Journal of Nepal Agricultural Research Council
Vol.10:37-49, May 2024
ISSN: 2392-4535 (Print), 2392-4543 (Online)
DOI: https://doi.org/10.3126/jnarc.v10i1.73265

Principal Component Analysis of Nutrient Content of Root and Tuber Crops of Nepal

Pravin Ojha[1@], Roman Karki[1], Utshah Manandhar[1] and Tek Prasad Gotame[2]

[1] National Food Research Centre, Nepal Agricultural Research Council, Khumaltar, Lalitpur, @ projha84@gmail.com; https://orcid.org/0000-0002-1103-5984
[2]Nepal Agricultural Research Council, Planning Division, Singhdurbar, Plaza, Kathmandu; RK: rmn_karki@narc.gov.np; UM: utshah32@gmail.com; TPG:gotame@gmail.com

Received: June 07, 2024, Revised: Nov. 21, 2024, Accepted: Published: December, 2024.

The authors declare that there is no conflict of interest.

ABSTRACT

The research aimed to cluster the nutrient profile of six root and tuber crops including chayote root. Potato cv. Janakdev and Khumal Upahar (from Lalitpur) and cv. MS-42-3 (from Nuwakot); sweet potato cv. Suntale 1 and landraces Bensishar White and Kimichaur Seto (from Lalitpur); landrace of elephant foot yam and taro (from Pokhara); landrace of chayote root (from Dhankuta) and cassava (from Udaypur) were collected and Principal Component Analysis (PCA) was done based on their nutrient compositions. The evaluated nutrients included moisture, protein, fat, ash, crude fiber, carbohydrates, iron, calcium, phosphorus, sodium, potassium, zinc, vitamin C, and reducing sugar. PCA revealed that the first four principal components accounted for 85.7% of the total variation. Ash, calcium, and reducing sugar positively correlated with PC1 and PC2, while vitamin C, zinc, and protein were negatively associated with these components. There was a negative correlation of sodium concentration with potassium, phosphorus, and iron concentrations. Additionally, there was a negative correlation between carbohydrates and fat with the moisture content of the potatoes and tubers. The collected potatoes and tubers were classified into four clusters based on similarity and distance levels. The potatoes and sweet potatoes were placed in two distinct clusters. Elephant foot yam and chayote root were categorized in one cluster, distinguishing them from cassava. These findings highlighted the diverse nutritional profiles of the examined root and tuber crops, offering valuable insights for further research. Furthermore, the results can pave the way for commercial exploration of these potatoes and tubers based on their nutritional profiles.

Keywords: Cluster, Potato and tubers, Principal Component Analysis, Vitamin C, Zinc

सारांश

दुई वटा आलुका जातहरु खुमल उपहार र जनकदेव (ललितपुर संकलन) र एउटा जात एम एस ४३-३ (नुवाकोट संकलन), सखरखण्डको तीन जातहरु सुन्तले सखरखण्ड १, बेंसीसहर सेतो र किमिचौर सेतो (ललितपुर संकलन), स्कुसको जरा (धनकुटा संकलन), पोखराको स्थानीय पिडालु र ओल (घर तरुल) (पोखरा संकलन) र सिमल तरुल (उदयपुर संकलन) को पौष्टिक तत्वहरुको बनौटको आधारमा क्लस्टर गरिएको थियो । यी जरा तथा आलु बालीहरुमा पानीको मात्रा, प्रोटिन, बोसो, खरानी, कच्चा रेसा पदार्थ, कार्बोहाईड्रेट, फलाम, क्याल्सियम, फोस्फोरस, सोडियम, पोटास, जिंक, भिटामिन सी र कम चिनीको मात्रा ल्याबमा बिश्लेषण गरिएको थियो । प्रिन्सिपल कम्पोनेन्ट बिश्लेषणको नतिजा अनुसार चौथोसम्मको प्रिन्सिपल कम्पोनेन्टले ८५.७% जम्मा भिन्नता देखायो । जम्मा खरानीको भाग, क्याल्सियम र कम चिनी प्रिन्सिपल कम्पोनेन्ट पहिलो र दोश्रोसंग धनात्मक सम्बन्ध र भिटामिन सी, जिंक र प्रोटिन संग ऋणात्मक सम्बन्ध रहेको पाइयो । फलाम, पोटासियम र फोस्फोरससंग सोडियमको सम्बन्ध ऋणात्मक सम्बन्ध भएको पाइयो । त्यस्तै कार्बोहाईड्रेट र बोसोको सम्बन्ध पानीको मात्रसंग ऋणात्मक सम्बन्ध रहेको पाइयो । नजिक तथा टाढाको सम्बन्धको आधारमा संकलित आलु तथा कन्दमुल बालीहरु ४ वटा क्लस्टरमा छुटाईयो । त्यस्तै स्कुसको जरा र घर तरुल (ओल) लाई सिमलतरुल भन्दा फरक प्रिन्सिपल कम्पोनेन्टमा छुट्टिएको पाईयो । यस अनुसन्धानले आलु तथा कन्दमुल बालीहरुमा पौस्टिक तत्वहरुको बिबिधतालाई उजागरण गरेको छ ।

INTRODUCTION

Root and tuber (R&T) crops are second in their importance to cereals as a global source of carbohydrates and particularly due to high production potential in the areas unsuitable for cereal crops Khatri et al 2017). Root and tuber crops are vital components of agricultural systems worldwide, playing a crucial role in global food security and socio-economic development. In Nepal, these underground crops have gained increasing attention due to their importance as staple foods, sources of income, and contributors to agricultural biodiversity.

Taro (*Colocasia esculenta* L. Schott), yam (*Dioscorea alata* L.), cassava (*Manihot esculenta* Crantz), Ele phant foot yam (*Amorphophallus paeoniifolius* (Dennst.) Nicolson), potato (*Solanum tuberosum* L.) and s weet potato (*Ipomoea batatus* L.) are the major cultivated R&T crops in Nepal (Khatri et al 2017). The country's diverse agro-climatic conditions, ranging from the lowland Terai region to the high Himalayas, provide a unique environment for cultivating a wide variety of R&T crops. Potatoes and other tuber crops, such as sweet potatoes, yams, taro, and cassava, have been integral to Nepal's traditional farming systems for centuries. The adoption of these crops has been shaped by factors such as climate, soil types, and cultural practices, resulting in a rich diversity of tubers across different regions. Similarly, chayote roots are used in a variety of curries along with the fruit and shoots of the chayote plant (*Sedium edule* Jacq) in a hilly region of Nepal.

The tuberous root of chayote is used as an alternative to potato and is tastier than other root and tuber crops (Joshi et al 2020). In recent years, Nepal's potato production has seen substantially increased, significantly contributing to the country's agricultural output. The cultivation practices surrounding potatoes and tubers in Nepal are closely linked to the socio-economic fabric of rural communities. Smallholder farmers, who represent the majority of the agricultural workforce, often rely on tuber crops for sustenance and income. The adaptability of these crops to varying altitudes and climatic conditions further enhances their importance, helping farmers mitigate risks associated with climate variability. Potatoes, sweet potatoes, yams, and other locally adapted tubers serve not only as staple foods but also as valuable sources of essential nutrients. Understanding the nutritional content of these crops is crucial for addressing malnutrition and promoting public health initiatives in the country.

Principal Component Analysis (PCA) is a multivariate analysis technique used to reduce the dimensionality of data. This reduction facilitates a clearer visual representation of the data in fewer dimensions and helps establish relationships between variables (Grane and Jach 2014). PCA represents data along axes where each axis corresponds to a principal component (PC). The first principal component (PC1) accounts for more variability in the data than the second (PC2), which in turn explains more variability than the third (PC3), and so forth (Granato et al 2018). PCA can be performed using either covariance or correlation, although PCA using correlation is typically preferred when dealing with variables on different scales. Correlation-based PCA focuses primarily on the relationships between variables, while variance-based PCA emphasizes the variance of those variables (Mishra et al 2017). Another important concept in PCA is eigenvalues, which represent the variance captured by each principal component. Higher eigenvalues indicate that a corresponding principal component accounts for greater variability in the data. Eigenvalues also aid in determining the optimal number of principal components, thereby assisting in dimensionality reduction (Johnson 1998). Additionally, clustering techniques in statistics group data objects based on their similarities and differences. K-means clustering is particularly advantageous for managing large datasets while providing high-quality clusters (Kaushik and Mathur, 2014).

In the field of food science, tools like Principal Component Analysis (PCA) and clustering help uncover hidden connections among various aspects of food. They enable us to understand how different elements such as dietary patterns, chemical compositions, and sensory qualities are interrelated. The use of PCA in our research simplifies data visualization, making it much easier to interpret complex datasets. Essentially, this tool untangles the web of information, allowing us to see how different nutrients are interconnected and how they vary across different types of food. However, in Nepal, there has been a gap in research where these statistical tools, particularly multivariate analysis, have not been adequately utilized to categorize foods based on their nutritional content. By employing PCA, we can condense large amounts of nutrient data into a few dimensions, capturing the most significant variations and relationships among different nutrients. Additionally, clustering tools allow us to group foods based on their nutritional profiles.

This approach has practical implications across various sectors: it aids industries in sourcing specific materials, assists government agencies in setting food policies and nutritional guidelines, and helps the public make informed choices about their diets based on diverse nutritional values. One interesting application of clustering is in the creation of low-cost, nutrient-rich diets, where foods are grouped according to their nutritional benefits. This not only contributes to improving public health but also has economic and policy implications, influencing how we approach food production and consumption.

MATERIALS AND METHODS

Two potato samples (cv. Janakdev and cv. Khumal Upahar) and three sweet potato samples (cv. Suntale 1, Besisahar White, and Kimichaur Seto) were collected from Khumaltar, Lalitpur. Additionally, a potato sample (cv. MS-42-3) was obtained from Nuwakot, while elephant foot yam and taro were sourced from Pokhara. The cassava and chayote roots were collected from Udaypur and Pakhribas, respectively. A minimum of two kilograms of each sample was obtained from one plot. To achieve homogenization, samples from three plots were carefully combined to create a uniform blend that incorporated the unique characteristics of each plot. The methodology involved triplicate analyses, with each analysis repeated three times for every sample.

Chemical analysis

The moisture content was determined using the hot-air oven method, which involved drying the sample at 105°C until a constant weight was achieved, by AOAC method number 930.15 (AOAC 2005). Protein content was calculated based on nitrogen content, which was measured using the Kjeldahl method outlined in AOAC method number 920.152 (AOAC 2005). The fat content of potatoes and tubers was determined by solvent extraction with a Soxhlet apparatus, using petroleum ether as the solvent, following AOAC method number 991.36 (AOAC 2005). To ascertain the crude fiber content of potatoes and tubers, we employed AOAC method number 934.01. The total ash content was determined using AOAC method number 945.46 (AOAC 2005).

The total carbohydrate content was calculated by the difference method using the formula:
Carbohydrate (% dry basis) = 100 − (crude protein + total ash + crude fiber + crude fat).
The calcium, iron, and phosphorus contents of potatoes and tubers was assessed according to AOAC (2005) methods, utilizing a Cary UV-Vis spectrophotometer (Agilent, USA). Sodium, potassium, and zinc levels in the food matrix were quantified using Flame Atomic Absorption Spectroscopy (Agilent AA 240FS, Germany), following the AOAC (2005) protocol. Vitamin C was analyzed using a titrimetric method as described in AOAC (2005). Reducing sugar content was also analyzed using Flame Atomic Absorption Spectroscopy (Agilent AA 240FS, Germany), as per AOAC (2005) guidelines.

Principal Component Analysis (PCA) and K-means Clustering

The PCA using correlations and K-means clustering was conducted with JMP Pro 17 software. The results are illustrated based on the first two principal components, which are showcased in the summary plot, scree plot, loading plot, and contribution plot. Additionally, the ten products were grouped using K-means clustering. Only four principal components were chosen as they characterize 80% of the variance.

RESULTS

The nutritional profile of potatoes and tubers is comprehensively summarized in **Table 1**. To explore the nutritional characteristics of these foods, a Principal Component Analysis (PCA) was performed, with the results interpreted through the lens of Eigenvalues. The PCA revealed that the first principal component accounted for 39.8% of the total variation within the dataset, indicating that this component captures a substantial portion of the underlying patterns in the nutritional data. Following this, the second principal component contributed 20.7%, while the third component accounted for 15.5%, and the fourth component contributed 9.7%. The selection of these components was guided by the Kaiser criterion, which stipulates that only those components with Eigenvalues greater than 1 should be retained for analysis. This methodological approach ensures that only the most significant factors influencing nutritional variation are considered. The findings from the PCA are visually represented in **Figure 1.a** and **Figure 1.b**, which illustrate the relationships between the principal components and their contributions to data variance. Additionally, a more detailed breakdown of these components and their respective contributions is provided in **Table 2**, offering further insights into the complex interplay of nutritional factors in potatoes and tubers.

Table 1. Nutritional value of root and tuber crops including chayote roots

Sample	Potato_ Janakdev	Potato_Khumal Upahar	Potato_ MS-42-3	Sweet potato_ Suntale 1	Sweet potato_ Besisahar white	Sweet potato_ Kimichaur seto	Elephant Yam_ Nawalparasi	Chayote root local	Taro	Cassava
Moisture (g/100 g)	82.33±1.28	80.57±1.19	77.98±2.55	70.45±0.76	69.25±0.83	70.11±0.22	76.73±4.41	72.21±3.21	73.86±1.85	64.27±0.24
Protein (g/100 g)	1.33±0.25	1.59±0.13	1.28±0.19	1.21±0.14	0.89±0.19	1.33±0.03	0.78±0.15	0.62±0.03	1.47±0.18	1.56±0.07
Fat (g/100 g)	0.08±0.01	0.08±0.01	0.35±0.05	0.38±0.01	0.29±0.02	0.23±0.02	0.10±0.02	0.14±0.02	0.28±0.05	0.35±0.03
Ash (g/100 g)	0.73±0.05	0.73±0.05	0.70±0.07	1.08±0.06	0.95±0.15	1.30±0.15	1.15±0.18	1.26±0.28	1.49±0.32	0.55±0.03
Crude fibre (g/100 g)	0.59±0.05	0.55±0.08	0.78±0.04	1.29±0.04	0.88±0.07	0.97±0.05	1.02±0.04	0.42±0.18	1.04±0.2	0.66±0.04
Carbohydrate (g/100 g)	14.95±1.43	16.48±1.12	18.91±2.53	25.59±0.61	27.74±0.87	26.05±0.08	20.22±4.40	25.35±3.47	21.87±1.66	32.61±0.28
Iron (mg/100 g)	0.54±0.04	0.50±0.03	0.62±0.06	0.26±0.01	0.30±0.03	0.41±0.02	0.64±0.04	1.12±0.07	0.87±0.08	0.49±0.03
Calcium (mg/100 g)	9.95±0.4	9.37±0.21	9.21±0.06	31.40±1.4	24.89±1.71	25.97±2.46	42.87±1.89	34.71±4.98	53.56±2.03	20.66±0.97
Phosphorous (mg/100 g)	47.18±2.45	45.35±2.04	55.19±3.01	35.43±4.02	35.40±2.55	36.87±1.84	33.13±1.98	62.01±1.96	52.16±2.28	43.11±3.23
Sodium (mg/100 g)	3.95±0.09	4.01±0.04	3.87±0.16	24.46±0.38	25.75±0.08	26.52±0.06	15.61±0.08	12.63±0.27	12.29±0.14	10.15±0.3
Potassium (mg/100 g)	461.14±9.96	453.83±3.42	460.25±6.21	345.75±8.64	346.26±5.58	347.99±3.94	494.91±4.25	379.62±8.04	435.31±3.09	353.46±9.57
Zinc (mg/100 g)	0.29±0.04	0.36±0.03	0.33±0.03	0.14±0.02	0.16±0.02	0.16±0.02	0.28±0.07	0.15±0.04	0.19±0.03	0.38±0.02
Vitamin C (mg/100 g)	25.64±1.24	26.30±2.20	29.94±1.99	22.63±3.57	10.56±0.76	10.23±1.95	16.49±1.14	6.92±0.47	13.77±1.85	10.24±0.41
Reducing sugar (g/100 g)	0.41±0.04	0.36±0.04	0.46±0.05	0.89±0.03	0.87±0.03	0.85±0.03	0.47±0.03	0.43±0.01	1.37±0.1	0.32±0.04

Note: The values presented are the mean of triplicate analysis (n=3)

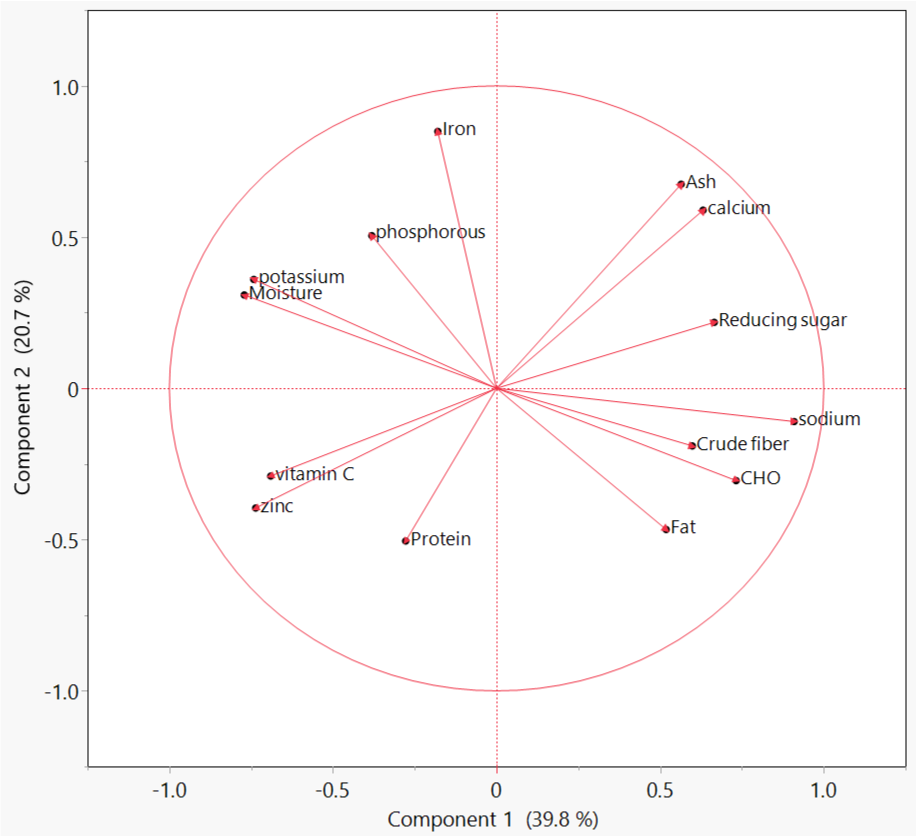

Figure 1a. PCA correlation circle of the nutritional profile of root and tuber crops
CHO=Carbohydrate

Table 2. Eigenvalue of principal components

Number	Eigenvalue	Per cent
1	5.567106	39.765
2	2.903930	20.742
3	2.171972	15.514
4	1.362677	9.733
5	0.794184	5.673
6	0.619231	4.423
7	0.202317	1.445
8	0.140645	1.005
9	0.092039	0.657
10	0.067038	0.479
11	0.041110	0.294
12	0.028249	0.202
13	0.009503	0.068

The four principal components identified in the analysis collectively account for a substantial 85.7% of the total variance within the dataset. This significant proportion underscores the effectiveness of the PCA in capturing the underlying structure of the nutritional data. The analysis revealed a positive association between several variables, specifically ash content, calcium, reducing sugar, sodium, crude fiber, carbohydrates, and fat, with the first principal component. These variables were found to significantly contribute to the variance explained by this component, indicating their importance in defining the nutritional profile of potatoes and tubers.

Similarly, the second principal component exhibited positive correlations with moisture, potassium, phosphorus, iron, ash, calcium, and reducing sugar. Each of these variables played a crucial role in influencing the variance captured by this component. Notably, ash, calcium, and reducing sugar emerged as positively correlated with both principal components, suggesting their dual significance across different aspects of nutritional profiling. Conversely, vitamin C, zinc, and protein displayed negative associations with these components, indicating that fluctuations in their values could substantially impact the dataset's overall characteristics.

Furthermore, the results illustrated that an increase in fat, carbohydrates, crude fiber, and sodium contents in potatoes and tubers was associated with a decrease in moisture, potassium, phosphorus, and iron levels. This inverse relationship highlights the multifaceted interplay among these nutritional components. Similarly, higher levels of ash, calcium, and reducing sugar correlate with lower levels of vitamin C, zinc, and protein. These findings emphasize the complex balance within the nutritional composition of potatoes and tubers and suggest that changes in one parameter can significantly affect others within this system.

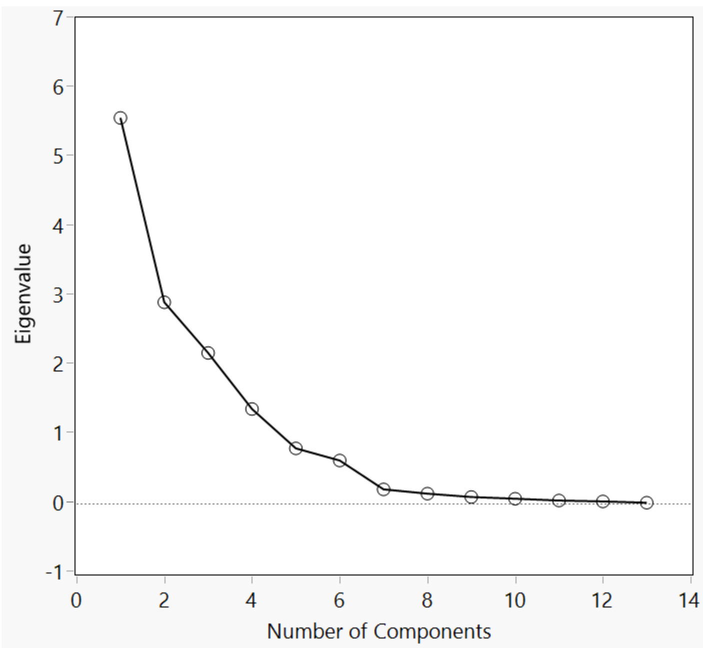

Figure 1b. Scree plot of eigenvalues against the corresponding number of principal components.

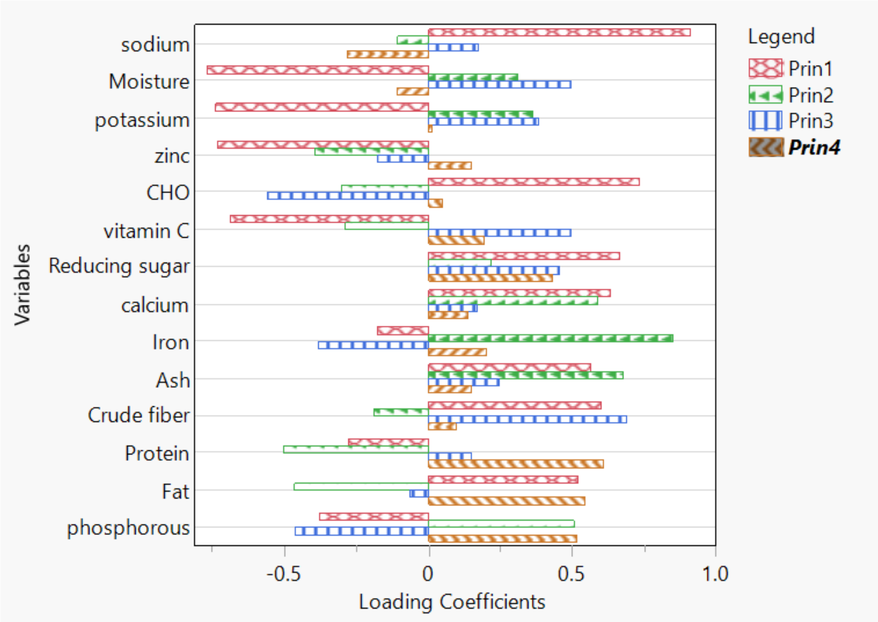

Figure 2. Positive and negative correlation of parameters with 4 principal components
CHO=Carbohydrate

Table 3 and **Figure 2** provide a detailed image of the distances of various parameters about the first four principal components. In this context, a negative value signifies a negative correlation between the parameter and the principal component, while a positive value indicates a positive correlation. The magnitude of these values is particularly important; higher values reflect greater distances from the origin to the lines representing the principal components, suggesting a stronger influence of those parameters on the respective components. Conversely, lower values correspond to reduced distances, indicating a lesser degree of influence. Furthermore, the data analysis revealed a consistent trend in the loading values across the principal components. Specifically, the loading values for most parameters were the highest in Principal Component 1, demonstrating that this component captures the most significant variance within the dataset. As one moves to subsequent components, there is a gradual decrease in loading values, indicating that each following principal component accounts for progressively less variance. This pattern underscores the hierarchical structure of the data and emphasizes the importance of the first principal component in capturing key relationships among the variables analyzed.

Table 3. Loading matrix chart

Parameters	PC1	PC2	PC3	PC4
Moisture (g/100 g)	-0.77	0.31	0.49	-0.11
Protein (g/100 g)	-0.28	-0.50	0.15	0.61
Fat (g/100 g)	0.52	-0.47	-0.07	0.54
Ash (g/100 g)	0.56	0.67	0.24	0.15
Crude fibre (g/100 g)	0.60	-0.19	0.69	0.09
Carbohydrate (g/100 g)	0.73	-0.31	-0.56	0.05
Iron (mg/100 g)	-0.18	0.85	-0.38	0.20
Calcium (mg/100 g)	0.63	0.59	0.17	0.14
Phosphorous (mg/100 g)	-0.38	0.50	-0.46	0.52
Sodium (mg/100 g)	0.91	-0.11	0.17	-0.28
Potassium (mg/100 g)	-0.74	0.36	0.38	0.01
Zinc (mg/100 g)	-0.74	-0.40	-0.18	0.15
Vitamin C (mg/100 g)	-0.69	-0.29	0.50	0.19
Reducing sugar (g/100 g)	0.67	0.22	0.46	0.43

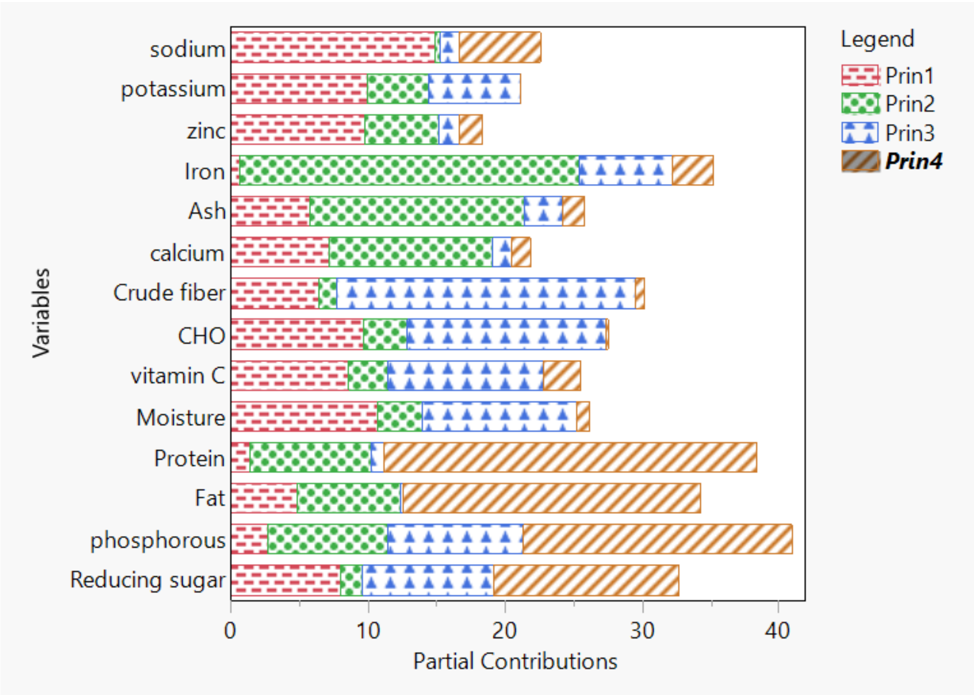

Figure 3. **Variables contributing to the principal components (1 to 4)**
CHO=Carbohydrate

Table 4 and **Figure 3** provide a comprehensive overview of the contributions of various variables to principal components 1 through 4. The cut-off point for significance was established at 6.66%, derived from a calculation that involves multiplying 100 by the ratio of one parameter to the total number of parameters (15). Utilizing this cut-off point, several key parameters emerged as major contributors to each principal component. For Principal Component 1, eight parameters were identified as significant contributors: moisture, carbohydrate, calcium, sodium, potassium, zinc, vitamin C, and reducing sugar. In Principal Component 2, six parameters played a critical role: protein, fat, ash, iron, calcium, and phosphorus. Principal Component 3 revealed another set of eight important parameters: moisture, crude fiber, carbohydrate, iron, phosphorus, potassium, and vitamin C. Finally, Principal Component 4 consisted of four notable contributors: protein, fat, phosphorus, and reducing sugar. This detailed breakdown underscores the importance of each variable in shaping the respective principal components, illustrating how different factors collectively influence the overall data structure and interpretation.

Figure 4 illustrates the clustering of the collected potatoes and tubers, comprising a total of 10 samples, categorized into four distinct clusters based on their nutritional profiling. Especially, the potatoes are classified within Cluster 1, indicating a specific grouping based on their nutritional characteristics. In contrast, sweet potatoes are assigned to Cluster 3, highlighting their unique nutritional profile compared to other tubers. The remaining tubers, except cassava, are aggregated in Cluster 4, suggesting that they share similar nutritional attributes that differentiate them from both potatoes and sweet potatoes. Interestingly, cassava is distinctly positioned in Cluster 2, setting it apart from the other tubers and indicating a unique nutritional composition. This clustering analysis provides valuable insights into the nutritional relationships among various potato and tuber samples, emphasizing the diversity within this group and the potential implications for dietary choices and agricultural practices.

Table 4. Contribution of variables to the principal components (1 to 4)

Parameters	PC1	PC2	PC3	PC4
Moisture (g/100 g)	10.68	3.28	11.27	0.86
Protein (g/100 g)	1.39	8.77	0.99	27.19
Fat (g/100 g)	4.83	7.50	0.21	21.71
Ash (g/100 g)	5.72	15.68	2.73	1.64
Crude fibre (g/100 g)	6.42	1.25	21.84	0.64
Carbohydrate (g/100 g)	9.63	3.20	14.50	0.17
Iron (mg/100 g)	0.58	24.82	6.76	2.96
Calcium (mg/100 g)	7.15	11.94	1.32	1.39
Phosphorous (mg/100 g)	2.61	8.77	9.94	19.57
Sodium (mg/100 g)	14.85	0.42	1.35	5.93
Potassium (mg/100 g)	9.89	4.46	6.74	0.01
Zinc (mg/100 g)	9.74	5.39	1.49	1.64
Vitamin C (mg/100 g)	8.56	2.89	11.30	2.76
Reducing sugar (g/100 g)	7.95	1.64	9.54	13.53

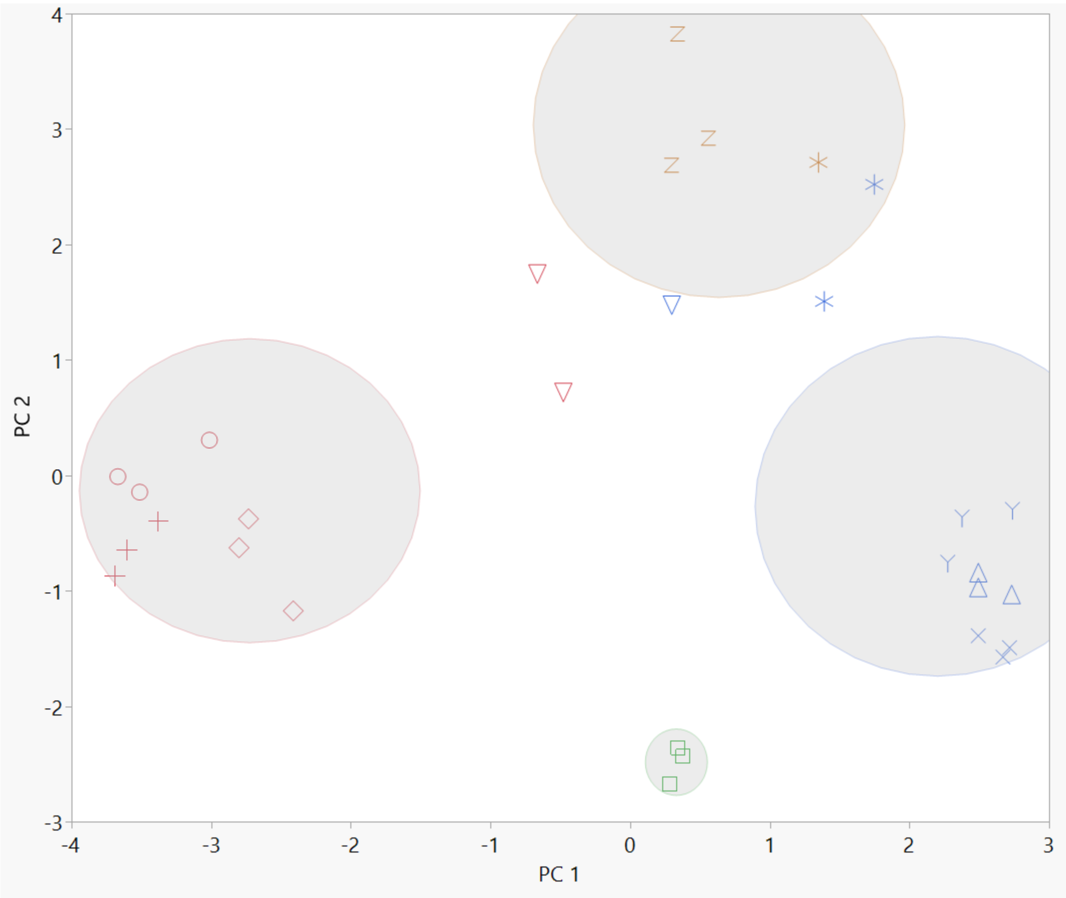

Figure 4. K-means clustering based on principal component 1 and principal component 2

DISCUSSIONS

As reviewed by Cozzolino et al (2019), Principal Component Analysis (PCA) is utilized in food science to organize sensory data and chemometrics generated by various sophisticated machines, and it can even be applied to cases of food fraud and adulteration. Ares et al (2009) interpreted the sensory quality indices of strawberries, focusing on consumer preferences regarding their appearance and odor. Additionally, Selvraj et al (2023) employed PCA to correlate variations in nutritional content and phytochemicals with sensory values based on legume sources. Granato et al (2018) explored PCA tools and hierarchical clustering to establish the relationship between bioactive components and the functional properties of fruit juices. Atsaam et al (2021) categorized 76 types of cereals in the food composition database of West Africa into six groups. Meanwhile, Akbay et al (2000) clustered 155 lamb samples based on fatty acid composition, cholesterol, and energy, dividing them into two major clusters. Laurie et al (2022) demonstrated an inverse relationship between carbohydrates and moisture, indicating that as carbohydrate levels increase, moisture content decreases. The authors noted that carbohydrates were associated with Principal Component 1 (PC1), while moisture positively influenced Principal Component 2 (PC2). Fat, ash, and minerals were positively correlated with PC1 and PC2. In contrast, Aweke and Roba (2016) reported that protein and carbohydrates correlated positively with PC1, while fat was positively correlated with PC2. They pointed out that dry matter positively correlated with PC1 and PC2, whereas fiber and ash were negatively correlated with both principal components. These results contradicted our findings, which might have been influenced by various factors such as the number of parameters and samples (Grane and Jach 2014).

PCA simplifies complex data by integrating it into principal components (a smaller set of data) that capture most of the information (Jolliffe and Cadima 2016). It facilitates easier data visualization, helps identify trends in the dataset, and highlights outliers. Clustering foods aids in grouping them based on similar nutritional properties (Granato et al 2018). Furthermore, PCA is valuable for classifying foods by their nutritional profiles, assisting in the development of dietary guidelines and fortification strategies (Greenfield and Southgate 2003). K-means clustering classifies foods based on the average nutrient values, proving useful in diet planning and food substitution based on preferences (Greenfield and Southgate 2003; Granato et al 2018). Raigond et al (2017) clustered potatoes based on their glycemic properties, while Pandey et al (2023) grouped them into five clusters according to their mineral content. The insights gained from this clustering can significantly benefit the food industry, aiding product development and optimizing marketing strategies. Furthermore, the nutritional information derived from this research can play a crucial role in creating tailored diets for various demographic segments.

CONCLUSION

The investigation into the nutritional composition of potatoes and tubers has revealed significant insights regarding nutrient interrelationships. Results indicate a negative correlation between sodium concentration and levels of potassium, phosphorus, and iron, suggesting that higher sodium may inhibit the absorption of these essential minerals crucial for human health. Additionally, the negative correlation between fat and carbohydrate content and moisture levels highlights the complex interplay of macronutrients in these tubers, potentially reflecting metabolic adaptations for nutrient storage and water retention. Through comprehensive nutritional profiling, potatoes and tubers were categorized into four distinct clusters based on their nutrient composition. This classification enhances authors' understanding of potato diversity and identifies potential breeding avenues to enhance specific nutritional traits. The clustering reveals patterns that could inform dietary choices, illustrating how different varieties contribute uniquely to nutrient intake. These findings underscore the importance of dietary choices in achieving a balanced nutrient profile. The interrelationships among sodium, potassium, phosphorus, iron, fat, carbohydrates, and moisture contents demonstrate how dietary components influence each other's availability in human nutrition. This research contributes valuable knowledge to nutritional science and suggests that strategic selection of potato varieties could help address nutritional deficiencies in populations reliant on these staple foods. Future studies should further explore these correlations and their implications for agriculture and public health.

ACKNOWLDEGEMENT

The authors would like to acknowledge the Nepal Agricultural Research Council, the Asian Food and Agriculture Cooperation Initiative (AFACI), and the Rural Development Administration (RDA) of Korea for their technical and financial support.

REFERENCES

Akbay A, AT Elhan, C Ozcan and S Demirtas. 2000. Hierarchical cluster analysis as an approach for systematic grouping of diet constituents on basis of fatty acid, energy and cholesterol content: application on consumable lamb products. *Medical Hypotheses*, ***55***(2), 147-154.

AOAC. 2005. Official Methods of Analysis of AOAC International. Ed. W Horwithz and GW Latimer, 18th Edn.

Ares G, S Barrios, C Lareo and P Lema. 2009. Development of a sensory quality index for strawberries based on correlation between sensory data and consumer perception. *Postharvest Biology and Technology*, *52*, 97-102.

Atsaam DD, SS Oyelere, OS Balogun, R Wario and NV Blamah. 2021. K-means cluster analysis of the West African species of cereals based on nutritional value composition. *African Journal of Food, Agriculture, Nutrition and Development*, ***21***(1), 17195-17212.

Aweke A, and B Roba. 2016. Evaluation of proximate composition of sweet potato (Ipomoea batatas (L.) Lam) for better utilization through agro-food processing chain. *Global Journal of Animal Breeding and Genetics*, ***4***(4), 249-255.

Bhandari S, S Neupane, and R Tripathi. 2019. Socio-economic impact of potato production in Nepal. *International Potato Journal*, ***46***(2), 1-12.

Cozzolino D, A Power and JJ Chapman. 2019. Interpreting and reporting principal component analysis in food science analysis and beyond. *Food Analytical Methods*, ***12***, 2469-2473.

FAO. 2022. The state of food and agriculture. 2022. Building resilience for peace and food security. Rome, Italy: FAO.

Ghimire MB, DD Subedi and HB Katuwal. 2013. Assessment of potato diversity and its utilization in Nepal. *Journal of Agricultural Research and Development*, ***1***(1), 1-13.

JICA. 2020. Potato Seed Tuber Production Techniques Manual. Kathmandu, Nepal: JICA.

Joshi BK, S Shrestha, B Adhikari and M Bhattarai. 2020. Traditional practices and genetic diversity on chayote landraces and their conservation. Sustain. Dev. 10, 272–288. **DOI**: https://doi.org/10.31924/nrsd.v10i2.060.

Khatri-Chhetri N, PR Bhandari and S Kumar. 2017. Status and prospects of root and tuber crops in Nepal. *Horticulture Nepal*, ***15***(1), 164-178.

Khatri BB, BP Luitel, S Ghimire and BK Joshi. 2017. Diversity and utilization status of root and tuber crops genetic resources in Nepal. In: Conservation and Utilization of Agricultural Plant Genetic Resources in Nepal (BK Joshi, HB KC and AK Acharya, eds). Proceedings of 2nd National Workshop on CUAPGR, 22-23 May 2017 Dhulikhel; NAGRC, FDD, DoA and MoAD; Kathmandu, Nepal. 282-289.

Katuwal HB, DD Subedi and MB Ghimire. 2019. Potato production in Nepal: An overview. In Potato Production, Processing, and Utilization. Springer, Cham, **pp**. 105-121.

MOALD. 2019. Statistical information on Nepalese agriculture. Kathmandu, Nepal: MOALD.

Granato D, JS Santos, GB Escher, BL Ferreira and RM Maggio. 2018. 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*, ***72***, 83-90.

Grane A and A Jach. 2014. Applications of principal component analysis (PCA) in food science and technology. *Mathematical and Statistical Methods in Food Science and Technology*, ***13***(1), 57-86.

Greenfield H and DAT Southgate. 2003. Food composition data: Production, management, and use. Food and Agriculture Organization of the United Nations (FAO).

Johnson DE. 1998. Applied multivariate methods for data analysis. Duxbury Press; 1st edition.

Jolliffe IT and J Cadima. 2016. Principal component analysis: a review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202.

Kaushik M and B Mathur. 2014. Comparative study of K-means and hierarchical clustering techniques. *International Journal of Software & Hardware Research in Engineering*, ***2***(6), 93-98.

Laurie SM, MW Bairu and RN Laurie. 2022. Analysis of the nutritional composition and drought tolerance traits of sweet potato: Selection criteria for breeding lines. *Plants*, ***11***(14), 1804.

Mishra SP, U Sarkar, S Taraphder, S Datta, D Swain, R Saikhom, S Panda and M Laishram. 2017. Multivariate statistical data analysis-principal component analysis (PCA). *International Journal of Livestock Research*, *7*(5), 60-78.

Selvaraj P, A Sanjeevirayar and A Shanmugam. 2023. Application of principal component analysis as properties and sensory assessment tool for legume milk chocolates. *American Journal of Computational Mathematics*, ***13***(1), 136-152.

||-------|||-------||