TRENDS IN CURRENT BIOLOGY
ISSN 2584-2080 (online)
Vol 2(4): 35–38
https://doi.org/10.14719/tcb.4398

HORIZON e-Publishing Group
in association with atom e-Publishing

RESEARCH COMMUNICATION

K-means Clustering assisted ATR-FTIR study for rapid quality evaluation of arrowroot starch

Rinkal N Patel[1], Kunjal L Vegad[1], Sarita J Gagat[2], Shailvi B Joshi[2], Nikunj D Patel[2*]

[1] Department of Pharmaceutical sciences, Parul College of Pharmacy and Research, Parul University, India

[2] Department of Pharmaceutical analysis and Quality assurance, School of Pharmacy, National Forensic Sciences University, An institution of National importance, Ministry of Home Affairs, Govt. of India, Gandhinagar, India

*Email: ndthummar@gmail.com, ndthummar123@yahoo.in

OPEN ACCESS

ARTICLE HISTORY

Received: 19 July 2024
Accepted: 07 September 2024

Available online
Version 1.0 : 26 November 2024

Additional information

Peer review: Publisher thanks Sectional Editor and the other anonymous reviewers for their contribution to the peer review of this work.

Reprints & permissions information is available at https://horizonepublishing.com/journals/index.php/TCB/open_access_policy

Publisher's Note: The article processing was done by atom e-Publishing, Thiruvananthapuram, India. Horizon e-Publishing Group remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

CITE THIS ARTICLE

Patel RN, Vegad KL, Gagat SJ, Joshi SB, Patel ND. K-means Clustering assisted ATR-FTIR study for rapid quality evaluation of arrowroot starch. Trends in Current Biology. 2024; 2(4): 35-38. https://doi.org/10.14719/tcb.4398

Abstract

To solve the issues, arise in quality evaluation of arrowroot starch due to substitution and adulteration, rapid and robust k means assisted Attenuated Total Reflectance -Fourier Transform Infrared (ATR-FTIR) method is developed. Arrowroot starch, locally known as "Aerukka" or "Hulankeeriya" is the rhizomes of plants, *Curcuma angustifolia* Roxb. (Zingiberaceae) and *Maranta arundinacea* Linn. (Marantaceae). Here, ATR-FTIR technique was used to assess the arrowroot starch samples. All the samples were first analyzed in transmittance mode and the spectrum of respective samples was recorded. Resultant unique fingerprint region data was then treated with K-means clustering of machine learning. Coding for K-means clustering was executed in python. K-means is the fastest unsupervised machine learning algorithm to break down data points into groups even when very little information is available and it effectively organizing and extracting insights from complex datasets. The results showcase the effectiveness of K-means clustering as a valuable tool in analytical procedures, aiding in data-driven decision -making, and facilitating meaningful data interpretations which has distinguished all the samples in three different groups. This rapid method can be used as a routine tool to differentiate samples of arrowroot starch from its adulterants and substitutes for their quality assessment.

Keywords

Arrowroot starch; K-means clustering; machine learning; quality; rapid

Introduction

Arrowroot plant, locally known as "Aerukka" or "Hulankeeriya" is a rhizomes of tuber crop that spread horizontally below the surface. Arrowroot plant is having two varieties East Indian Arrowroot (*Curcuma angustifolia* Roxb., Zingiberaceae) and West Indian Arrowroot (*Maranta arundinacea* Linn., Marantaceae). Approximately 20% starch content is available in the rhizome of arrowroot and around 20–30% of this starch is composed of amylose (1). This starch is used as a soothing agent, to treat the digestive issues with celiac disease, and in the food industry it is used as porridge and it is an alternative to wheat flour with no gluten. It is also used to treat Amlapitta and various other diseases. Apart from this, it is a key ingredient in plentiful Ayurvedic formulations, like Eladi Churna, Narikela Khanda Paka, Pippalyady Avaleha, Chyavana Prasha, Vajikarana Ghrita, Sitopaladi Churna, Bala Ghrita, Dadimashtaka Churna, Drakshavaleha and many others (2). Arrowroot starch with excellent gelling ability, is also used for the production of films with good functional properties (3,4).

Both arrowroot starch has proven no toxicity in rat studies indicating their safety profiles. Though, it is critical to note that arrowroot starch is adulterated with starches from tapioca, sweet potato, and rice flour. Moreover, arrow root starch is often misinterpreted and substituted by the siliceous material (Vanskapur). Since the mankind exists, traditional plants have been used as a treatment for different human disorders as a fresh or dried part/s of plant (5). These herbal medicines are popular as they are easily available, economic and harmless in nature and that's why the herbal medicines have been widely utilized for centuries in many countries including India, Korea, China, Japan, etc (6).

Almost, 60% or more phytochemical scrutiny executed in Indian market for herbal plants is based on adulterated samples according to one suggestion. In previous few decades, extensive usage and wide-reaching revival is seen in the area of herbs and formulations and Quality assessment is the need of an hour as a prime concern for marketed herbs and formulations (7). Notably, arrowroot starch found to be adulterated, highlighting the importance of accurate identification and quality control. Thus, the quality control is a key matter for arrow root starch.

Routine quality assessment of arrowroot starch using scanning highly sophisticated techniques is affluent and restrictive for daily sample analysis (8,9). Compare to that, infrared spectroscopy is rapid, non-destructive, cost-effective, and easy to handle method and when this spectroscopy is used with attenuated total reflectance (ATR) then it is particularly beneficial, simplifying the spectra acquisition process and increasing adaptability (10, 11).

In Microscopic studies result is based on hunks of the herbs and certain organelle by comparing them with the existing atlas of the medicinal herbs and these pharmacopeia's hitches are dependent on observers. On other, macroscopic or organoleptic studies are limited by one's sensory organ (12). In compare to that unique spectroscopic pattern is witnessed for each specimen in spectroscopic techniques. In these unique spectra both primary and secondary metabolites accessible in those specimens are reflected. For quality evaluation, Chinese State Food and Drug Administration has permitted both supervised and unsupervised approach of machine learning (13). K-means clustering, as an unsupervised learning algorithm, offers a powerful tool for the quality assessment of herbal drugs using FTIR spectra. By leveraging the inherent patterns within the spectra, K-means clustering categorizes the data into distinct clusters based on similarities, thereby facilitating the effective categorization of herbal drugs according to their quality. This method enables a streamlined and objective approach to assess and differentiate herbal samples, contributing to an efficient and reliable quality control process in the field of herbal drug analysis. By employing K -means clustering, the complex FTIR results are simplified, making them easily interpretable for individuals without specialized expertise (14). This approach enhances the accessibility and comprehension of the obtained results, contributing to a more user-friendly and practical application in quality assessment. Here we have taken 10 samples of arrowroot starch and their possible substitutes for ATR- FTIR analysis and data obtained were treated for K-means clustering (15, 16).

Materials & methods

Material

Arrowroot starch samples (sample 4 and 5 in Table 1) were purchased from the market of Kerala, India and were authenticated. Well-known substitutes like Tikhur and vanslochana samples were purchased from market of Gujarat, India (sample 1 to 3 and 6 to 10 in Table 1).

Table 1: Samples of Arrowroot starch and relevant substitutes

Sr no.	Sample details
1	Tikhur Gandhinagar market, India
2	Tikhur from Ahmedabad market, India
3	Tikhur from LVG store, Ahmedabad, India
4	West Indian Arrowroot starch from Kerala, India
5	East Indian Arrowroot from Kerala, India
6	Vanshlochan (Gandhinagar Ayurvedic Stores)
7	Vanshlochan (Atharva ayurveda, Ahmedabad)
8	Vanshlochan from Ahmedabad market, India
9	Vanshlochan from Usha kirana stores, Ahmedabad
10	Vanshlochan from Jagrutti pharmacy

Before analysis, All the samples were sieved with 60 mesh size sieve and then coarse powder samples were taken for ATR-FTIR analysis. First of all, 10 different spectra were taken for sample 5 (East Indian Arrowroot) from Table 1 to check the reproducibility with ATR-FTIR analysis. All the spectra were showing the replica for the respective peak values and the pattern remained the same throughout the recorded spectra. By considering this, all the samples were taken for spectroscopic analysis to record their chemical fingerprint by means of ATR-FTIR studies. In this study, all the samples were taken one by one on the ATR assembly of FTIR spectrophotometer (Bruker vertex-7). For all the samples, Baseline was corrected and then acquisition for all data was performed to obtain the peaks. From all the range of 400-4000 cm^{-1}, Fingerprint area (500-1500 cm^{-1}) is unique and therefore, the values of transmittance from this area were considered to distinguish the compounds.

Fourier Transform Infrared Spectroscopy (FTIR) has emerged as a powerful analytical technique for studying the chemical composition of herbal drugs, offering valuable insights into the functional groups present in complex mixtures. In the pursuit of ensuring the quality and authenticity of herbal drugs, FTIR spectra have become a focal point for researchers. The application of advanced statistical methods, such as k-means clustering, has further elevated the efficacy of FTIR spectroscopy in providing comprehensive assessments of herbal drug quality. FTIR spectroscopy involves the measurement of the absorption, emission, or reflection of infrared light by a

sample, producing a spectrum that serves as a molecular fingerprint. This technique enables the identification of key functional groups, allowing for the characterization of herbal drugs based on their chemical composition. However, the vast and intricate nature of spectral data necessitates robust methods for data analysis and interpretation.

K-means clustering, a widely used unsupervised machine learning algorithm, has gained prominence in the field of FTIR analysis for herbal drugs. This method classifies spectral data into distinct groups or clusters based on similarities in their absorption patterns. The application of k-means clustering to FTIR spectra offers several advantages in the study of herbal drug quality like Pattern Recognition, Quality Discrimination, Batch-to-Batch Consistency, Identification of Adulterants, and Data Reduction etc. K-means clustering facilitates the identification of inherent patterns within complex FTIR spectra. By grouping similar spectra together, it enables the recognition of characteristic features associated with specific herbal constituents or variations. The algorithm can effectively discriminate between different qualities or varieties of herbal drugs based on their FTIR signatures. This discrimination is crucial for ensuring adherence to quality standards and identifying any potential adulteration or variability in herbal formulations. K-means clustering aids in assessing the consistency of herbal drug batches. By clustering spectra from different batches, it becomes possible to evaluate the reproducibility of manufacturing processes and identify any deviations that may impact the overall quality of the product. The algorithm can assist in detecting the presence of adulterants or contaminants in herbal drugs. Deviations from the expected spectral patterns can raise red flags, prompting further investigation into the integrity of the herbal formulation. K-means clustering contributes to data reduction by categorizing similar spectra into clusters. This simplifies the analysis process and enhances the interpretability of complex FTIR datasets. We explore the application of k-means clustering in conjunction with FTIR spectroscopy for the comprehensive evaluation of herbal drug quality. By leveraging the strengths of both techniques, researchers can gain deeper insights into the chemical composition, consistency and authenticity of herbal formulations, fostering advancements in the field of herbal medicine quality control.

Results

All of the samples of table-1 were analysed with ATR-FTIR spectroscopic method and data was treated for Kohonen K -means Clustering. K-means clustering is a widely used unsupervised machine learning algorithm that partitions data into distinct groups or clusters based on similarities in their features. The idea behind K-means clustering is to divide a dataset into a specified number of clusters (k), where all the points within the same cluster are similar to one another and those in different clusters are different. It starts by randomly assigning each data point to a cluster, and then it iteratively improves the clusters by moving the data points to the cluster center that is closest to them. This logic continues until the cluster assignments stop changing, or a maximum number of iterations is reached. Thus, All the samples were discriminated into three different clusters coded by different colours.

Discussion

Data obtained from ATR-FTIR was treated for K-means clustering and the code was written in python. In K-means clustering analysis, 10 samples were discriminated in three colours Green, Yellow and blue. Where, Colour 1-Blue is for Tikhur showing three clusters, Colour 2- Yellow is for Arrowroot starch showing two clusters and Colour 3-Green is for Vanshlochan showing five clusters. This analysis was cross validated with Hierarchical cluster analysis (HCA) and dendogram obtained from HCA is in concordance with the results of K-means clusters well depicted in figure 2.

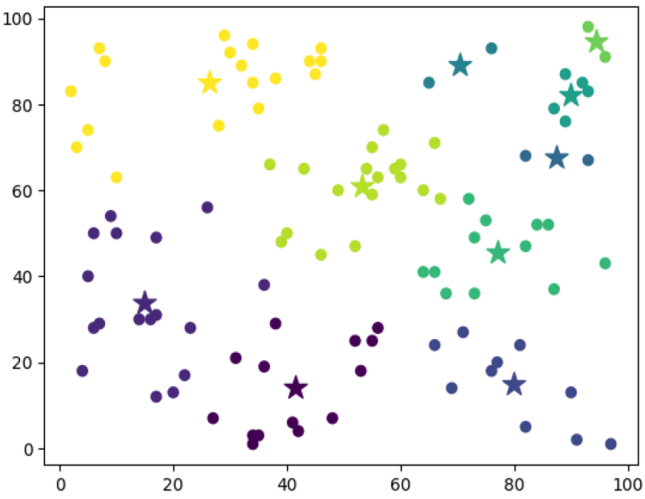

Fig 1: K means clustering of samples for arrowroot starch

Conclusion

In comparison to analytical techniques like HPTLC, RP-HPLC, LC/MS, NMR, which are generally time consuming, expensive, and complex; the proposed K-means clustering assisted ATR-FTIR method is simple, faster, easy to handle, non-destructive and efficient for quality evaluation of arrowroot starch. Prime concern of the study was to keep the Indian intangible culture intact and to produce the strong quality assessment tool that gives the support to traditional text. This method can be used for day-to-day quality assessment of samples of arrowroot starch from its adulterants and substitutes.

Authors' contributions

Rinkal participated in Writing and Instrumental analysis. Kunjal performed analysis especially standardization parameters and supervision. Coding part was carried out by Sarita and Shailvi. Concept and idea, Review and supervision was done by Nikunj. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest: Authors do not have any conflict of interests to declare.

Ethical issues: None.

References

1. Malki MK, Wijesinghe JA, Ratnayake RH, Thilakarathna GC. Characterization of arrowroot (Maranta arundinacea) starch as a potential starch source for the food industry. Heliyon. 2023 ;9(9).
2. Patel ND, Kanaki NS. Quality Evaluation of the Tugaksheeree Samples by ATR-FTIR Spectroscopy Using Multicomponent Analysis. The Natural Products Journal. 2021;11(3):377-82.https://doi.org/10.2174/2210315510999200408125555
3. Nogueira GF, Fakhouri FM, de Oliveira RA. Extraction and characterization of arrowroot (*Maranta arundinaceae* L.) starch and its application in edible films. Carbohydrate Polymers. 2018 ; 186:64-72. https://doi.org/10.1016/j.carbpol.2018.01.024
4. Romero-Bastida CA, Bello-Perez LA, Velazquez G, Alvarez-Ramirez J. Effect of the addition order and amylose content on mechanical, barrier and structural properties of films made with starch and montmorillonite. Carbohydrate Polymers. 2015; 127:195-201.https://doi.org/10.1016/j.carbpol.2015.03.074
5. Rinkal NP, Kunjal LV, Nikunj DP. Discriminating the leaves of *Ocimum sanctum* Linn. conferring to their age by gas chromatographic fingerprint analysis assisted with multivariate analysis. Trends in Current Biology. 2024;2(1). https://doi.org/10.14719/tcb.2947
6. Patel PR, Patel ND, Patel SG, Kanaki NS, Patel AJ. Fingerprint Analysis of *Vitex Negundo* by HPLC Coupled with Multi-Components Analysis. Current Pharmaceutical Analysis. 2020;16 (6):743-51. https://doi.org/10.2174/1573412915666190312161325
7. Patel ND, Kanaki NS. Fingerprint analysis of Shankhpushpi for species discrimination by HPLC coupled with chemometric methods. Journal of Liquid Chromatography & Related Technologies. 2020;43(13-14):455-63. https://doi.org/10.1080/10826076.2020.1728695
8. Patel ND, Kanaki NS. Rapid authentication of Sitopaladi churna samples by attenuated total reflectance-Fourier transform infrared spectroscopy using multivariate analysis. Spectroscopy Letters. 2021;54(6):405-18.https://doi.org/10.1080/00387010.2021.1931784
9. Patel N, Kanaki N, Movaliya V. Assessment of diurnal variation in *Ocimum sanctum* Linn. by gas chromatographic fingerprint analysis coupled with chemometric methods. Plant Science Today. 2018;5(3):131-4.https://doi.org/10.14719/pst.2018.5.3.413
10. Patel ND, Prajapati PN, Kanaki NS. Quality Assessment of Sitopaladi Churna Using High-Performance Liquid Chromatography Coupled with Multivariate Analysis. Journal of chromatographic science. 2020;58(10):961-8. https://doi.org/10.1093/chromsci/bmaa070
11. Patel N, Kanaki N. Arbitrating Identification Issues of *Picrorhiza kurroa* Samples by ATR-FTIR Spectroscopy Using Multivariate Analysis. Pharmaceutical Chemistry Journal. 2021; 55:585-90. https://doi.org/10.1007/s11094-021-02469-y
12. Patel ND, Patel RN, Vegad KL, Kanaki NS. HPLC Fingerprint Analysis Coupled with Multivariate Analysis for Quality Assessment of *Picrorhiza kurroa* Rhizomes. Current Chromatography. 2022;9(1):95-102. https://doi.org/10.2174/2213240609666220930121841
13. Patel N, Kanaki N. Geographical classification of *Adhatoda vasica* based on HPLC fingerprint analysis coupled with multivariate analysis. Journal of Pharmacognosy and Phytochemistry. 2018;7 (4):2488-90.
14. Pelleg D, Moore A. Accelerating exact k-means algorithms with geometric reasoning. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining 1999; (277-281). https://doi.org/10.1145/312129.312248
15. Saikia MJ. K-means Clustering Machine Learning Approach Reveals Groups of Homogeneous Individuals with Unique Brain Activation, Task and Performance Dynamics using fNIRS. IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2023. https://doi.org/10.1109/TNSRE.2023.3278268
16. Patel ND, Dave KT, Patel RN, Vegad KL, Kanaki NS. HPLC based metabolomic fingerprint profiling as a comprehensive quality assessment tool, why not for standardization of Ayurvedic herbs and formulations? Pharmacognosy research, 2024, 16(4): 677-687.