint and Discriminant Analysis of Plantago major Leaves from Diverse Origins in Indonesia," *PJ*, vol. 11, no. 6s, pp. 1483–1489, Dec. 2019, doi: 10.5530/pj.2019.11.229.

[26] Z. Zilhadia, Y. Anggraeni, Y. F. Apriyanti, M. Mustafidah, and I. Jaswir, "Analysis of lard in cod liver oil emulsion using FTIR spectroscopy combinedprincipal component analysis," *Food Res.*, vol. 8, no. 3, pp. 424–431, Jun. 2024, doi: 10.26656/fr.2017.8(3).329.

[27] M. Mustafidah, I. Irnawati, E. Lukitaningsih, and A. Rohman, "Authentication analysis of milkfish fish oil using the combination of FTIR spectroscopy and chemometrics," *Food Res.*, vol. 5, no. 2, pp. 272–278, Mar. 2021, doi: 10.26656/fr.2017.5(2).607.

[28] R. Bro and A. K. Smilde, "Principal component analysis," *Anal. Methods*, vol. 6, no. 9, pp. 2812–2831, 2014, doi: 10.1039/C3AY41907J.

[29] A. Biancolillo and F. Marini, "Chemometric Methods for Spectroscopy-Based Pharmaceutical Analysis," *Front. Chem.*, vol. 6, p. 576, Nov. 2018, doi: 10.3389/fchem.2018.00576.

[30] I. Widyastuti, H. Z. Luthfah, Y. I. Hartono, R. Islamadina, A. T. Can, and A. Rohman, "Antioxidant Activity of Temulawak (Curcuma xanthorrhiza Roxb.) and its Classification with Chemometrics," *Indonesian J. Chemom.Pharm. Anal*, p. 29, Jul. 2020, doi: 10.22146/ijcpa.507.

[31] N. Kadiasi, R. Tako, A. Ibraliu, V. Stanys, and N. S. Gruda, "Principal Component and Hierarchical Cluster Analysis of Major Compound Variation in Essential Oil among Some Red Oregano Genotypes in Albania," *Agronomy*, vol. 14, no. 7, p. 1419, Jun. 2024, doi: 10.3390/agronomy14071419.