Vietnam Journal of Food Control
A Scientific Journal of National Institute for Food Control
Journal homepage: https://vjfc.nifc.gov.vn

Research Article

Preliminary study on distinguishing commercial orange juices based on chemical composition analysis at the laboratory scale

Le Viet Ngan[1*], Do Truc Quynh[1], Trinh Yen Nhi[1], Vu Thi Nhat Le[1], Do Thi Hong Thuy[1], Tran Huy Hieu[1], Vu Thi Trang[1], Le Thi Hong Hao[1], Ta Thi Thao[2], Nguyen Duc Thanh[3]

[1]National Institute for Food Control, Hanoi, Vietnam
[2]University of Science, Vietnam National University, Hanoi, Vietnam
[3]Vietnam Military Medical University, Hanoi, Vietnam

(Received: 04 Aug 2025, Revised: 20 Oct 2025, Accepted: 21 Oct 2025)

Abstract

Orange juice is among the most consumed fruit juices worldwide, valued for its nutritional and sensory qualities. However, due to its high economic value, orange juice is also one of the most vulnerable commodities to adulteration, particularly through dilution or substitution with cheaper ingredients. This study aimed to differentiate between natural and commercial orange juice and to evaluate the ability to detect adulteration by dilution with sucrose and citric acid solutions using multivariate statistical approaches. A total of 50 samples, including fresh-squeezed orange juice, commercial juices, and artificially adulterated juices at different levels (10%, 20%, 30%, 50%, and 90%), were analyzed for organic acid composition (oxalic acid, malic acid, ascorbic acid, citric acid, and furmaric acid), flavonoids (hesperidin, narirutin) analyzed by HPLC-PDA and sugars (fructose, glucose, and sucrose) analyzed by HPLC-RID. Principal Component Analysis (PCA) was first applied to reduce data dimensionality and to identify major compounds contributing to the variation among samples. Subsequently, Discriminant Analysis based on PCA scores (PCA-DA) was used for classification and model validation. The results showed that PCA successfully separated natural juices from commercial ones, with citric acid, glucose, fructose, hesperidin, and narirutin being the main discriminant markers. For the adulterated juice experiment, PCA-DA achieved clear separation of groups, with 100% correct classification when three principal components were included in the model. These findings confirm that combining chemical profiling with PCA and PCA-DA is a reliable and efficient approach for authenticity testing of orange juice. The method provides a promising analytical tool for quality control laboratories and regulatory agencies to combat economically motivated adulteration in fruit juices.

Keywords: *PCA, PCA-DA, fresh orange juice, commercial orange juice, HPLC.*

1. INTRODUCTION

Orange juice is defined as the liquid extracted from the fruit *Citrus sinensis*, widely consumed worldwide due to its appealing sensory properties and high nutritional value, including abundant vitamin C, antioxidants, and carotenoids. The production process of orange juice is primarily carried out through industrial processes to maintain quality, extend shelf life, and prevent spoilage as well as enzyme activity [1]. According to the World Integrated Trade Solution database, in 2023, the global import volume and value of fresh or dried oranges (HS code 080510) reached 6.439 million tons and 6.409 billion USD, respectively. Among juice products, not-from-concentrate (NFC) orange juice is considered the highest quality type, holding a large share in orange juice consumption in developed countries and gradually expanding its market share in emerging markets [2, 3]. Although considered premium, NFC still differs in flavor and quality compared to directly

*Corresponding author: Le Viet Ngan (Email: nganbmt113@gmail.com)
https://doi.org/10.47866/2615-9252/vjfc.4632

squeezed fresh orange juice. Exploiting this, a form of Economically Motivated Adulteration (EMA) may occur when manufacturers label NFC on products made from low-quality fresh orange juice or blends, to gain higher profits.

Methods for analyzing the authenticity of fruit juices are classified into two types: targeted analysis (detecting specific markers) and untargeted analysis (not searching for specific markers) [5, 6].

In targeted analysis, characteristic compounds such as sugars, organic acids, phenolic acids, and flavonoids are often selected as markers, then analyzed using modern separation techniques like High-performance liquid chromatography (HPLC), Gas chromatography (GC), Thin-layer chromatography (TLC), or Capillary electrophoresis [5, 7]. These techniques allow for rapid and accurate qualitative and quantitative analysis of characteristic compounds, thereby serving as a basis for evaluating juice authenticity. For example, HPLC-DAD-ESI-MS/MS has been applied to differentiate citrus juices and detect adulteration based on polyphenol profiles [8]. Additionally, LC-MS combined with machine learning has been used to classify NFC orange juice and concentrated orange juice [9].

In addition to chromatography, spectroscopic techniques are also extensively exploited for juice authentication due to their advantages of being fast, non-destructive, and capable of generating multidimensional data for statistical analysis. Typical examples include Infrared (IR, NIR), Fluorescence, UV-Vis, and Raman spectroscopy [10, 11]. Studies have demonstrated the ability to quantify adulteration ratios or detect the presence of adulterants based on spectral signals, when combined with multivariate data processing methods such as SIMCA, PLS-DA, or soft-PLS-DA.

A recent trend is the application of untargeted analysis, particularly with the support of chromatography coupled with mass spectrometry (LC-MS, GC-MS), to build comprehensive metabolic fingerprints for juices. This method allows for detecting fraud at very low levels (down to 1 - 5%) and clearly distinguishing pure juices from those adulterated with various other fruits [12-16].

In Vietnam, research on classification and traceability is still new, with no studies on orange juice authentication. Some related studies include: Classification of honey based on ^{1}H-NMR analysis and using MCA and PCA for data processing [17]; Using ICP-OES/AAS and Linear Discriminant Analysis (LDA) to group teas based on metals [18]; NIR for classifying tea origins and quantifying methylxanthines [19].

Therefore, the objective of this study is to develop a simple, sensitive targeted analysis method to differentiate original orange juice from adulterated ones, through data processing of chemical composition (sugars, organic acids, and phenolics) using PCA and PCA-DA.

2. MATERIALS AND METHODS

2.1. Reagents and samples

Standards: Acid ascorbic (Sigma-Aldrich, purity ≥ 99%); acid citric (Sigma-Aldrich, purity ≥ 99%); acid malic (Sigma-Aldrich, purity ≥ 99%); acid oxalic (Sigma-Aldrich, purity ≥ 99%); acid fumaric (Sigma-Aldrich, purity ≥ 99%); fructose (Sigma-Aldrich, purity ≥ 99%); glucose (Sigma-Aldrich, purity ≥ 99%); sucrose (Sigma-Aldrich, purity ≥ 99%); Hesperidin (Sigma-Aldrich, purity ≥ 95%); narirutin (Sigma-Aldrich, purity ≥ 95%).

Chemicals: Potassium dihydrogen phosphate (Merck, purity ≥ 98%); Ortho-phosphoric acid (Merck, purity ≥ 98%); Meta-phosphoric acid (Merck, purity ≥ 98%); Acetonitrile (Merck, purity ≥ 98%); Dichloromethane (Merck, purity ≥ 98%); Methanol (Merck, purity ≥ 98%); n-Hexane, Ethyl acetate (Merck, purity ≥ 98%). Ultrapure water was obtained from a Milli-Q water filtration system (Merck).

This study used 30 fresh orange samples and 20 commercial orange juice products purchased from the market. All samples were collected from different supermarkets in Hanoi.

Sample preparation: Fresh orange samples were squeezed, blended, and filtered. Commercial juice samples were centrifuged for 10 min at 6000 rpm.

In the orange juice adulteration study, to simulate the production process, some fresh orange juice samples (n = 4) were adulterated with a mixture of 6% sucrose and 6% citric acid at ratios: 10%, 20%, 30%, 50%, and 90% to obtain solutions with total sugar content from 4 - 10% and citric acid from 0.5% to 1.5%. Each sample was analyzed in duplicate, and the average value was taken.

2.2. Equipment

HPLC system coupled with Refractive Index Detector (HPLC-RID) (Shimadzu 20A), HPLC-PDA (Shimadzu 10A), HPLC-PDA (Waters), High-speed centrifuge (MIRKO-220R).

2.3. Analytical methods

Orange juice contains various components, among which sugars and organic acids determine flavor, stability, and material characteristics, while flavonoids represent biological value and antioxidant capacity. This set of indicators also supports origin authentication, fraud detection, and monitoring changes during processing and preservation.

2.3.1. Sugar composition analysis

The method for analyzing sugars (fructose, glucose, and sucrose) was developed by the National Institute for Food Control and has been accredited according to ISO 17025 [21] with limit of determination (LOD): 0.015 g/100 mL.

Samples were filtered through a 0.45 µm membrane, then directly analyzed on an HPLC-RID system (Shimadzu 20A) using a Sugar SZ5532 column (6.0 mm × 150 mm; 6 µm) (Shodex). Mobile phase: acetonitrile: water mixture (8 : 2, v/v); flow rate: 0.8 mL/min; column temperature: 60°C; injection volume: 20 µL. Each sample was analyzed in duplicate, and the mean value was used.

2.3.2. Organic acid content analysis

The method for analyzing organic acids (oxalic acid, malic acid, ascorbic acid, citric acid, and fumaric acid) was developed by the National Institute for Food Control and has been accredited according to ISO 17025 [21] with LOD is equal to 8 mg/L.

Samples were filtered through a 0.45 µm membrane, then directly analyzed on an HPLC-PDA system (Shimadzu 10A), wavelength 210 nm. Using SUPELCOGEL™ C-610H column (300 mm × 7.8 mm; 9 µm), SUPELCOGEL™ H Guard Column (50 mm × 4.6 mm, 9 µm) (Supelco); mobile phase: 0.1% orthophosphoric acid solution in water; column temperature: 30°C; flow rate: 0.6 mL/min; and injection volume: 20 µL. Each sample was analyzed in duplicate, and the average value was taken for the final result.

2.3.3. Flavonoid content analysis

The method for analyzing flavonoids (hesperidin, narirutin) was referenced from Belajová, E. and Suhaj, M. [22]. Samples were diluted with a methanol: water mixture (7 : 3, v/v), then filtered through a 0.45 µm membrane and directly analyzed on an HPLC-PDA system (Waters) at wavelength 285 nm (hesperidin, narirutin). LOD is equal to 1 mg/L.

Using XBridge® C18 column (4.6 mm × 150 mm, 5 µm; Waters), with mobile phase consisting of: Channel A: 0.1% orthophosphoric acid solution in water; and Channel B: methanol.

Elution program: from 0 to 7 min, Channel B increases from 10% to 20%; from 7 to 16 min, Channel B increases from 20% to 35%; from 16 to 22 min, Channel B increases from 35% to 100%; from 22 to 26 min, Channel B decreases from 100% to 10% and maintains at 10% until 30 min; column temperature 30°C, flow rate 1.0 mL/min, injection volume 20 µL. Each sample was analyzed in duplicate, and the average value was taken for the final result.

2.4. Data collection

The results of sugar, organic acid, and flavonoid analyses were processed using Microsoft Excel 2016 software and statistical processing using XLSTAT 2025 software.

3. RESULTS AND DISCUSSION

3.1. Sugar, organic acid, and flavonoid analysis in fresh orange juice and commercial orange juice

The results of analyzing sugar content, organic acids, and flavonoids in fresh orange juice (F) and commercial orange juice (C) are shown in **Table 1**.

The analysis of 30 natural orange juice samples (F) and 20 commercial orange juice samples (C) showed clear differences in organic acid, sugar, and flavonoid compositions.

Table 1. *Sugar content, organic acids, and flavonoids in fresh orange juice and commercial orange juice*

Sample	Parameter	Organic acid content (mg/L)					Sugar content (mg/100 mL)			Flavonoid content (mg/L)	
		OA	MA	AA	CA	FA	GLU	FRU	SUC	HES	NAR
Fresh orange juice (n = 30)	Min	122.9	116.6	103.5	3946.1	13.3	1600.0	1510.0	3224.0	36.7	4.3
	Max	482.4	910.0	998.2	9951.4	347.8	4060.0	3960.0	7500.0	512.5	56.4
	Mean	242.0	468.3	529.9	7502.0	57.6	2440.7	2497.3	5407.1	263.0	35.0
	SD	84.8	232.4	238.1	1580.2	65.0	626.8	596.1	1181.6	119.3	13.7
Commercial orange juice (n = 20)	Min	15.0	10.9	17.6	3574.8	15.4	170.0	160.0	4157.0	20.7	3.3
	Max	94.0	92.5	464.0	7643.4	95.5	2140.0	2870.0	8838.0	141.2	15.8
	Mean	49.8	52.6	202.9	5128.5	55.9	637.4	677.1	7492.7	76.0	9.5
	SD	25.1	24.9	113.9	1252.9	26.3	448.1	597.6	1127.4	38.4	4.4

Note: OA: oxalic acid, MA: malic acid, AA: ascorbic acid, CA: Citric acid, FA: furmaric acid, GLU: glucose, FRU: fructose, SUC: sucrose, HES: hesperidin, and NAR: narirutin; Min: minimum, Max: maximum, and SD: standard deviation.

Organic acids: The contents of oxalic acid, malic acid, ascorbic acid (vitamin C), and citric acid were significantly higher in group F compared to group C (Mann-Whitney U, $p < 0.05$). These are important indicators contributing to the sour taste and nutritional value of natural orange juice. In contrast, fumaric acid showed no significant difference between the two groups (Mann-Whitney U, $p > 0.05$), indicating that this variable is not a suitable indicator for distinguishing orange juice origins in this study.

Sugar content: Glucose and fructose in group F were approximately 3-4 times higher than in group C (Mann-Whitney U, $p < 0.05$), reflecting the characteristics of natural orange juice, which is rich in monosaccharides. Meanwhile, sucrose was significantly higher in group C (Mann-Whitney U, $p < 0.05$), consistent with commercial orange juice often being supplemented with sugar during production. Therefore, the ratio (glucose + fructose) / sucrose can be used as a characteristic indicator to differentiate the two groups of orange juice.

Flavonoids: The two main compounds, hesperidin and narirutin, had much higher concentrations in natural orange juice (Mann-Whitney U, $p < 0.05$). This is an important feature because flavonoids are concentrated in the pulp and membranes of the fruit, which are easily removed or degraded during industrial filtration, concentration, and sterilization processes. Therefore, flavonoids, especially hesperidin, can be considered typical biological markers for the authenticity of orange juice.

Overall, the indicators of organic acids (oxalic, malic, ascorbic, and citric), monosaccharides (glucose, fructose), and flavonoids (hesperidin, narirutin) showed very large and statistically significant differences between the two groups, indicating they are strong discriminant variables between natural and commercial orange juice. In contrast, fumaric acid has little value in classification. Particularly, the trend of increased sucrose and decreased flavonoids and vitamin C in the commercial group aligns with the realities of concentrated orange juice production technology and flavor adjustment.

When evaluating the correlation matrix between the two orange juice groups with the 10 variables above, the correlation matrix is shown in **Table 2**.

Pearson correlation analysis showed that glucose and fructose have a strong relationship (r = 0.975), and both vary in parallel with oxalic acid and flavonoids (hesperidin, narirutin). In contrast, sucrose has an inverse relationship with glucose and fructose, reflecting differences in sugar metabolism between samples. Meanwhile, fumaric acid is almost unrelated to the other components, showing its distinct variation pattern.

Table 2. *Pearson correlation matrix between the 10 analyzed compounds*

Variable	OA	MA	AA	CA	FA	GLU	FRU	SUC	HES	NAR

OA	1	0.590	0.405	0.405	-0.068	0.732	0.720	-0.629	0.701	0.725
MA	0.590	1	0.510	0.504	-0.073	0.665	0.641	-0.446	0.416	0.520
AA	0.405	0.510	1	0.330	0.139	0.481	0.436	-0.450	0.449	0.440
CA	0.405	0.504	0.330	1	-0.167	0.500	0.478	-0.462	0.355	0.434
FA	-0.068	-0.073	0.139	-0.167	1	-0.159	-0.155	-0.063	-0.153	-0.136
GLU	0.732	0.665	0.481	0.500	-0.159	1	0.975	-0.513	0.602	0.662
FRU	0.720	0.641	0.436	0.478	-0.155	0.975	1	-0.535	0.600	0.671
SUC	-0.629	-0.446	-0.450	-0.462	-0.063	-0.513	-0.535	1	-0.568	-0.543
HES	0.701	0.416	0.449	0.355	-0.153	0.602	0.600	-0.568	1	0.932
NAR	0.725	0.520	0.440	0.434	-0.136	0.662	0.671	-0.543	0.932	1

Note: OA: oxalic acid, MA: malic acid, AA: ascorbic acid, CA: Citric acid, FA: furmaric acid, GLU: glucose, FRU: fructose, SUC: sucrose, HES: hesperidin, and NAR: narirutin.

Principal Component Analysis (PCA) with the Scree Plot is shown in **Figure 1**.

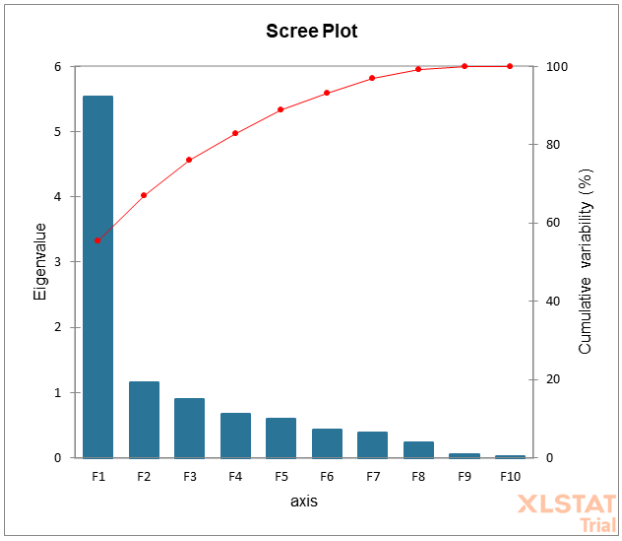

Figure 1. *Scree Plot showing eigenvalues of Principal Components in PCA analysis between fresh and commercial orange juice*

Principal Component Analysis (PCA) shows that F1 is the most important axis, with an eigenvalue of approximately 5.5, explaining 55.4% of the data variance. This component is closely related to compounds such as citric acid, malic acid, oxalic acid, glucose, fructose, hesperidin, and narirutin, reflecting the general nutritional characteristics of orange juice dominated by major organic acids, monosaccharides, and flavonoids. The principal component F2 has an eigenvalue of about 1.2, explaining an additional 11.6% variance, mainly influenced by fumaric acid, sucrose, and ascorbic acid, indicating the role of these substances in creating secondary differences between samples. Thus, the two components F1 and F2 together explain about 67% of the total data variation. Components from F3 onward have eigenvalues < 1 and contribute < 10% variance, so they do not provide much additional information. The elbow on the Scree plot appears after F2, so only two principal components need to be retained to describe the main trends and visualize the data.

To visualize the relationships between variables and sample separation according to the two principal components, a biplot was constructed in **Figure 2.**

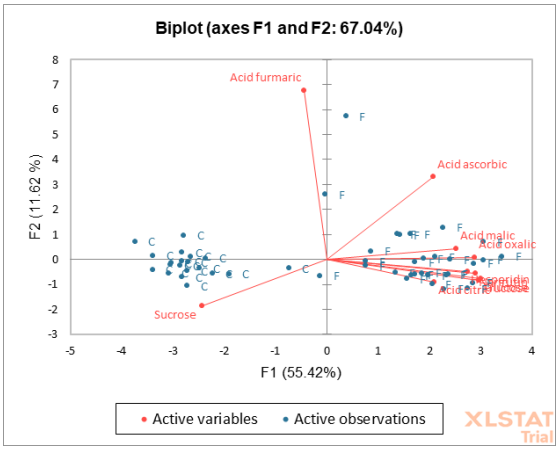

Figure 2. *PCA Biplot showing relationships between chemical variables (organic acids, sugars, flavonoids) and distribution of fresh and commercial orange juice samples*

The PCA biplot above shows a clear separation between the two groups of orange juice samples on the two principal components. Fresh orange juice samples (F) are concentrated on the positive side of F1, associated with high contents of organic acids (citric, malic, oxalic, and ascorbic), monosaccharides (glucose, fructose), and flavonoids (hesperidin, narirutin). In contrast, commercial orange juice samples (C) are distributed on the negative side of F1, mainly dominated by sucrose and fumaric acid. F1 explains 55.4% and F2 explains 11.6% of the data variance, total of 67%, indicating that these two components represent most of the variation and are sufficient to distinguish the two sample groups.

To evaluate supervised classification ability, the first two principal components (F1 and F2) explaining 67% of the total data variance were used as input for Linear Discriminant Analysis (PCA-DA). This method allows testing the separation ability between the two sample groups (natural orange juice and commercial orange juice) as well as determining the accuracy in sample classification. The PCA-DA results are shown in **Figure 3.**

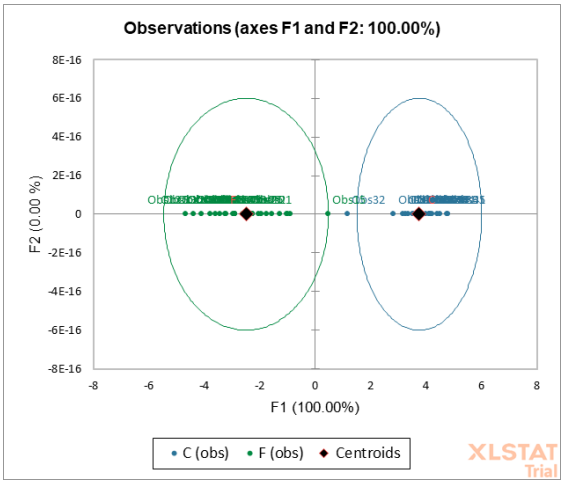

Figure 3. *PCA-DA Scatter Plot showing separation between fresh and commercial orange juice based on two principal components*

The PCA-DA scatter plot with centroids shows that the two groups of natural orange juice samples (F) and commercial (C) are completely and clearly separated on the F1 axis (100% variance). The observation points of each group are tightly clustered around their respective centroids with no overlap between the two groups. This demonstrates that the constructed PCA-DA model has the ability to distinguish the two types of orange juice, reflecting clear differences in chemical composition between natural and commercial orange juice. The confusion matrix for the training set shows that all 20 commercial samples and all 30 fresh samples were correctly classified into their respective groups, achieving 100% classification accuracy.

The application of combining PCA and PCA-DA allows for clear and accurate differentiation between fresh orange juice and commercial orange juice. PCA shows that the first two principal components explain 67% of the total variance, with F1 closely related to organic acids (citric, malic, and oxalic), monosaccharides (glucose, fructose), and flavonoids (hesperidin, narirutin), while commercial orange juice is mainly related to higher sucrose and fumaric acid contents. The subsequent PCA-DA analysis confirmed this difference, achieving 100% correct classification rate, as shown in both the scatter plot with group centroids and the confusion matrix. These results prove that hesperidin, narirutin, and citric acid, along with the monosaccharide-to-sucrose ratio, play roles as strong marker compounds for authenticity verification. Fresh orange juice is characterized by naturally higher flavonoid and organic acid contents, while commercial orange juice has higher sucrose and lower flavonoid contents, reflecting industrial processing and formulation. Therefore, the chemometric method combining PCA and PCA-DA has proven to be a powerful tool for authenticating and evaluating orange juice quality.

3.2. Orange juice adulteration experiment analysis

Analysis was performed on 4 orange juice samples adulterated with a mixture of 6% sucrose and 6% citric acid at ratios: 10%, 20%, 30%, 50%, and 90% to obtain solutions with total sugar content from 4 - 10% and citric acid from 0.5% to 1.5%. Each sample was analyzed in duplicate, and the average value was taken. The results are shown in **Table 3**, where the ratio of natural orange juice decreases from 100% to 10%.

Table 3. *Results of analyzing sugar content, organic acids, and flavonoids in the adulteration experiment*

Sample	Mixture ratio (%)	Organic acid content (mg/L)					Sugar content (mg/100 mL)			Flavonoid content (mg/L)	
		OA	MA	AA	CA	FA	GLU	FRU	SUC	HES	NAR
Orange juice 1	0	310	559	776	6408	103	2790	2830	5600	226	33.6
	10	285	546	680	6947	97.4	2448	2312	5241	207	29.9
	20	240	441	625	6682	90.8	2349	2170	5349	164	29.2
	30	221	383	562	5867	67	1772	1840	5631	162	25.8
	50	168	287	1232	6185	50.7	1493	1304	5959	236	19.5
	90	30	59	304	6076	11.8	< 300	< 300	5937	49.2	4.2
Orange juice 2	0	193	853	582	7186	36.9	2140	2390	7010	106	12.1
	10	287	497	743	5821	83.9	2646	2695	5912	220	28.5
	20	245	468	597	6238	85.4	2081	2155	5229	163	26.9
	30	147	596	413	7471	28.1	1512	1803	7254	73.9	9.05
	50	94	417	292	6159	17.2	1002	1108	6305	57.2	5.97
	90	18	78	62	6406	ND	< 300	< 300	6139	9.89	< 3
Orange juice 3	0	229	184	801	8291	13.3	3270	3250	6549	249	42.3
	10	274	482	737	6127	91.9	2628	2725	5912	207	31.6
	20	249	444	666	6486	81.8	2185	2348	5664	184	25.7
	30	150	136	610	8055	9.6	2352	2285	6732	169	30.1
	50	113	83	376	7196	ND	1511	1746	6414	123	23.1
	90	25	19	86	6500	ND	< 300	< 300	6371	25.9	4.05
Orange juice 4	0	293	502	258	8186	17.1	2050	2570	5846	312	56.4
	10	288	461	763	6438	96	2632	2595	5677	211	29.9
	20	249	418	648	6837	84.3	2334	2458	5499	180	24.5
	30	212	336	189	7553	11.6	1315	1807	5552	240	35.7
	50	151	269	139	7003	ND	1000	1296	6023	159	25.7
	90	28	53	25	6475	ND	< 300	< 300	5763	33.8	5.39

Note: ND: Not detected, OA: oxalic acid, MA: malic acid, AA: ascorbic acid, CA: Citric acid, FA: furmaric acid, GLU: glucose, FRU: fructose, SUC: sucrose, HES: hesperidin, and NAR: narirutin.

After analyzing pure orange juice samples and various adulteration levels (10%, 20%, 30%, 50%, and 90%), multivariate analysis methods were applied to evaluate the ability to distinguish between sample groups similar to the evaluation in **Section 3.1** for distinguishing the two orange juice groups. PCA was used to reduce data dimensions and identify main variation trends, then combined with Discriminant Analysis (DA) to test classification ability and determine compounds acting as characteristic markers for differences between orange juice groups. The results are shown in **Figure 4**, **Figure 5**, **Figure 6**, **Table 4**, and **Table 5**.

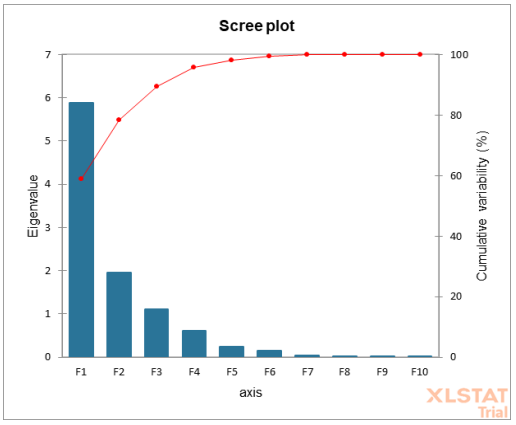

Figure 4. *Scree Plot showing eigenvalues of Principal Components in PCA Analysis for the orange juice adulteration experiment*

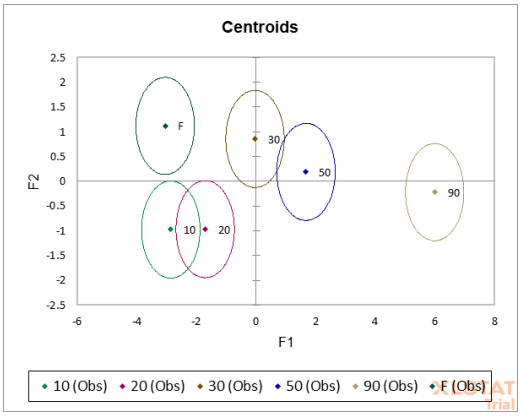

Figure 5. *PCA-DA Scatter Plot with group centroids, showing separation of orange juice adulteration levels based on two principal components*

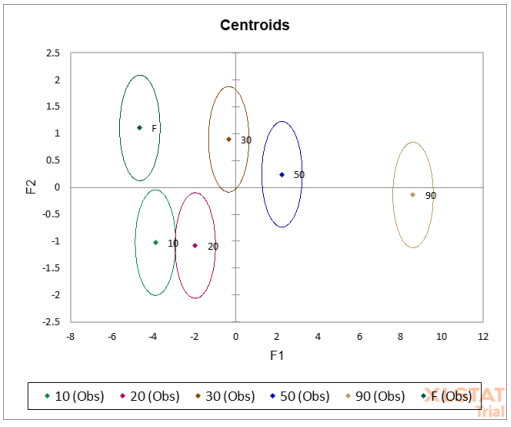

Figure 6. *PCA-DA Scatter Plot with group centroids, showing separation of orange juice adulteration levels based on three principal components*

Table 4. *Confusion matrix for training in PCA-DA analysis with 02 principal components*

Variable	10	20	30	50	90	F	Total	Exactly (%)
10	4	0	0	0	0	0	4	100
20	0	4	0	0	0	0	4	100
30	0	0	3	1	0	0	4	75.0
50	0	0	1	3	0	0	4	75.0
90	0	0	0	0	4	0	4	100
F	0	0	1	0	0	3	4	75.0
Total	**4**	**4**	**5**	**4**	**4**	**3**	**24**	**87.5**

Table 5. *Confusion matrix for training in PCA-DA analysis with 03 principal components*

Variable	10	20	30	50	90	F	Total	Exactly (%)
10	4	0	0	0	0	0	4	100
20	0	4	0	0	0	0	4	100
30	0	0	4	0	0	0	4	100
50	0	0	0	4	0	0	4	100
90	0	0	0	0	4	0	4	100
F	0	0	0	0	0	4	4	100
Total	**4**	**4**	**4**	**4**	**4**	**4**	**24**	**100**

The minimum number of principal components was determined based on the Scree plot (**Figure 4**), where the elbow indicates that at least 02 principal components should be retained. When performing DA analysis with 02 principal components, the results obtained the group centroid plot (**Figure 5**) and confusion matrix for the training set (**Table 4**) show that errors still exist in identifying samples adulterated at 30% and 50% levels, with a correct classification rate of 75%. In contrast, when performing DA analysis with 03 principal components, the group centroid plot (**Figure 6**) and corresponding confusion matrix (**Table 5**) show classification ability achieving 100% accuracy at all adulteration levels. Therefore, adding more principal components to the PCA-DA model significantly improves classification ability, from incomplete accuracy when using only 02 components to 100% when using 03 principal components. This indicates that selecting the appropriate number of components is a key factor in optimizing the model's reliability for identifying orange juice adulteration levels.

4. CONCLUSION

This study has demonstrated that combining chemical composition analysis (organic acids, sugars, and flavonoids) with multivariate statistical methods (PCA and PCA-DA) is an effective approach for evaluating authenticity and detecting adulteration in orange juice. The results show that PCA helps reduce data and identify marker compounds, while PCA-DA allows accurate classification of sample groups, especially when using three principal components, compared to 75% accuracy when using two principal components, indicating that three principal components significantly improve classification ability. For commercial and natural orange juice, the model showed clear differences in chemical characteristics, and in the adulteration study, PCA-DA achieved absolute accuracy with three principal components in identifying dilution levels at laboratory scale. The study mainly focused on method development at laboratory scale and is a preliminary study, so it is still limited in sample size. In subsequent phases, the research will continue to expand the number of samples and diversify origins to ensure representativeness and analytical reliability, as well as apply the method to analyze types of orange juice on the market.

REFERENCES

[1]. G. Galaverna, and C. Dall'Asta, "Chapter 21 - Production Processes of Orange Juice and Effects on Antioxidant Components," in *Processing and Impact on Antioxidants in Beverages*, pp. 203-214, 2014.

[2]. L. Vervoort, T. Grauwet, T. Kebede, I. van der Plancken, R.A.H. Timmermans, A. Van Loey, "Headspace fingerprinting as an untargeted approach to compare novel and traditional processing technologies: A case-study on orange juice pasteurisation," *Food Chemistry*, vol. 134, no. 4, pp. 2303-2312, 2012.

[3]. Sellami, V. Mall, and P. Schieberle, "Changes in the Key Odorants and Aroma Profiles of Hamlin and Valencia Orange Juices Not from Concentrate (NFC) during Chilled Storage," *Journal of Agricultural and Food Chemistry*, vol. 66, no. 28, pp. 7428-7440, 2018.

[4]. A. Różańska, T. Dymerski, and J. Namieśnik, "Novel analytical method for detection of orange juice adulteration based on ultra-fast gas chromatography," *Monatshefte für Chemie - Chemical Monthly*, vol. 149, pp. 1615-1621, 2018.

[5]. L. Xu, Z. Xu, and X. Liao, "A review of fruit juice authenticity assessments: Targeted and untargeted analyses," *Critical Reviews in Food Science and Nutrition*, vol. 62, no. 22, 2022.

[6]. M.E. Dasenaki, and N.S. Thomaidis, "Quality and Authenticity Control of Fruit Juices - A Review," *Molecules*, vol. 24, no. 6, pp. 1014, 2019.

[7]. S. A. Socaci, C. Socaciu, M. Tofană, I. V. Rat¸i, and A. Pintea, "In-tube extraction and GC-MS analysis of volatile components from wild and cultivated sea buckthorn (Hippophae rhamnoides L. ssp. Carpatica) berry varieties and juice," *Phytochemical Analysis*, vol. 24, no. 4, pp. 319-328, 2019.

[8]. B. Abad-García, S. Garmón-Lobato, and M.B. Sánchez-Ilárduya, L.A. Berrueta, B. Gallo, F. Vicente, R.M. Alonso-Salces, "Polyphenolic contents in Citrus fruit juices: Authenticity assessment," *European Food Research and Technology*, vol. 238, no. 5, pp. 803-818, 2014.

[9]. R. Sun, R. Xing, J. Zhang, N. Yu, Y. Ge, W. Zhang, and Y. Chen, "UPLC-QTOF-MS coupled with machine learning to discriminate between NFC and FC orange juice," *Food Control*, vol. 145, pp. 109487, 2023.

[10]. X. Gao, D. Fan, W. Li, X. Zhang, Z. Ye, Y. Meng, and T. Cheng-yi Liu, "Rapid quantification of the adulteration of pomegranate juices by Raman spectroscopy and chemometrics," *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, vol. 302, no. 5, pp. 123014, 2023.

[11]. S. Ehsani, H. Yazdanpanah, and H. Parastar, "An innovative screening approach for orange juice authentication using dual portable/handheld NIR spectrometers and chemometrics," *Microchemical Journal*, vol. 194, pp. 109304, 2023.

[12]. P. Oliveri, and R. Simonetti, "Chemometrics for food authenticity applications. Advances in Food Authenticity Testing, Woodhead Publishing Series in Food Science," *Technology and Nutrition*, pp. 701-728, 2016.

[13]. L. Vaclavik, A. Schreiber, O. Lacina, T. Cajka, and J. Hajslova, "Liquid chromatography-mass spectrometry-based metabolomics for authenticity assessment of fruit juices," *Metabolomics*, vol. 8, pp. 793-803, 2012.

[14]. Z. Jandric, D. Roberts, M.N. Rathor, A. Abrahim, M. Islam, and A. Cannavan, "Assessment of fruit juice authenticity using UPLC QToF MS: A metabolomics approach," *Food Chemistry*, vol. 148, pp. 7-17, 2014.

[15]. M. E. Dasenaki, S. K. Drakopoulou, R. Aalizadeh, and N. S. Thomaidis, "Targeted and untargeted metabolomics as an enhanced tool for the detection of pomegranate juice adulteration," *Foods*, vol. 8, no. 6, pp. 212, 2019.

[16]. Z. Jandric, and A. Cannavan, "An investigative study on differentiation of citrus fruit/fruit juices by UPLC-QToF MS and chemometrics," *Food Control*, vol. 72, pp. 173-80, 2017.

[17]. Nguyen The Anh, Pham Quang Trung, Ta Thi Thao, "Classification and identification of vietnamese honeys using chemometrics based on 1H-NMR data," *Journal of analytical sciences*, vol. 26, no. 3A, pp. 201-207, 2021 [in Vietnamese].

[18]. Nguyen Thi Thao, Hoang Quoc Tuan, Cung Thi To Quynh, and Vu Hong Son, "Research and develop a method for testing and authenticating the origin of green tea in Vietnam," *Conference proceeding - Food Control Conference 2022*, pp. 24, 2022.

[19]. Tran Thi Hue, Bui Duc Tho, Nguyen Van Ri, Ta Thi Thao, "Geographic origin classification and simultaneous determination of methylxanthines in Vietnamese tea using chemometrics based on the near infrared reflectance spectroscopy," *Conference proceeding - The 6th Analytica Vietnam Conference*, pp. 264-273, 2019.

[20]. Quality Accreditation Office, Laboratory Code 203, "Decision No: 2269/QĐ-VPCNCL dated December 14, 2023, List of Accredited Tests (No. 4)" [online]: https://nifc.gov.vn/chi-dinh-cua-van-phong-cong-nhan-chat-luong/2269-2023-q-vpcncl-post253.html [Accessed: 30/6/2025].

[21]. E. Belajová, and M. Suhaj, "Determination of phenolic constituents in citrus juices: Method of high performance liquid chromatography," *Food Chemistry*, vol. 86, pp. 339-343, 2004.