

Authenticating the Origin of Pakchoi Using Metal Content in the Edible Parts of Pakchoi at Hanoi Market, Vietnam

Mac The Vinh^{1,2}, Truong Ngoc Minh³, Nguyen Thi Huong⁴, Nguyen Thi Minh Tu^{1,5*}

¹School of Chemistry and Life Sciences, Hanoi University of Science and Technology, Ha Noi, Vietnam

²Hanoi University of Industry, Ha Noi, Vietnam

³Center for High Technology Research and Development - Vietnam Academy of Science and Technology, Ha Noi, Vietnam

⁴Hanoi University of Science and Technology, Ha Noi, Vietnam

⁵New Industry Creation Hatchery Center (NICHe), Tohoku University, Sendai, Miyagi, Japan

*Corresponding author email: tu.nguyenthiminh@hust.edu.vn

Abstract

Pakchoi (*Brassica rapa* subsp. *chinensis*) is a familiar vegetable and is widely grown in Vietnam and many other Southeast Asian countries. Like other Brassica plants, pakchoi has long been known for its ability to accumulate a variety of metals. Using the metal content data in the edible parts of pakchoi from Chuong My, Me Linh, and Dan Phuong (3 areas specializing in growing pakchoi serving Hanoi market) during the period from September 2021 to April 2024, this study evaluated the ability to authenticate the origin of pakchoi using multivariate statistical analysis methods (Principal Component Analysis-PCA and Linear Discriminant Analysis-LDA). 27 kinds of metal in the edible parts of pakchoi from these areas, including Li, Bo, Mg, Al, Ti, V, Cr, Mn, Zn, Fe, Co, Ni, Cu, As, Rb, Sr, Nb, Mo, Ag, Cd, Sb, Cs, Ba, Hg, Tl, Pb, and Bi, were determined by Inductively Coupled Plasma Mass Spectrometry (ICP-MS). The PCA and LDA methods were used and LDA method was successful in distinguishing the three geographical areas. Of the total 27 metals, the characteristic metals used for distinction in LDA were shown, including: As, Ba, Co, Cs, Li. In addition, the model constructed by LDA showed the ability to accurately recognize 27/27 prepared prediction samples. This research shows the potential of using metal content results in building a model to authenticate the geographical origin of pakchoi in particular and other vegetable crops in general in the food market in Hanoi, Vietnam.

Keywords: Pakchoi, Hanoi, metal content, ICP-MS, LDA, PCA.

1. Introduction

Pakchoi (*Brassica rapa* subsp. *chinensis*) is a vegetable originating from China that has long been widely used in Southeast Asian countries such as Vietnam, Laos, Indonesia, etc. Pakchoi has a yield of 10-20 tons/ha for small varieties and 20-30 tons/ha for large varieties [1]. With the ability to withstand heat, humidity, and pests, pakchoi has become one of the popular vegetables on the daily dining table of many Vietnamese people. Depending on characteristics, habitat, and morphology, pakchoi in the world is divided into many types: Tai Sai pakchoi, White pakchoi, Green Fortune pakchoi [2], Dwarf Carton White pakchoi, Joi pakchoi, Meiquin pakchoi, etc. In Vietnam, Green Fortune pakchoi is grown and used the most.

In addition to being rich in vitamins (A, B, C) and flavonoids (633-982 µg/g carotenoids, of which lutein accounts for 40-43%, violaxanthin 17-28%, neoxanthin 13%, and β-carotene 19-27%),

pakchoi also contains minerals: calcium, manganese, potassium, zinc, iron, sodium, magnesium, selenium, phosphorus, etc. [3]. This makes the determination of metal content components in pakchoi potentially a sign to support the geographical traceability of pakchoi.

Nowadays, the problem of overuse of chemicals and pesticides in fruits and vegetables is increasingly appearing, along with the appearance of fruit and vegetable products, food of unknown origin competing directly with domestic products. To regain consumer trust, tracing the origin of food products in general and fruits and vegetables in particular is more necessary than ever.

2. Materials and Methods

2.1. Sampling and Storing Samples

Pakchoi is grown and collected from the stems and leaves (edible parts) in Chuong My, Dan Phuong, and Me Linh areas with specific addresses listed in Table 1 (3 locations per area, 9 locations total). These

are the main areas with agricultural cooperatives mainly growing pakchoi to serve the Hanoi market from September 2021 to the end of April 2024. The edible parts of pakchoi, including stems and leaves, were sampled in the 5th week of growth (harvest time) and dried at 70 °C until the moisture content was 10-11%. All samples were then ground into powder (particle size is less than 400 µm) before being stored in clean polyethylene zip bags and finally stored in a dry place.

2.2. Digestion of Samples

0.1 g of pakchoi edible parts samples were pre-digested with 4 mL HNO₃ (65%) and 2 mL H₂O₂ (30%) and for at least 24 hours in a laboratory fume hood before being digested using the microwave digestion system-MARS 6. The pakchoi edible parts digestion program is shown in Fig. 1. Pakchoi edible parts samples were measured for absolute moisture content in parallel with the digestion of the samples to obtain results based on absolute dry matter content. Samples after digestion were then collected for metal content analysis by inductively coupled plasma mass spectrometry (ICP-MS).

The 65% nitric acid (HNO₃) and 30% hydrogen peroxide (H₂O₂) used were from Merck, USA. Ultrapure deionized water with a resistivity of 18.2 mΩcm was supplied from a Milli-Q plus water purification system (Millipore, Bedford, MA, USA). The standards used in the ICP-MS analysis included the metals Li, B, Mg, Al, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, Zn, As, Rb, Sr, Nb, Mo, Sb, Cs, Ba, Hg, Tl, Pb, and Bi.

The data are the average values of three analyses of metal content in pakchoi samples in one location.

The metal content of one sample was calculated as below:

$$\text{Con. (mg/kg)} = \frac{C \times V}{W \times S} \quad (1)$$

where:

C: Value obtained from the device (ppb or mg/L)

V: Volume of sample after digestion (L)

W: Wet sample weight (kg)

S: As % dry matter/100

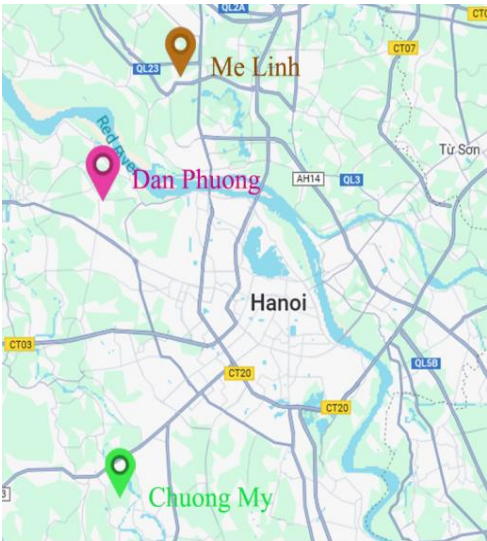
% dry matter = (g absolute dry sample)/(g sample) × 100

The data from samples in different areas were then processed by Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Data were analyzed using Microsoft Excel 2023, upgraded with XLSTAT software version 2024. The data used for prediction in the LDA method is the metal content of 27 extra samples from 3 areas, including Chuong My, Dan Phuong, and Me Linh.

2.3. Methods

PCA is a multivariate statistical method developed by Pearson (1901) [4] and Hotelling (1933) [5]. Building a new coordinate plane from the original variables, PCA helps clarify the data structure. Therefore, PCA is commonly used to reduce data dimensionality. In multivariate statistical analysis, PCA serves as an exploratory method. In food chemistry studies, PCA has become increasingly useful due to its ease of interpretation and discussion, especially when large data sets are analyzed [6].

Table 1. List of pakchoi cultivation locations

	Chuong My	20°51'45.6"N, 105°39'40.0"E 20°51'55.3"N, 105°39'25.9"E 20°52'02.0"N, 105°39'40.0"E
	Dan Phuong	21°05'06.0"N, 105°42'22.3"E 21°05'08.9"N, 105°42'25.4"E 21°05'11.8"N, 105°42'24.2"E
	Me Linh	21°10'19.7"N, 105°45'38.3"E 21°10'27.6"N, 105°45'31.7"E 21°10'36.4"N, 105°45'39.9"E

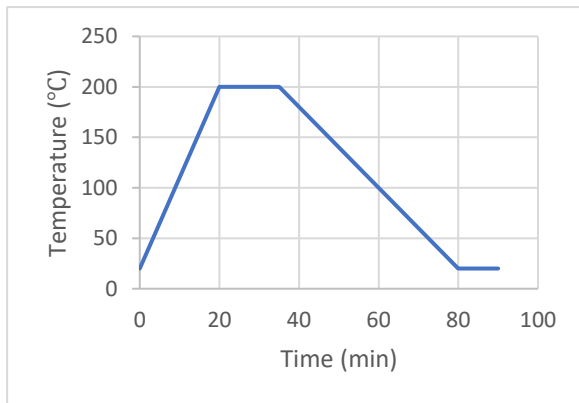


Fig. 1. Heating program of Mars 6 Microwave Digestion for pakchoi's edible samples

LDA is a predictive discriminant analysis technique - an extension of Fisher's linear discriminant [7]. LDA is similar to analysis of variance (ANOVA) and regression analysis in that the goal is to describe a dependent variable by a linear combination of components or other measures [8]. LDA is also similar to PCA in that it is a multivariate analysis method with the goal of reducing dimensionality by constructing principal components from the original combination of variables. However, while PCA aims to maximize variance in principal components, LDA aims to maximize separation of groups [9]. Therefore, in multivariate statistical analysis, LDA serves as a classification method.

3. Results and Discussion

3.1. PCA on Metal Content in the Pakchoi's Edible Parts

After determining the metal content in the edible parts of pakchoi using the ICP-MS method, the results obtained include the average metal content after 3 repetitions of the edible parts of pakchoi from 3 areas, including Chuong My, Dan Phuong, and Me Linh (3 sampling locations per area with specific locations listed in Table 1). The results are used to run the PCA method. The results of the PCA method were obtained as charts in Fig. 2a, Fig. 2b, and Fig. 2c.

As can be seen in the scree plot chart (Fig. 2a) and the observations chart (Fig. 2b), the first 2 functions account for 51.09% of the total sample variation, with 35.87% of Function 1 and 15.22% of Function 2. Sample groups together with their locations can be seen on the observations chart (Fig. 2b) and biplot chart (Fig. 2c) of the PCA.

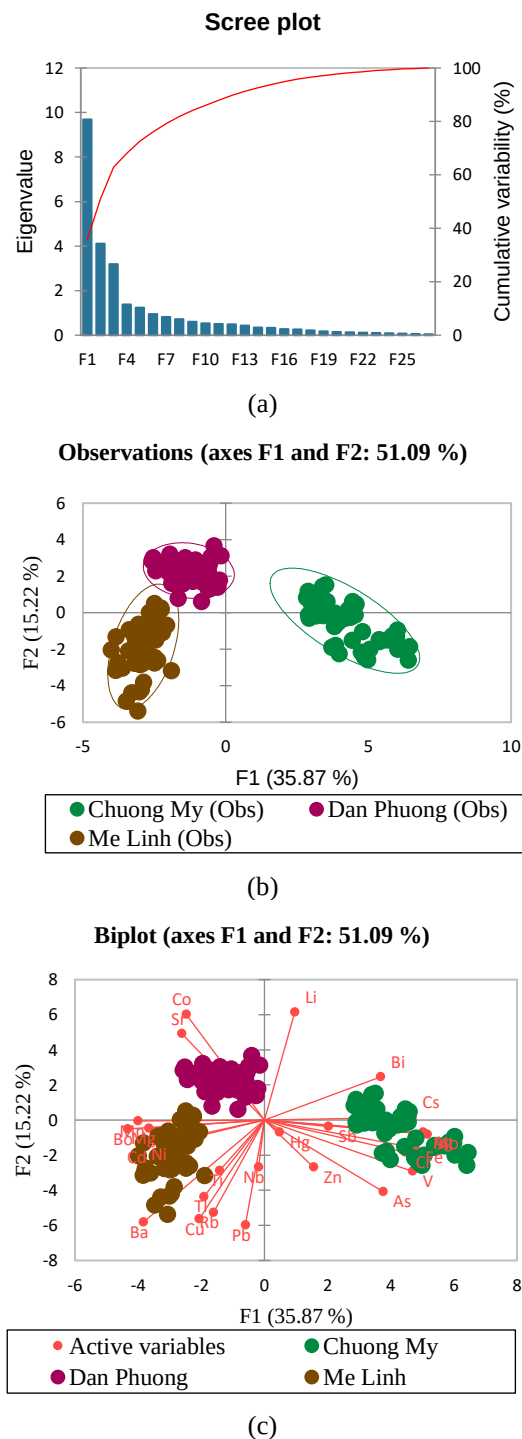


Fig. 2. PCA results on metal content obtained in the edible parts of pakchoi from Chuong My, Dan Phuong, and Me Linh areas from September 2021 to April 2024: Scree plot chart of PCA (a), observations chart of PCA (b), and biplot chart of PCA (c)

The biplot chart shows more clearly the difference in metal content in the edible parts of pakchoi according to each growing district. Chuong My is characterized by the following metals: Mo, Zn, As, V, Al, Fe, Cs, Sb; Dan Phuong is characterized by the metals Li, Co, Sr, Tl; and Me Linh with metals: Ni, Ti, Ba, Cu, Rb, Nb, and Pb. Thus, by using PCA to analyze the metal content data of the edible parts of pakchoi from the three areas, the new plane provided by Function 1 and Function 2 only contained 51.09% of the total data set. In addition, the samples in the three pakchoi growing areas were not clearly differentiated with the confidence ellipses (confidence interval of 95%) of the Dan Phuong and Me Linh areas overlapping. This leads to the fact that using PCA alone is not enough to help differentiate the pakchoi growing locations. Therefore, we used another method, LDA, to help distinguish the areas.

3.2. LDA on Metal Content in the Pakchoi's Edible Parts

The metal content results of the edible parts of pakchoi over 4 years were used for LDA. 27 separate samples from 3 pakchoi growing areas were used to validate the LDA model. The LDA results obtained are shown in Fig. 3.

According to the results obtained from LDA, Function 1 accounts for 62.14% of the data variance, and Function 2 accounts for 37.88% of the data (Table 2, Fig. 3a). Standardized canonical discriminant function coefficients of LDA are shown in Table 3. As followed, the F1 values of As, Co, Cs, and Li are higher than most metals in the F1 column. Thus, the above metal variables are mainly associated with Function 1. Similarly, for the case of Function 2, the corresponding associated metal variable is Ba.

Table 2. Eigenvalues of LDA

	F1	F2
Eigenvalue	48.491	26.346
Discrimination (%)	64.795	35.205
Cumulative (%)	64.795	100.000

The LDA model was built using the results of metal concentrations in the edible parts of pakchoi from samples collected from 2021 to 2024 as training samples, combined with 27 separate test samples from the respective pakchoi growing areas. The results are shown in Fig. 3b and Fig. 3c. From Fig. 3b, different pakchoi growing areas are separated from each other and distinguished according to their relationship with As, Co, Cs, Li, and Ba. Accordingly, Ba is a characteristic metal of the Me Linh area. Co and Li are characteristic metals of the Dan Phuong area. As for the case of Chuong My, the corresponding

characteristic metals are As and Cs. Thus, Function 1 distinguishes well the Chuong My area from the other two areas. Meanwhile, Function 2 distinguishes well the Dan Phuong and Me Linh areas.

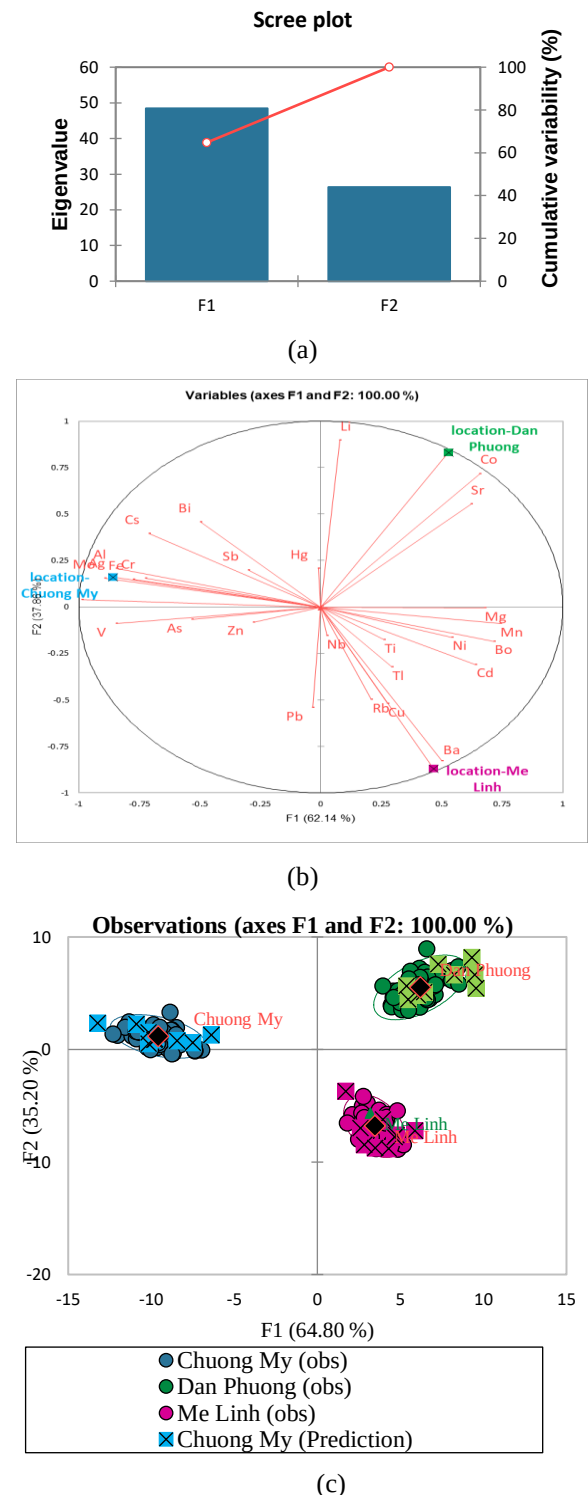


Fig. 3. LDA results on metal content obtained in the edible parts of pakchoi from Chuong My, Dan Phuong, and Me Linh areas from September 2021 to April 2024: Scree plot chart of LDA (a), Variables chart of LDA (b), and observations chart of LDA (c)

Table 3: Standardized canonical discriminant function coefficients of LDA

	F1	F2
Ag	-0.261	-0.030
Al	-0.276	-0.017
As	0.598	-0.106
Ba	0.110	-0.684
Bi	-0.127	0.000
Bo	0.364	-0.111
Cd	0.161	-0.020
Co	0.602	0.628
Cr	0.078	0.178
Cs	-0.804	0.471
Cu	0.147	-0.111
Fe	0.031	0.084
Hg	-0.208	0.122
Li	0.614	0.467
Mg	0.519	0.255
Mn	0.402	-0.129
Mo	-0.478	0.222
Nb	0.231	-0.064
Ni	0.024	-0.274
Pb	-0.313	-0.236
Rb	0.383	0.047
Sb	-0.290	-0.016
Sr	0.125	0.077
Ti	0.232	-0.301
Tl	0.439	0.179
V	-0.583	-0.171
Zn	-0.141	-0.001

Arsenic is mostly found in nature in the form of arsenite and arsenate. They can cause various types of cell damage due to their carcinogenic effects [10].

Potential sources of arsenic contamination are thought to include mining and agricultural activities using chemical pesticides, insecticides, herbicides, desiccants, and leaf killers [11]. The concentration of arsenic in edible parts of plants depends on the prevalence of arsenic in the soil and the plant's ability to accumulate and transport it [12]. This could potentially result in the Chuong My area's soil or irrigation water containing more As than the other two areas. However, the average As content in the edible parts of pakchoi in the growing areas all resulted in results within the permissible limits for As content in dried vegetables (1 mg.Kg^{-1}) according to the National technical regulation on the limits of heavy metals contamination in food of Vietnam. There is no known role for Cs in plant nutrition, but excessive Cs can be toxic to plants [13]. However, the Cs content in the edible parts of pakchoi in all three areas was quite small. This leads to the possibility that Cs is not toxic to the plant and has a great impact on the morphology and characteristics of the plant. The As and Cs content in the edible parts of pakchoi in the Chuong My area is higher than that in other areas. This not only helps to distinguish the pakchoi growing areas clearly but also shows the typical accumulation of metals in the environment, such as soil, irrigation water, etc., in the Chuong My area.

Barium is present in many places in the soil, and all plants contain small amounts, usually at levels of about 4 to 50 $\text{mg.Kg}^{-1}\text{DW}$ [14]. Ba is a component of geothermal brine and therefore has the potential to spill into the environment if accidentally leaked. Following table 4, the Ba content of pakchoi in the two areas of Chuong My and Dan Phuong are $19.06 \pm 2.89 \text{ Kg}^{-1}\text{DW}$ and $38.13 \pm 5.16 \text{ mg.Kg}^{-1}\text{DW}$. The samples from Me Linh have a much higher Ba content than the other two areas ($343.15 \pm 49.79 \text{ mg.Kg}^{-1}\text{DW}$), indicating the influence of the characteristic Ba available in the soil. Ba has been shown to be toxic to plants when accumulated at high concentrations due to limited CO_2 assimilation due to limited photosynthetic activity [15]. However, research results on *Brassica juncea*, a plant of the same Brassica family, show that this plant can tolerate high concentrations of Ba in the soil and even has a positive impact on the plant's reproductive growth [16]. Among the total edible samples of pakchoi, samples from Me Linh origin had significantly higher Ba content, more than 10 times higher than the Ba content from the other two regions. This also leads to the possibility that the soil in the Me Linh area has a higher Ba content than other regions.

Cobalt is believed to be a beneficial metal for plants, but its benefits to plants are still unclear [17]. However, the Co content in the edible parts of pakchoi in 3 cultivated areas is quite small. This leads to the Co content being used to clarify the characteristics of pakchoi in the areas but is not enough to prove that it can influence the growth and morphology of the plant.

Table 4: Average characteristic metal content by areas (mg.Kg⁻¹DW)

	Chuong My	Dan Phuong	Me Linh
As	0.82 ± 0.20	0.45 ± 0.17	0.54 ± 0.06
Ba	19.06 ± 2.89	38.13 ± 5.16	343.15 ± 49.79
Co	0.12 ± 0.01	0.95 ± 0.08	0.28 ± 0.05
Cs	0.43 ± 0.11	0.22 ± 0.04	0.10 ± 0.01
Li	0.52 ± 0.20	0.99 ± 0.22	0.04 ± 0.01

Lithium in soil can come from many kinds of sources: Industrial wastewater and nearby water resources near to Li-related industry [18], disposal of Li batteries [19], etc. Therefore, Li mainly exists in the surface soil layer but less in the deeper soil layer. The uniform Li content in topsoil samples in different regions may reflect consistency in cultivation and irrigation.

Twenty-seven separate edible samples of pakchoi taken from 3 areas (9 samples per area) were used to obtain the metal content results as prediction samples (test samples). The results of running the LDA model are shown in Fig. 3c and Table 5. Accordingly, the samples were predicted accurately and were shown as square boxes with a black crossline in Fig. 3c. Thus, the quality of the built model is appropriate.

Table 5 shows the results of the model for the 27 prepared prediction samples. The model constructed by LDA showed the ability to accurately recognize 27/27 prepared prediction samples. Accordingly, the samples were predicted accurately and were shown as square boxes with a black crossline in Fig. 3c. Thus, the quality of the built model is appropriate.

4. Conclusion

Chuong My, Me Linh, and Dan Phuong are 3 areas specializing in growing pakchoi serving Hanoi's market during the period from September 2021 to April 2024. Metal concentrations in the edible parts of pakchoi from these areas were collected and determined by ICP-MS. The results obtained were used to determine the geographical origin of pakchoi. The PCA method used in the analysis of metal content in the edible parts of pakchoi showed the ability to distinguish the Chuong My area from the Dan Phuong and Me Linh areas. However, the two areas, Dan Phuong and Me Linh, showed overlap with each other on the plane constructed from the two principal components of PCA. The LDA method was used and was successful in distinguishing the three geographical areas by using As, Ba, Co, Cs, and Li content. The model's predictive accuracy for 27 prepared samples

was 100%. This shows the potential of using metal content results in building a model to authenticate the geographical origin of pakchoi in particular and other vegetables in general.

Table 5: Results for the prediction sample:

Predicted class	Pr (Chuong My)	Pr (Dan Phuong)	Pr (Me Linh)
Chuong My	1	0	0
Chuong My	1	0	0
Chuong My	1	0	0
Dan Phuong	0	1	0
Dan Phuong	0	1	0
Dan Phuong	0	1	0
Me Linh	0	0	1
Me Linh	0	0	1
Me Linh	0	0	1
Chuong My	1	0	0
Chuong My	1	0	0
Chuong My	1	0	0
Chuong My	1	0	0
Chuong My	1	0	0
Chuong My	1	0	0
Dan Phuong	0	1	0
Dan Phuong	0	1	0
Dan Phuong	0	1	0
Dan Phuong	0	1	0
Dan Phuong	0	1	0
Me Linh	0	0	1
Me Linh	0	0	1
Me Linh	0	0	1
Me Linh	0	0	1
Me Linh	0	0	1
Me Linh	0	0	1

Acknowledgments

This research was funded by the Vietnam Academy of Science and Technology under grant number NCXS 01.02/23-25.

References

- [1] C.W. Li, Chinese cabbages of China, China: Agricultural Publishing Company, 1984.
- [2] A. Kalisz, A. Kostrzewa, J. Sekara, and A. Grabowska Stanislaw, Yield and nutritional quality of several non-heading Chinese cabbage (*Brassica rapa* var. *chinensis*) cultivars with different growing period and its modelling, *Korean Journal of Horticultural Science and Technology*, vol. 30, iss. 6, pp. 650-656, 2012. <https://doi.org/10.7235/hort.2012.12108>
- [3] P. Hanson, R. Yang, L. Chang, L. Ledesma, and D. Ledesma, Contents of carotenoids, ascorbic acid, minerals and total glucosinolates in leafy brassica pakchoi (*Brassica rapa* L. *chinensis*) as affected by season and variety, *Journal of the Science of Food and Agriculture*, vol. 89, iss. 5, pp. 906-914, Mar. 2009. <https://doi.org/10.1002/jsfa.3533>
- [4] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 2, iss. 11, pp. 559-572, Jun. 2010. <https://doi.org/10.1080/14786440109462720>
- [5] H. Hotelling, Analysis of a complex of statistical variables into principal components, *Journal of Educational Psychology*, vol. 24, pp. 498-520, 1993. <https://doi.org/10.1037/h0070888>
- [6] D. Granato, J. S. Santos, G. B. Escher, B. L. Ferreira, and R. M. Maggio, Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in foods: a critical perspective, *Trends in Food Science & Technology*, vol. 72, pp. 83-90, Feb. 2018. <https://doi.org/10.1016/j.tifs.2017.12.006>
- [7] R. A. Fisher, The use of multiple measurements in taxonomic problems, *Annals of Eugenics*, vol. 7, iss. 2, pp. 179-188, Sep. 1936. <https://doi.org/10.1111/j.1469-1809.1936.tb02137.x>
- [8] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, Wiley Interscience, 2004, 552 pp.
- [9] A. M. Martinez, PCA versus LDA in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, iss. 2, pp. 228-233, Feb. 2001. <https://doi.org/10.1109/34.908974>
- [10] N. Bolan, S. Mahimairaja, A. Kunhikrishnan, B. Seshadri, and R. Thangarajan, Bioavailability and ecotoxicity of arsenic species in solution culture and soil system: implications to remediation, *Environmental Science and Pollution Research*, vol. 22, iss. 12, pp. 8866-8875, 2015. <https://doi.org/10.1007/s11356-013-1827-2>
- [11] S. Bhattacharya, P. Sharma, S. Mitra, I. Mallick, and A. Ghosh, Arsenic uptake and bioaccumulation in plants: A review on remediation and socio-economic perspective in Southeast Asia, *Environmental Nanotechnology, Monitoring & Management*, vol. 15, May. 2021. <https://doi.org/10.1016/j.enmm.2021.100430>
- [12] R. Q. Huang, S. F. Gao, W. L. Wang, S. Staunton, and G. Wang, Soil arsenic availability and the transfer of soil arsenic to crops in suburban areas in Fujian Province, southeast China, *Science of the Total Environment*, vol. 368, iss 2-3, pp. 531-41, Sep. 2006. <https://doi.org/10.1016/j.scitotenv.2006.03.013>
- [13] P. J. White and M. R. Broadley, Mechanisms of caesium uptake by plants, *New Phytologist*, pp. 241-256, Jul. 2008. <https://doi.org/10.1046/j.1469-8137.2000.00704.x>
- [14] F. M. Chaudhry, A. Wallace, and R. T. Mueller, Barium toxicity in plants, *Communications in Soil Science and Plant Analysis*, vol. 8, iss. 9, Nov. 2008, pp. 795-797. <https://doi.org/10.1080/00103627709366776>
- [15] R. Suwa, K. Jayachandran, N. T. Nguyen, A. Boulouar, K. Fujita, and H. Saneoka, Barium toxicity effects in soybean plants, *Archives of Environmental Contamination and Toxicology*, vol. 55, pp. 397-403, Feb. 2008. <https://doi.org/10.1007/s00244-008-9132-7>
- [16] H. Bouslimi, R. Ferreira, N. Dridi, P. Brito, S. Martins-Dias, I. Caçador, and N. Sleimi, Effects of Barium Stress in *Brassica juncea* and *Cakile maritima*: The indicator role of some antioxidant enzymes and secondary metabolites, *Phyton*, vol. 90, no. 1, pp. 145-158, 2021. <https://doi.org/10.32604/phyton.2020.011752>
- [17] S. Khalid, M. Shahid, R. Naz, N. K. Niazi, A. A. Al-Kahtani, Z. A. Alothman, R. Fatima, and B. Murtaza, Cobalt in soil-plant-human system: a data-analysis review, *Environmental and Earth Sciences*, Sep. 2023. <https://doi.org/10.20944/preprints202309.1436.v1>
- [18] S. L. Hull, U. V. Oty, and W. M. Mayes, Rapid recovery of benthic invertebrates downstream of hyperalkaline steel slag discharges, *Hydrobiologia*, vol. 736, pp. 83-97, May. 2014. <https://doi.org/10.1007/s10750-014-1894-5>
- [19] S. Al-Thyabat, T. Nakamura, E. Shibata, and A. Iizuka, Adaptation of minerals processing operations for lithium-ion (LiBs) and nickel metal hydride (NiMH) batteries recycling: critical review, *Minerals Engineering*, vol. 45, pp. 4-17, May. 2013. <https://doi.org/10.1016/j.mineng.2012.12.005>