



Application of fuzzy clustering algorithm and hyperspectral images for rice authentication

Mahsa Edris¹ , Mahdi Ghasemi- Varnamkhasti² , Sajad Kiani³ , Hassan Yazdanpanah^{4&5} , Zahra Izadi⁶ 

1. Mechanical Engineering of Biosystems Department, Faculty of Agriculture, Shahrekord University, Shahrekord, I.R.

Iran. E-mail: mahsa.edris@gmail.com

2. Corresponding Author, Mechanical Engineering of Biosystems Department, Faculty of Agriculture, Shahrekord University, Shahrekord, I.R. Iran. E-mail: ghasemymahdi@gmail.com

3. Biosystems Engineering Department, Faculty of Agriculture, Sari University of Agricultural Sciences and Natural Resources, Sari, I.R. Iran. E-mail: s.kiani@sanru.ac.ir

4. Food Safety Research Center, Shahid Beheshti University of Medical Sciences, Tehran, I.R. IRAN

5. Toxicology and Pharmacology Dept., School of Pharmacy, Shahid Beheshti University of Medical Sciences, Tehran, I.R. Iran, E-mail: yazdanpanah@sbmu.ac.ir

6. Mechanical Engineering of Biosystems Department, Faculty of Agriculture, Shahrekord University, Shahrekord, I.R. Iran. E-mail: z.zahraizadi@gmail.com.

Article Info

Article type: Research Article

Article history:

Received: Oct. 28, 2024

Revised: Jan. 6, 2025

Accepted: Jan. 19, 2025

Published online: Summer 2024

Keywords:

Adulteration,

Clustering,

Non-destructive methods,

Pre-processing

ABSTRACT

Rice is a vital and strategic product that is used as a major food source. The high demand for purchasing and consuming rice leads to the adulteration of this product globally. Hence, a non-destructive and rapid method is needed to verify the authenticity of rice. Hashemi rice, a high-priced and high-quality rice in the market, is combined with rice such as Neda and Shiroudi, which are very similar in shape but lower in quality and price than Hashemi rice. This study used hyperspectral imaging (HSI) coupled with a fuzzy clustering algorithm to assess adulteration in Hashemi rice samples. First, to reduce the data's dimensionality, the principal component analysis method was applied to the preprocessed data using the multiplicative dispersion correction and Savitzky-Golay methods. Then, the fuzzy unsupervised clustering algorithm was applied using the whole spectrum wavelength (400-1000 nm). It was able to separate the original sample from the adulterated samples well. Also, the fuzzy membership diagram separated the original and self-adulterated samples, mixing 5% to 50%, confirming the correctness and capability of the fuzzy method. Therefore, the HSI system with fuzzy unsupervised algorithms can be used as a reliable and out-of-laboratory method for rapid rice authenticity evaluation.

Cite this article: Edris, M., Ghasemi-Varnamkhasti, M., Kiani, S., Yazdanpanah, H., Izadi, Z., (2024) Application of fuzzy clustering algorithm and hyperspectral images for rice authentication, *Iranian Journal of Biosystem Engineering*, 55 (2), 79-92.

<https://doi.org/10.22059/ijbse.2025.384493.665571>

© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/ijbse.2025.384493.665571>





EXTENDED ABSTRACT

Introduction

After wheat, rice is the second strategic crop in the world and is one of the main sources of dietary fiber, fat, protein, and other rare nutrients. Starch, protein, and fat are the main components of rice grain. Due to the high rice demand, this product is very vulnerable to adulteration at the global level. Considering the increasing scale and nature of adulteration, using reliable and fast methods to deal with rice adulteration is necessary. Fast, non-destructive, and non-contact hyperspectral imaging (HSI) method is a combination of imaging and spectroscopic technologies. With this method, spatial and spectral information can be obtained simultaneously for each point of the rice sample. This study aims to develop the HSI method and the clustering method to investigate the ability of a fast and non-destructive method for authenticating and predicting the percentage of Iranian rice adulteration.

Materials and methods

In this research, a model was developed to detect rice adulteration by HSI device along with the fuzzy clustering method. HSI is a non-destructive and fast method and in the wavelength range of 400 to 1000 nm, all original (Hashmi) and adulterated (Shiroudi and Neda) samples and combinations of fake rice with Hashemi rice from 5 to 50% were scanned. Then they were pre-processed using MSC and SG algorithms and the best wavelengths were selected and checked using evolutionary wavelength selection. Finally, using the FCM method, a model was presented to detect adulteration in rice.

Results and discussion

The results showed that the clustering method (fuzzy) together with the HSI system (as a non-destructive, fast, and accurate system) for predicting adulteration in rice samples shows satisfactory results. PCA after applying MSC + SG pre-processing, three samples of Hashemi, Shiroudi, Neda and are well separated from each other and can be identified. PC1, PC2, and PC3 after MSC + SG pre-processing for Hashemi, Shiroudi, and Neda cultivars were calculated as 60, 17 and 7%, respectively, with a total explained variance of 84% which was distributed by three PC1, PC2, and PC3. Then, after PCA, the pre-processed spectra were used for the C-means fuzzy clustering model. In the fuzzy clustering results, the original sample (in green color) and the fake sample (in red color) can be seen, and the fuzzy membership chart was used to confirm the results and further analysis. Two Hashmi rice samples (original) in green color and the Neda sample and its counterfeits (counterfeit) in red color (100 pixels or the number of first samples) belong to original and counterfeit clusters with high membership degree (almost complete). After the amount of 100 samples (100-400), two samples of Hashemi and Neda are mixed, which proves the combination of these two samples (from 5% to 50%). Fuzzy results in the full wavelength of its transparency over the selected wavelength.

Conclusion

Results proved that the HSI system coupled with the fuzzy clustering method can satisfactorily evaluate the rice authenticity. The fuzzy membership graph can predict adulteration in rice samples well.

Credit authorship contribution statement

Mahsa Edris: Writing – original draft, Methodology, Data curation, Software, Formal analysis.

Mahdi Ghasemi Varnamkhasti: Review and edit, conceptualize, supervise, and manage project administration.

Sajad Kiani: Writing – review and editing, Formal analysis, Investigation, Data curation, Validation.

Hasan Yazdanpanah: Supervision, Project administration, Methodology, Resources.

Zahra Izadi: Supervise, project administration, methodology, and resources.

Data Availability Statement

Data available on request from the authors. All the data used in this original research are presented throughout the text and in the form of Tables and Figures.

Ethical considerations

The authors avoided data fabrication, falsification, plagiarism, and misconduct.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

The authors extend their sincere appreciation to Shahrekord University, Sari University of Agricultural Sciences and Natural Resources, Rice Research Center of Iran, and Food and Drug Administration for their support throughout this project.

Conflict of interest

The author declares no conflict of interest.

کاربرد الگوریتم خوشه‌بندی فازی و تصاویر ابرطیفی به منظور اصالت‌سنجی برنج

مهسا ادريس^۱، مهدی قاسمی ورنامخواستی^۲✉، سجاد کیانی^۳، حسن یزدان‌پناه^۴ و زهرا ایزدی^۵۱. گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی دانشگاه شهرکرد، شهرکرد، ایران mahsa.edriis@gmail.com۲. نویسنده مسئول، گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران ghasemymahdi@gmail.com۳. گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه علوم کشاورزی و منابع طبیعی ساری، ساری، ایران s.kiani@sanru.ac.ir۴. گروه داروشناسی-سم‌شناسی، دانشکده داروسازی دانشگاه شهید بهشتی، تهران تهران yazdanpanah@sbmu.ac.ir مرکز تحقیقات سلامت

غذا، دانشگاه علوم پزشکی شهید بهشتی، تهران، ایران

۵. گروه مهندسی مکانیک بیوسیستم، دانشکده کشاورزی، دانشگاه شهرکرد، شهرکرد، ایران z.zahraizadi@gmail.com

چکیده

اطلاعات مقاله

نوع مقاله: مقاله پژوهشی

برنج یک محصول حیاتی و راهبردی است که به عنوان منبع اصلی غذایی مورد استفاده قرار می‌گیرد. تقاضای بالا برای خرید و مصرف برنج، منجر به تقلب در این محصول در سطح جهانی می‌شود. از این رو روشی غیرمخرب و سریع برای احراز اصالت برنج نیاز است. برنج هاشمی که به عنوان برنج گران قیمت و با کیفیت بالای بازار ایران معروف است با برنج‌های همچون ندا و شیرودی که از لحاظ شکل بسیار شبیه ولی از لحاظ کیفیت و قیمت پایین‌تر از آن هستند، ترکیب می‌شود. در این پژوهش از تصویربرداری ابرطیفی همراه با الگوریتم خوشه‌بندی فازی برای ارزیابی تقلب در نمونه‌های برنج هاشمی استفاده شد. ابتدا به منظور کاهش ابعاد داده‌ها روش تجزیه و تحلیل مولفه‌های اصلی بر روی داده‌های پیش‌پردازش شده به‌وسیله روش‌های اصلاح پراکندگی ضربی و ساویتزکی-گولای اعمال شد. سپس الگوریتم خوشه‌بندی بدون نظارت فازی با استفاده از طول موج کامل طیف (۴۰۰-۱۰۰۰ نانومتر) به خوبی توانست نمونه اصلی را از نمونه‌های تقلبی جدا کند. همچنین نمودار عضویت فازی نمونه‌های اصلی و تقلبی و اختلاط ۵ درصد تا ۵۰ درصد را به خوبی از هم جدا کرد و درستی روش فازی را تایید کرد. بنابراین، سامانه تصویربرداری ابرطیفی همراه با الگوریتم‌های بدون نظارت فازی را می‌توان به عنوان روشی مطمئن و خارج از آزمایشگاه برای اصالت‌سنجی سریع برنج هاشمی و امکان‌سنجی وجود تقلب در آن را استفاده کرد.

تاریخ دریافت: ۱۴۰۳/۸/۷
تاریخ بازنگری: ۱۴۰۳/۱۰/۱۷
تاریخ پذیرش: ۱۴۰۳/۱۰/۳۰
تاریخ انتشار: تابستان ۱۴۰۳

واژه‌های کلیدی:

پیش‌پردازش،
تقلب،
خوشه‌بندی،
روش‌های غیرمخرب.

استناد: ادريس؛ مهسا، قاسمی ورنامخواستی؛ مهدی، کیانی؛ سجاد، یزدان‌پناه؛ حسن، ایزدی؛ زهرا، (۱۴۰۳) کاربرد الگوریتم خوشه‌بندی فازی و تصاویر ابرطیفی به منظور

اصالت‌سنجی برنج، مجله مهندسی بیوسیستم ایران/ ایران، ۵۵ (۲)، ۹۲-۷۹. <https://doi.org/10.22059/ijbse.2025.384493.665571>

© نویسندگان.

ناشر: مؤسسه انتشارات دانشگاه تهران.

DOI: <https://doi.org/10.22059/ijbse.2025.384493.665571>

مقدمه

برنج (*Oryza sativa* L.) یکی از محصولات غذایی اصلی در دنیا است. برنج حاوی بسیاری از مواد مغذی همچون چربی، پروتئین، فیبر غذایی، آهن، کلسیم و ویتامین B1 است که برای سلامتی انسان مفید است. بیشترین میزان برنج در جهان در کشورهای آسیایی تولید می‌شود (Edris et al., 2024; Teye et al., 2019). برنج با توجه به ساختار و مواد مغذی مشابه دارای کیفیت غذایی مختلفی است. یکی از چالش‌های رایج در صنعت برنج، تقلب است. برخی تاجران با جایگزینی محصول درجه یک با برنج با کیفیت پایین‌تر برای سود بیشتر، تقلب انجام می‌دهند (McGrath et al., 2021). یکی دیگر از روش‌های تقلب این است که برنج کهنه را با برنج تازه مخلوط می‌کنند که در شکل ظاهری، طعم و رنگ آن تأثیر چندانی نمی‌گذارد، اما به شدت به منافع مصرف‌کننده آسیب می‌رساند (Lai et al., 2014). یکی از موارد تقلب رایج در عرضه برنج در ایران به این صورت است که انواع برنج پر محصول (به طور مثال رقم شیرودی) با قیمت و کیفیت پایین با انواع برنج مرغوب (به طور مثال رقم هاشمی) که از نظر ظاهری شباهت زیادی به هم دارند، مخلوط می‌شوند (Fathi & Nabipour., 2019). بنابراین، احراز اصالت برنج امری ضروری است.

روش‌های مختلفی از جمله کروماتوگرافی گازی- طیف‌سنجی جرمی^۱ (GC-MS)، روش‌های تجزیه و تحلیل مبتنی بر DNA و طیف‌سنجی جرمی پلاسمای جفت القایی (ICP-MS) برای تشخیص تقلب برنج استفاده می‌شوند با این حال این روش‌ها زمان‌بر و مخرب بوده و در عین حال به کارشناس حرفه‌ای نیازمند هستند. از این رو یک روش سریع و اقتصادی و غیرمخرب برای اصالت‌سنجی و تشخیص تقلب برنج الزامی است (Aznan et al., 2022). در سال‌های اخیر روش‌های غیرمخرب همچون طیف‌سنجی فرو سرخ نزدیک (NIR) (Shannon et al., 2021)، تصویربرداری ابرطیفی^۲ (HSI) (Weng et al., 2020; Edris et al., 2024; Van Haute et al., 2023) و بینی الکترونیکی^۳ (Ghasemi-Varnamkhasti et al., 2011) برای ارزیابی کیفیت و اصالت‌سنجی غلات همچون برنج استفاده شدند. روش سریع، غیرمخرب و غیر تماسی HSI، یک روش ترکیبی از تکنولوژی‌های تصویربرداری و طیف‌سنجی است. با این روش می‌توان اطلاعات مکانی و طیفی را به طور همزمان برای هر نقطه از نمونه مورد آزمون بدست آورد. در واقع با این روش امکان استخراج اطلاعات دقیق‌تری وجود دارد چرا که محدودیت‌هایی که در روش‌های طیف‌سنجی NIR که تنها طیف متوسط یک نقطه در نمونه را ارائه و اطلاعاتی در مورد توزیع مکانی اجزای تشکیل دهنده در نمونه را ندارد و یا تصویربرداری رنگی که فقط سه باند بصری (R، G و B) را ارائه می‌کند را رفع می‌کند (Kheiralipour, & Jayas., 2024; Kiani et al., 2018). با توجه به اینکه برنج یک محصول دانه‌ای است و امکان متفاوت بودن دانه‌ها به هر دلیلی وجود دارد سامانه HSI برای داده برداری طیفی و مکانی دانه‌های برنج بسیار مناسب است. از سوی دیگر ترکیب این سامانه با روش‌های داده کاوی فازی می‌تواند خوشه‌بندی دانه‌ها با درجات کیفی از نظر طیفی را به خوبی انجام دهد.

خوشه‌بندی فازی، یک روش یادگیری بدون نظارت است که هدف آن کشف گروه‌بندی طبیعی اشیاء بر اساس شباهت ویژگی‌های ذاتی، است (Theodoridis et al., 2010). الگوریتم خوشه‌بندی فازی درجات مختلفی از عضویت را به الگوهای هر خوشه اختصاص می‌دهد تا یک رابطه غیر دودویی بین آنها ایجاد کند (Adán & Wilfrido, 2015).

پیشینه پژوهش

در پژوهشی، از طیف‌سنجی NIR با دامنه طیفی ۱۰۰۰ تا ۲۵۰۰ نانومتر در ترکیب با تجزیه و تحلیل تفکیک حداقل مربعات جزئی^۴ PLS^۵ و ماشین بردار پشتیبان^۶ (SVM) برای تشخیص تقلب در برنج استفاده شد. در این تحقیق هر گروه شامل ۲۰ نمونه خالص و ۱۴۰ نمونه مخلوط با نسبت اختلاط بین ۵ تا ۵۰ درصد بود. مدل‌های PLS-DA و SVM توانستند برنج تقلبی را با محدودیت تشخیص ۵

1. Gas Chromatography-Mass Spectrometry
2. Inductively Coupled Plasma Mass Spectrometry
3. Near-infrared
4. Hyperspectral imaging
5. Electronic nose
6. Microextraction
7. Partial Least Square Discriminant Analysis
8. Support Vector Machines

درصد و با ($R^2_c = 0.97$, $RMSE_c = 2.78$, $R^2_p = 0.96$ و $RMSE_p = 2.93$) شناسایی کنند (Liu et al., 2020).

در پژوهشی برای مقابله با تقلب برنج باسماتی هندی یک سیستم تجزیه و تحلیل دو مرحله‌ای ایجاد شد. با استفاده از سامانه NIR قابل حمل با دامنه طیفی (۱۰۷۰-۷۴۰ نانومتر) و از روش ریز استخراج با فاز جامد در فضای فوقانی^۵ (HS-SPME) ترکیبات گازی استخراج و با استفاده از GC-MS به شناسایی تقلب در ترکیبات گازی نمونه برنج باسماتی پرداخته شد. مزیت اصلی سیستم دو مرحله‌ای این است که تعداد زیادی از نمونه‌ها را می‌توان در مزرعه با استفاده از سامانه NIR قابل حمل غربال کرد و نمونه‌های غیرمنطبق را می‌توان برای تجزیه و تحلیل تاییدی به آزمایشگاه ارسال کرد که این روش باعث کاهش زمان و هزینه می‌شود. علاوه بر این، عطر برنج، طبقه‌بندی بسیار خوبی ($R^2 < 0.9$ و دقت < 0.99) نشان داد. انواعی از گونه‌های بالقوه تقلبی با آلدئیدهایی که به عنوان ترکیبات نشانگر عطر بودند شناسایی شدند (Shannon et al., 2021).

در پژوهشی دیگر از ترکیب روش‌های غربالگری و تایید در یک سیستم آزمون چند لایه برای نظارت و کاهش تقلب در برنج استفاده شد. ترکیب روش‌های غربالگری و تایید در یک دستگاه آزمون چند لایه با غربالگری سریع NIR قابل حمل با دامنه طیفی ۱۰۷۰-۷۴۰ نانومتر، و با تایید مرحله دوم با استفاده از تجزیه و تحلیل مبتنی بر طیف‌سنجی جرمی (کروماتوگرافی مایع- طیف‌سنجی جرمی LC-MS، GC-MS و ICP-MS) نمونه‌های مشکوک در آزمایشگاه انجام شد. مدل‌های طیف‌سنجی NIR قابل حمل و روش‌های کروماتوگرافی گازی- طیف‌سنجی جرمی مبتنی بر آزمایشگاه برای تمایز بین گونه‌های برنج باسماتی با ارزش بالا و تقلب‌کننده‌های بالقوه آن‌ها ایجاد شدند. علاوه بر این، یک روش طیف‌سنجی جرمی پلاسمای جفت القایی برای تایید از انواع برنج چینی و همچنین یک روش طیف‌سنجی جرمی کروماتوگرافی مایع برای تمایز کیفیت در برنج ویتنام ایجاد شد. در نتیجه سیستم چند لایه با غربالگری سریع NIR توانست به خوبی سطح قابل توجهی از آزمایش را از طریق غربالگری سریع در خارج از آزمایشگاه برای تصمیم‌گیری و تشخیص تقلب نهایی فراهم کند (McGrath et al., 2021).

همچنین از سامانه HSI همراه با چند روش پیش‌پردازش داده‌ها و مدل‌های طبقه‌بندی برای شناسایی تازگی دانه‌های برنج استفاده شد. این سامانه همراه با مدل ماشین بردار پشتیبان با پیش‌پردازش Savitzky-Golay و هشت باند موج توانست به دقت طبقه‌بندی بالای ۹۳/۶۷ درصد دست یابد. نتایج نشان داد این روش توانایی بالایی برای شناسایی تازگی بذر برنج دارد (He et al., 2019).

در پژوهشی دیگر از سامانه NIR همراه با یک روش رگرسیون چند متغیره مناسب (PLS) برای تعیین محتوای آمیلوز برنج استفاده شد. نتایج نشان داد فناوری NIR با دقت بالا ($R = 0.94$; $RMSEP = 1.938$) در بازه‌های طیفی ($8194 - 8941$ و $5592 - 5045$ cm^{-1}) برای تعیین آمیلوز برنج به عنوان روشی سریع و غیرمخرب قابل استفاده است (Sampaio et al., 2018). از سامانه HSI در محدوده ۱۰۰۰-۴۰۰ نانومتر برای شناسایی برنج آلوده به پارافین استفاده شد. در این پژوهش از روش نظارت شده PLS-DA استفاده شد. نتایج نشان داد PLS-DA دقت بالایی (۹۴ درصد) برای تشخیص دانه‌های برنج آلوده به پارافین نشان داد. در واقع در این مطالعه پتانسیل ترکیب تصویربرداری فراطیفی با شیمی‌سنجی PLS-DA نشان داده شد (Faqeerzada et al., 2023). در جدیدترین پژوهش، با استفاده از سامانه HSI در محدوده ۴۰۰ تا ۹۵۰ نانومتر برای تشخیص آرد قنادی و آرد نان‌های سامون، سنگک و تافتون استفاده شد. از روش‌های تحلیل تفکیک خطی، ماشین بردار پشتیبان و شبکه عصبی مصنوعی در نرم‌افزار متلب استفاده شد. دقت طبقه‌بندی شبکه عصبی مصنوعی (۹۸ درصد) بالاتر از روش‌های دیگر بود. نتایج نشان داد تصویربرداری ابرطیفی همراه با شبکه عصبی مصنوعی توانایی بالایی برای تشخیص انواع مختلف آرد گندم را دارد (Nargesi et al., 2024).

در مطالعه‌ای از سیستم بین‌الکترونیکی و تصویربرداری فراطیفی همراه با شبکه عصبی کانولوشنال^۷ (CNN) برای طبقه‌بندی کیفیت برنج استفاده شد CNN با یک ماشین یادگیری افراطی گسترش جهانی^۸ (GE-ELM) که روشی است برای متعادل کردن

1. Coefficient of Determination
2. Rot Mean Square of Calibration
3. Predicted R-squared
4. Rot Mean Square Error of Prediction
5. Headspace Solid-Phase Microextraction
6. Liquid Chromatograph-Mass Spectrometer-Mass Spectrometer
7. convolutional neural network
8. global extension extreme learning machine

تفاوت‌های بین کلاس‌ها ترکیب شد در نتیجه CNN-GE-ELM به عنوان یک تکنیک تحلیلی موثر برای تفاوت کیفیت برنج با عملکرد طبقه‌بندی عالی ۹۸/۰۷ درصد گزارش شد. (Shi et al., 2021)

در پژوهشی از یک طیف‌سنج NIR جیبی همراه با داده‌های چند متغیره برای تشخیص تقلب رقم برنج یاس با رقم آگرا استفاده شد. در این پژوهش مدل^۱ SD-PLSDA دقت ۹۸ درصد در هر دو مجموعه کالیبراسیون و پیش‌بینی برای نمونه‌های پودر برنج در مقایسه با دانه برنج به دست آورد. همچنین برای تعیین کمیت تقلب، نمونه‌های پودر برنج در مقایسه با دانه برنج، دقت بالاتری $R^2 = ۰/۹۴$ توسط Si-PLS^۲ در هر دو مجموعه کالیبراسیون و پیش‌بینی و با RMSEP و RMSECV در ۰/۱۳ و ۰/۱۶ به دست آورد (Amuah., 2022). (Teye, &

در تحقیقات اخیر با استفاده از گروه‌های از پیش تعریف شده طبقه‌بندی ارقام برنج انجام شده است. اما تعداد خوشه‌های واقعی (ذاتی) با تعداد خوشه‌های از پیش تعریف شده می‌تواند متفاوت باشد بنابراین از طریق روش‌های بدون نظارت همچون فازی می‌توان محدودیت‌هایی بین تعداد خوشه‌های ذاتی واقعی و گروه‌های از پیش تعریف شده را از بین رود و عملکرد هوشمندتری برای تعیین وجود یا عدم وجود ناخالصی در یک خوشه اصلی را ایجاد کرد. هدف از این مطالعه توسعه روش HSI در ترکیب با روش خوشه‌بندی فازی برای ایجاد یک روش سریع و غیرمخرب به منظور احراز اصالت و شناسایی تقلب در برنج هاشمی است.

روش‌شناسی پژوهش

جمع‌آوری نمونه

در این پژوهش سه رقم برنج ایرانی (هاشمی (S1)، شیروودی (S2)، و ندا (S3)) در نظر گرفته شد. رقم‌های شیروودی و ندا (به عنوان رقم فرعی یا تقلب) به ترتیب هر کدام از ۵ تا ۵۰ درصد و با گام حرکتی ۵ درصد، با نمونه‌ی هاشمی (اصلی) مخلوط شدند. برای اطمینان از اختلاط، نمونه‌ها در سه تکرار آماده شدند. رقم اصلی ارزش اقتصادی بالایی دارند و رقم‌های فرعی که ارزش اقتصادی پایین‌تری دارند و از نظر شکل شبیه هستند به عنوان تقلب به آن اضافه می‌شوند. اصالت رقم‌های مورد استفاده از قبل توسط مرکز تحقیقات برنج کشور واقع در آمل تایید شده است.

سامانه تصویربرداری ابرطیفی

سامانه تصویربرداری HSI از یک دوربین اسکن خطی قابل حمل (Specim IQ با فناوری Specim push-broom)، که برای هدف مشخص، قابلیت برنامه‌نویسی و خودکارسازی دارد تشکیل شده است (Specim, Spectral Imaging Ltd, Oulu, Finland). همچنین این دوربین شامل یک لنز جلو، یک طیف‌نگار تصویربرداری و یک حسگر تصویر است. علاوه بر این، سامانه HSI شامل رایانه‌ای برای اجرای نرم افزار پردازش تصویر و داده کاوی است. تصاویر هر نمونه در محدوده طول موج ۴۰۰ تا ۱۰۰۰ نانومتر و فاصله طیفی ۲ تا ۳ نانومتر ایجاد می‌شود. سامانه روشنایی مبتنی بر هالوژن، که در محدوده کامل ۴۰۰ تا ۱۰۰۰ نانومتر را پوشش می‌دهد، مورد استفاده قرار گرفت (تدارک دیده شده توسط شرکت تولید کننده دوربین). دوربین یک مکعب ابر طیفی سه بعدی^۴ به نام ابر مکعب^۵ ایجاد می‌کند. یک تصویر ابر طیفی را می‌توان $I(X, Y, \lambda)$ توصیف کرد که مجموعه‌ای از تصاویر مکانی جدا شده $I(X, Y)$ را برای هر طول موج جداگانه (λ) در نظر می‌گیرد. بنابراین، با استفاده از خصوصیات مفید روش‌های تصویربرداری و طیف‌سنجی، سامانه تصویربرداری HSI هزاران طیف برای هر پیکسل در تصویر مهیا می‌کند. مکعب تصاویر ایجاد شده توسط سامانه HSI ویژگی‌های طیفی و مکانی از هر پیکسل هر دانه برنج ارائه می‌دهد که منحصر به همان نقطه از ماده است و می‌تواند برای توصیف و شناسایی هر ماده خاص استفاده شود. قبل از اسکن، دوربین با یک سه پایه مناسب ثابت شد (شکل ۱) و سپس از یک مرجع سفید و یک مرجع رنگی (تدارک دیده شده توسط شرکت تولید کننده دوربین) برای اصلاح تغییرات در روشنایی و کالیبره کردن استفاده شد. مقادیر بازتاب پیکسل در برابر طول موج، اطلاعات شیمیایی خاص در دو بعد را برای یک طیف کامل نشان می‌دهد. قبل از انجام اندازه‌گیری‌ها، همگنی سیستم روشنایی و حساسیت دوربین در نمای اسکن شده ارزیابی و تایید شد. با استفاده از این سامانه نمونه‌ها می‌توانند بدون هیچ گونه آماده‌سازی خاصی طیف‌برداری شده و همچنین

1. Second derivative- Partial least square
2. Synergy interval partial least square
3. Sample
4. Three-Dimensional Hyperspectral Cube
5. Hypercube

برای طیف‌برداری مجدد در حین عملیات فرآوری، انبارداری و همچنین طی مراحل رشد گیاهی (به منظور پیش‌بینی برخی از خواص گیاهان در طول رشد) مورد استفاده قرار گیرند. علاوه بر این، یک مجموعه داده‌های طیفی می‌تواند به خصوصیات فیزیکی مواد، توزیع ترکیبات شیمیایی و سایر اطلاعات داخلی مانند عیوب در سطح مواد خوراکی و سایر محصولات کشاورزی مرتبط شوند (Seo et al., 2020; Siripatrawan et al., 2011; Shaw et al., 2002; ElMasry & sun, 2010).



شکل ۱. تصویری از دوربین ابرطیفی همراه با نمونه‌های برنج و صفحه کالیبراسیون که یک مرتبه در ابتدای داده برداری نیاز است

روش‌های پیش‌پردازش طیفی

تجزیه و تحلیل داده‌های HSI بر محاسبات ریاضی متکی است. بنابراین، استفاده از روش مناسب برای کاهش ابعاد داده‌های طیفی با ابعاد بالا و تجسم تفاوت بین آنها بسیار معنادار است (Miao et al., 2018). در این پژوهش، مکعب ابرطیفی $(204 \times 31 \times 31)$ با ابعاد فضایی پیکسل‌های X و Y (31×31) پیکسل و طول موج‌های λ (204) طول موج روی سطح نمونه به دست آمد سپس به یک ماتریس دو بعدی (204×961) مرتب شد. در ابتدا از روش پیش‌پردازش اصلاح پراکندگی ضربی^۱ (MSC) که پرکاربردترین روش پیش‌پردازش برای طیف‌های NIR است استفاده شد. با خطی کردن هر طیف، اثر پراکندگی نامطلوب را حذف و یا اثر پراکندگی غیریکنواخت جبران شد. سپس از روش پیش‌پردازش ساویتزکی-گولای^۲ (SG) که هدف آن روشی برای هموار کردن و حذف نویز (شیب، وارونگی و غیره) استفاده شد در واقع محتوای اطلاعاتی را بدون خراب کردن سیگنال حفظ می‌کند (Edris et al., 2024; Savitzky & Golay, 1964; Teye et al., 2019). برای کاهش داده‌های طیفی، با میانگین‌گیری یک ماتریس جدید (35×204) تشکیل شد. تجزیه و تحلیل داده‌ها با استفاده از روش‌های مرسوم مختلف از جمله تجزیه و تحلیل مؤلفه‌های اصلی^۳ (PCA)، الگوریتم فازی c-means (FCM)^۴ با تمام طول موج‌ها و انتخاب طول موج‌های تکاملی انجام شد.

روش تجزیه و تحلیل مؤلفه‌های اصلی (PCA) از یک تبدیل متعامد برای تبدیل مجموعه‌ای از مشاهدات متغیرهای احتمالاً همبسته به مجموعه‌ای از مقادیر متغیرهای خطی غیر همبسته به نام مؤلفه‌های اصلی استفاده می‌کند. PCA برای کاهش ابعاد ماتریس داده، فشرده کردن اطلاعات به مؤلفه‌های اصلی PC، تجسم گروه‌بندی‌های طبیعی درون داده‌ها، ارزیابی تأثیر متغیرها در تمایز آنها، انجام پیش‌بینی‌ها در جهت حداکثر واریانس، شناسایی الگوهای پنهان و نقاط پرت در داده‌ها استفاده می‌شود (Abdi & Williams., 2010; Sampaio et al., 2020; Kiani et al., 2010).

1. Multiplicative scatter correction
2. Savitsky-Golay
3. Principal Component Analysis
4. Fuzzy c-means

انتخاب طول موج تکاملی

در مسائل با ابعاد بالا به منظور کاهش فضای جستجو و بهبود عملکرد یادگیری، انتخاب ویژگی موثرتر در داده‌کاوی اهمیت دارد. در واقع با انتخاب طول موج، مؤثرترین ویژگی‌ها از طیف‌های مربوط به عنصر مورد نظر مورد بررسی قرار می‌گیرد. روشی برای انتخاب بهترین طول موج یا فواصل طول موج از داده‌های طیفی NIR برای مدل‌سازی کالیبراسیون چند متغیره، انتخاب طول موج تکاملی^۱ است. این روش توجه زیادی را به دلیل توانایی جستجو در کل داده‌ها را به خود جلب کرده است. انتخاب طول موج تکاملی پیچیدگی مدل و فضای جستجو را کاهش می‌دهد و مدل را بهبود می‌بخشد. الگوریتم انتخاب طول موج تکاملی با استفاده از تکامل دیفرانسیل همراه با شاخص دیویس-بولدین^۳ (DB) برای حذف افزونگی در داده‌های طیفی است که طول موج‌های موثر در خوشه‌بندی را انتخاب می‌کند. این الگوریتم مبتنی بر تکامل دیفرانسیل به طور همزمان واریانس درون گروهی را به حداقل می‌رساند و جداسازی را به حداکثر می‌رساند. برای این منظور، معیارهای DB که واریانس درون خوشه‌ای و فاصله خوشه‌ها در مجموعه‌ای از داده‌ها و بردارها است محاسبه می‌شود. انتخاب طول موج اتخاذ شده به طور قابل توجهی زمان پردازش HSI را کاهش می‌دهد و آن را برای برنامه‌های بر خط خارج از آزمایشگاه مناسب می‌کند (Li et al., 2022, Izadi & Kiani., 2023).

مدل FCM

خوشه‌بندی فازی روشی با عملکرد انعطاف‌پذیر و قابل اعتماد است زیرا عضویت الگوها را بین گروه‌های همپوشانی به اشتراک می‌گذارد. خوشه‌بندی فازی، بر حسب مجموعه‌های فازی است که در آن هر الگو به بیش از یک خوشه در آن واحد تعلق دارد. شناخته شده‌ترین الگوریتم فازی، c-means (FCM) است که توسعه فازی روش k-means است. FCM تلاش می‌کند تا یک خوشه فازی برای مجموعه‌ای از نقاط داده پیدا کند در حالی که تابع هزینه را به حداقل می‌رساند. الگوریتم حاصل، ابرهای کروی از نقاط را در فضای چند بعدی شناسایی می‌کند. این خوشه‌ها تقریباً به یک اندازه فرض می‌شوند. هر خوشه با مرکز خود نمایش داده می‌شود. این نمایش خوشه‌ها را مدل می‌نامند زیرا اغلب تصور می‌شود که تمام داده‌های اختصاص داده شده به خوشه را نشان می‌دهد. به عنوان فاصله‌سنج، فاصله اقلیدسی بین یک نقطه و یک نمونه استفاده می‌شود. در انتخاب مرکز خوشه همانطور که از نام الگوریتم پیداست از مقدار میانگین استفاده می‌شود. برای محاسبه مرکز خوشه، مجموع درجات عضویت هر عنصر بر توان m در خود بر حاصل ضرب درجات عضویت به توان m تقسیم می‌شود (Pedrycz, W. 2021).

مراحل بالا با استفاده از نرم افزار MATLAB نسخه ۲۰۱۷؛ و Unscrambler X نسخه ۱۰،۴ انجام شد.

نتایج و بحث

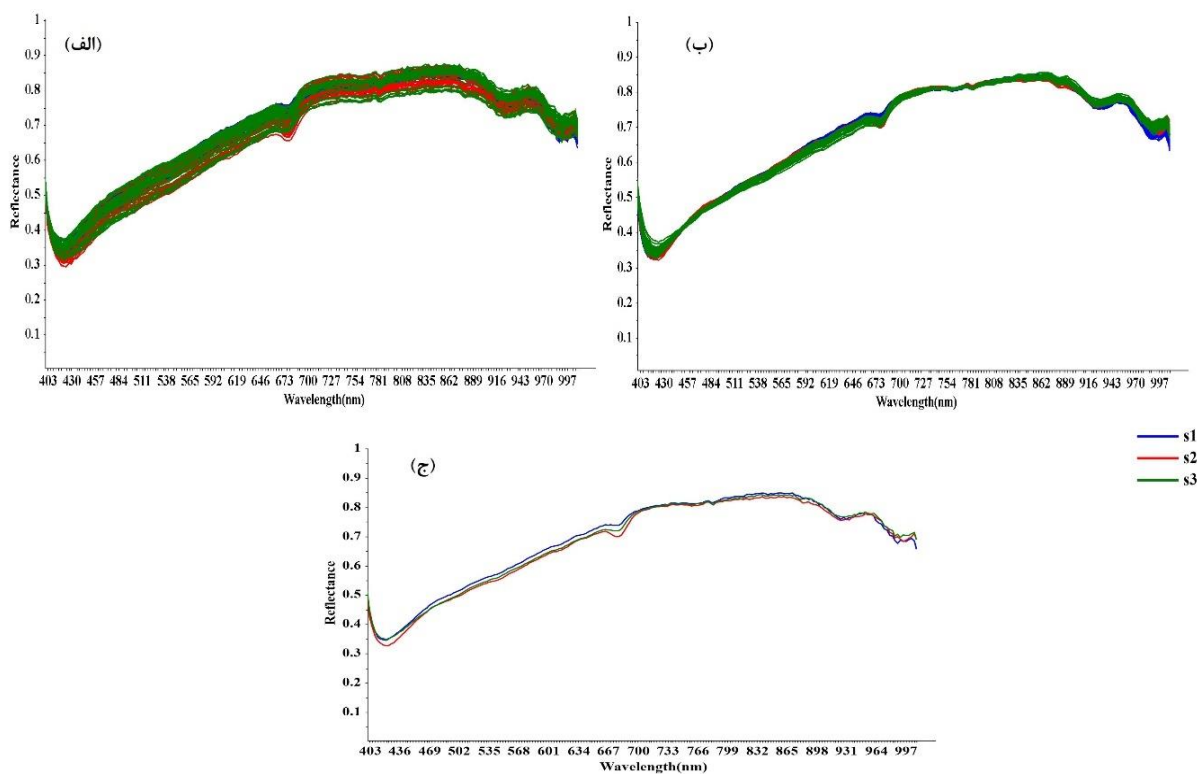
داده‌های طیفی

با استفاده از دوربین ابرطیفی در محدوده طیفی ۱۰۰۰-۴۰۰ نانومتر تمامی سه نمونه برنج هاشمی (S1)، شیرودی (S2) و ندا (S3) داده‌برداری طیفی شدند. سپس داده‌های طیفی هر سه نمونه به دست آمد. در شکل (۲) اطلاعات طیفی تمامی نمونه‌ها گزارش شده است. در شکل (۲الف) طیف خام سه رقم برنج نمایش داده شده است (۹۶۱ پیکسل در ۲۰۴ طول موج). در این شکل اطلاعاتی به دست نمی‌آید زیرا پدیده همپوشانی طیفی وجود دارد. بنابراین، در قسمت شکل ۲. ب) روش‌های پیش پردازش برای اصلاح پراکندگی نور و حذف شیب و وارونگی طیف‌ها بر روی داده‌های خام اعمال شد (۳۵ پیکسل در ۲۰۴ طول موج). در شکل ۲. ج) طیف میانگین هر سه نمونه با سه رنگ آبی، قرمز و سبز به ترتیب برای سه رقم هاشمی (S1)، شیرودی (S2) و ندا (S3) نمایش داده شده است. در ادامه از داده‌های اصلاح شده (پیش‌پردازش شده) برای نمایش PCA نمونه‌ها و مدل خوشه‌بندی فازی استفاده شد.

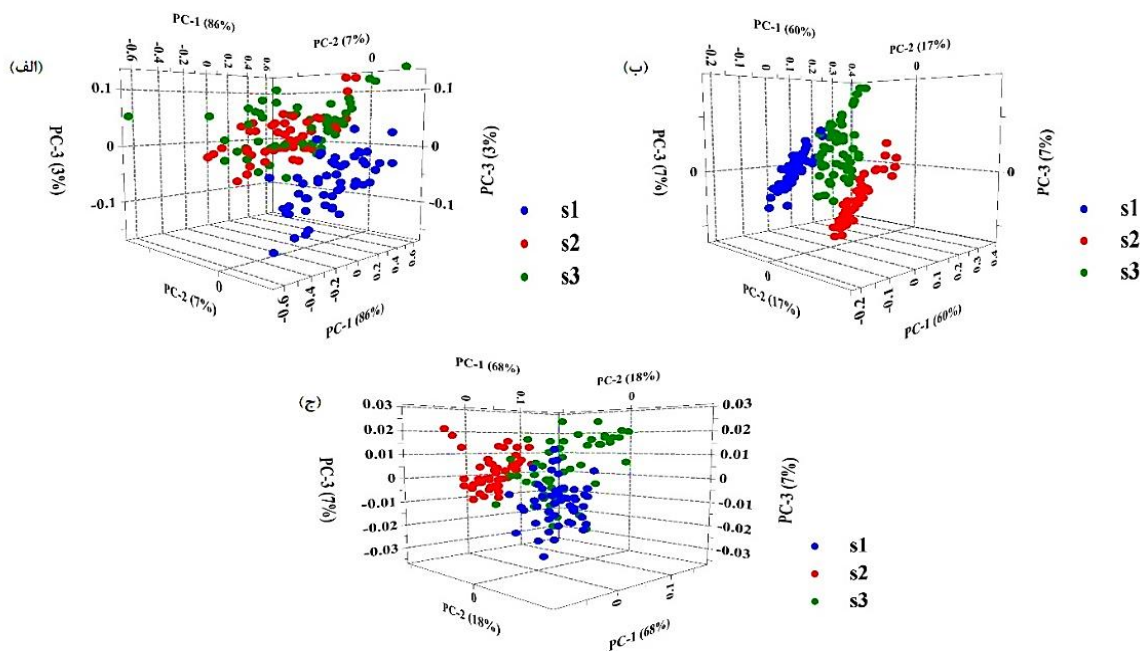
نتایج PCA

پس از پیش‌پردازش اولیه، از تحلیل PCA برای کاهش حجم ابعاد داده‌ها و همچنین آشکار کردن شباهت‌ها و تفاوت‌های ذاتی بین تمام نمونه‌ها استفاده شد (شکل ۳). همانطور که در شکل مورد نظر مشاهده می‌شود، نمایش داده‌ها از فضایی با ابعاد بالاتر به فضایی با ابعاد

کمر ارائه می‌شود. در شکل ۳. ب، پس از اعمال پیش‌پردازش MSC + SG، سه نمونه هاشمی، شیرودی، و ندا به خوبی از یکدیگر جدا شده و قابل شناسایی هستند. مقادیر PC1، PC2، و PC3 پس از پیش‌پردازش با MSC + SG برای ارقام هاشمی، شیرودی و ندا به ترتیب دارای ۶۰، ۱۷ و ۷ درصد تغییرات بین داده‌ها را پوشش داده است. در شکل ۳. ج با طول موج انتخابی (۴۷ عدد طول موج) نیز بر روی داده‌های پیش‌پردازش شده PCA اعمال شد، اما نتیجه‌ی بهتری نسبت به شکل ۳. ب دیده نشد و نتایج PCA در شکل ۳. ب کمی بهتر از نتایج شکل ۳. ج نمایان شد. در مرحله بعد به منظور کاهش حجم داده‌ها مولفه‌های اصلی یا PCs برای ایجاد مدل خوشه‌بندی فازی-C-means استفاده شد.



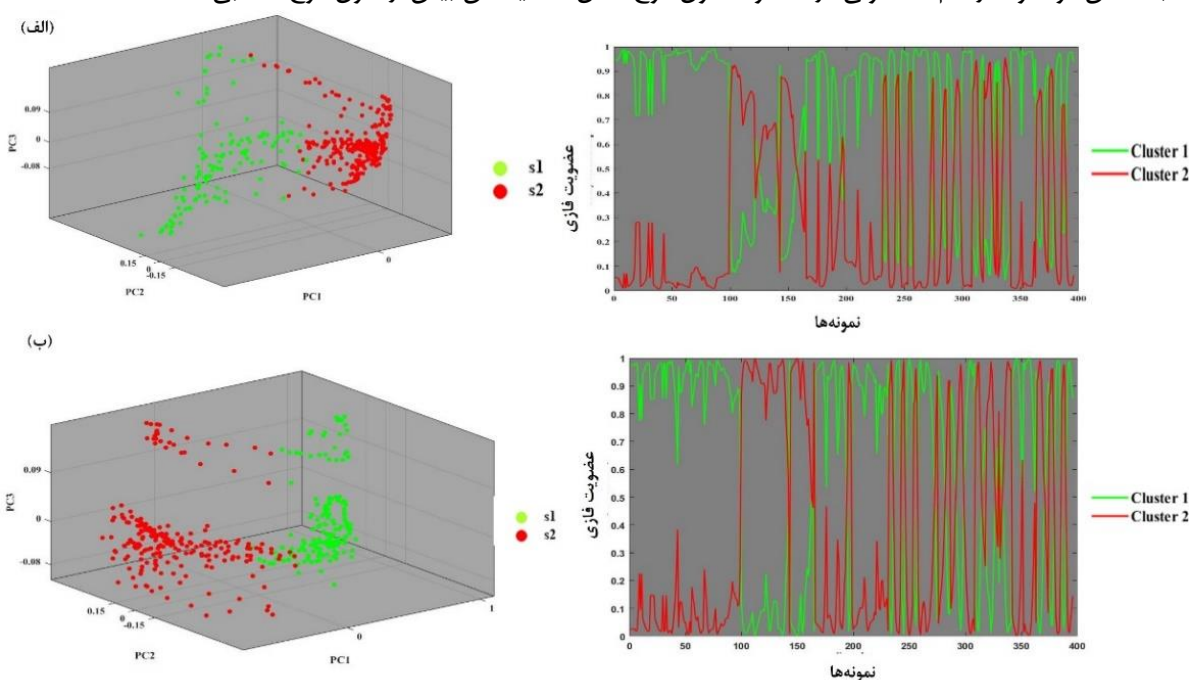
شکل ۲. الف) طیف خام و ب) پس از اعمال MSC+SG و ج) طیف متوسط برنج هاشمی، شیرودی و ندا.



شکل ۳. نمودارهای PCA الف) داده‌های خام و ب) داده‌های پیش‌پردازش شده MSC + SG و ج) MSC + SG با انتخاب طول موج برای برنج هاشمی، شیرودی و ندا

نتایج خوشه‌بندی به روش Fuzzy

دو نمونه اصلی و تقلبی با درصد اختلاط ۵ تا ۵۰ درصد در هر دو طول موج کامل و طول موج انتخابی در شکل‌های ۴ و ۵ به خوبی از هم جدا شده‌اند. برای تجزیه و تحلیل بهتر عضویت نمونه‌ها در خوشه‌ها، نمودار تکمیلی عضویت فازی نشان داده شده است (شکل الف-راست). در نمودار عضویت فازی نمونه اصلی برنج هاشمی (به رنگ سبز) و نمونه تقلبی ندا (به رنگ قرمز) قابل مشاهده است. همچنین این نتایج نشان می‌دهد که ۱۰۰ پیکسل اول نمایگر تعداد نمونه‌های اول است که متعلق به خوشه‌های اصلی و تقلبی با درجه عضویت بالا هستند (تقریباً کامل). این به این منظور است که ۳۵ پیکسل اول نمونه اصلی (هاشمی) و یا تقلبی (ندا) هستند که حاصل تصاویری است که ابتدا از اصل نمونه‌ها گرفته شده است و به همان صورت از ۵ درصد به بالا ادامه تصاویر گرفته شده است که به ترتیب ۳۵ پیکسل دوم و ۳۵ پیکسل سوم که به ترتیب ۵، ۱۰ درصد که تقلب در آن تشخیص داده نشده است و جزو نمونه‌های اصلی نمایش داده شده است. از ۱۰۰ پیکسل ۱۰ام به بعد (۴۰۰-۱۰۰) یعنی از تقلب ۱۵ درصد به بعد تقلب مشخص شده و دو نمونه هاشمی و ندا با هم مخلوط می‌شوند که نمودار عضویت فازی، ترکیب این دو نمونه (از ۱۵ درصد تا ۵۰ درصد) را نشان می‌دهد. در شکل ۴.الف و ۴.ب هر دو نتایج کامل نشان دهنده جدا شدن دو نمونه از هم است ولی در محدوده طول موج کامل شفافیت آن بیش از طول موج انتخابی است.

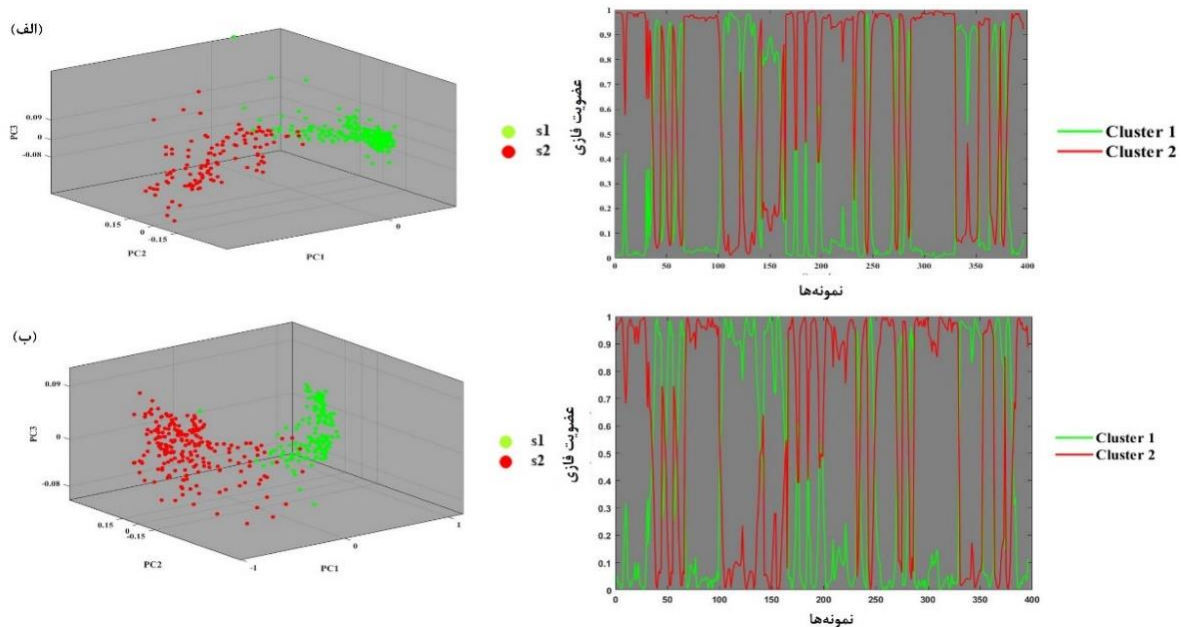


شکل ۴. خوشه‌بندی FCM و عضویت فازی نمونه‌های برنج هاشمی و ندا در محدوده (الف) طول موج کامل (ب) طول موج انتخابی

در نمودار عضویت فازی شکل (۵) نتایج جداسازی اختلاط برنج هاشمی و شیروودی نشان داده شده است تا ۵۰ پیکسل اول (تعداد نمونه) مربوط به دو نمونه مجزا و بعد از آن اختلاط دو نمونه را نشان می‌دهد. در واقع ۵ درصد را هم جزو نمونه‌های اصلی نشان می‌دهد و از ۱۰ درصد تقلب به بعد به خوبی اختلاط دو نمونه هاشمی و شیروودی نمایش داده شده است. در نتیجه نمودار فازی با طول موج کامل و انتخابی نمایش داده شده است و همچنین نمونه‌های اصلی و تقلبی از هم جدا شده و نمودار عضویت فازی جدا شدن نمونه‌ها و اختلاط نمونه‌ها را در شکل نشان می‌دهد.

هدف از بکارگیری سامانه HIS به همراه روش خوشه‌بندی فازی تعیین وجود تقلب یا عدم تقلب در نمونه هاشمی می‌باشد. در واقع با دو رنگ سبز و قرمز در نمودارهای فازی، نمونه اصلی (هاشمی) و تقلبی (ندا و شیروودی و ترکیب‌های آن با هاشمی) مشخص می‌شود. نمودار عضویت فازی تاییدی بر روش فازی مورد نظر است که به خوبی نمونه اصلی و تقلبی را جدا کرده است. در پژوهش‌های دیگر نیز با استفاده از داده‌های طیفی ۱۰۷۰-۷۴۰ نانومتر همراه با PLS-DA و مدل‌سازی مستقل نرم قیاس‌های طبقاتی (SIMCA)، نمونه‌های تقلبی و اصلی برنج مورد تجزیه و تحلیل قرار گرفت. ۲۰۰ نمونه برنج (۷۲ نمونه معتبر و ۱۲۸ نمونه تقلبی) مورد بررسی قرار گرفت. رویکرد PLS-DA نتایج رضایت‌بخشی را با نرخ کل طبقه‌بندی ۸۲/۶ درصد و ۸۲/۴ درصد برای نمونه‌های معتبر و تقلبی ارائه کرد، اما SIMCA

نتوانست مشکل بررسی شده را حل کند (Le Nguyen Doan et al., 2021). در مطالعه دیگری، از داده‌های طیفی ۱۷۰۰-۹۰۰ نانومتر و تحلیل PCA رگرسیون لجستیک^۱ (LR) برای کشف متغیر همبسته طیف در نمونه برنج و تعیین رابطه خطی تقلب استفاده شد. نمونه‌های برنج شامل ۳۱ نمونه برنج بدون تقلب و ۱۰ نمونه برنج تقلبی در ۳ مکان مختلف و در مجموع ۱۲۳ داده طیفی جمع آوری شدند. نتایج نشان داد که ترکیب PCA، NIRS و LR در تشخیص تقلب برنج با دقت ۹۴/۴ درصد در آزمون و ۹۹/۴ درصد در مجموعه آزمایشی ممکن و مؤثر است (Liew et al., 2020). در پژوهشی دیگر با استفاده از سامانه بینی الکترونیکی همراه با الگوریتم‌های بدون نظارت فازی، نقشه خودسازماندهی^۲ (SOM)، کلنی زنبورهای مصنوعی^۳ (ABC) و k-means برای مدل سازی تغییر عطر برنج در طول ذخیره سازی استفاده شد. نتایج نشان داد تقریباً تمام نمونه‌های معطر در گروه‌های مربوطه خود و برنج‌های غیر معطر، از آنجایی که بین گروه‌های ماه‌های نگهداری مجاور همپوشانی وجود داشت، برخی از نمونه‌های آن‌ها در خوشه‌های همسایه قرار گرفتند. در نتیجه سیستم بینی الکترونیکی همراه با الگوریتم‌های بدون نظارت به عنوان یک استراتژی موفق برای ارزیابی سریع و قابل اعتماد تغییر عطر برنج در طول ذخیره‌سازی موثر است (Rahimzadeh et al., 2022).



شکل ۵. خوشه‌بندی FCM و پارتیشن‌بندی فازی نمونه‌های برنج هاشمی و شیروودی در محدوده الف) طول موج کامل ب) طول موج انتخابی

تشخیص تقلب برنج‌هایی که از نظر ظاهری شبیه به هم هستند با چشم غیرمسلح به صورت دقیق امکان‌پذیر نبود، اما با کمک سامانه HSI و روش داده کاوی تجزیه و تحلیل فازی به راحتی قابل تشخیص بود. این سامانه به راحتی قابلیت استفاده برای اصلت‌سنجی برنج هاشمی در فروشگاه‌ها و مراکز توزیع در شمال کشور را دارد. همچنین با انجام مراحل کالیبراسیون می‌توان این سامانه را برای ارزیابی کیفی سایر نمونه‌های برنج اصلی و باکیفیت بکار برد. در خوشه‌بندی حاصل، همانطور که هر برنج طیف متفاوتی دارد، سیستم HIS به همراه روش‌های یادگیری ماشین توانستند با دو رنگ سبز و قرمز دو برنج اصلی و تقلبی را از هم جدا کند.

نتیجه‌گیری

در این پژوهش، یک سامانه شناسایی سریع و غیرمخرب مبتنی بر تصاویر ابرطیفی همراه با الگوریتم خوشه‌بندی فازی برای اصلت‌سنجی نمونه‌های برنج هاشمی توسعه داده شد. تمام طیف‌های نمونه برنج با دوربین HSI دریافت شد. مقایسه روش‌های مختلف پیش‌پردازش داده‌های طیفی، با استفاده از PCA نشان داد که روش MSC و Savitzky-Golay بهترین استراتژی پیش‌پردازش است که رقم‌های

برنج را از نظر بصری بهتر متمایز می‌کند. نتایج این تحقیق نشان داد که ترکیب روش PCA و روش خوشه‌بندی فازی همراه با سامانه HSI برای تشخیص تقلب در نمونه برنج هاشمی نتایج رضایت‌بخشی از خود نشان می‌دهد. تجزیه و تحلیل داده‌ها با استفاده از روش الگوریتم فازی با تمام طول موج‌ها و انتخاب طول موج‌های تکاملی انجام شد. نتایج نشان می‌دهد که سامانه HSI با الگوریتم غیر خطی فازی توسط نمودار عضویت فازی می‌تواند تقلب نمونه‌های برنج شیروودی و ندا را در نمونه‌های برنج هاشمی به خوبی پیش‌بینی کند. همچنین در تحقیقات آینده می‌توان از سامانه HSI به عنوان یک سامانه غیرمخرب و سریع با سایر روش‌های خوشه‌بندی شامل SOM، ABC و k-means، برای تشخیص تقلب و شناسایی برنج کهنه از برنج تازه استفاده کرد.

سپاسگزاری

نویسندگان این مقاله از دانشگاه شهرکرد، دانشگاه علوم کشاورزی و منابع طبیعی ساری، سازمان غذا و دارو و مرکز تحقیقات برنج کشور برای حمایت از این پروژه صمیمانه قدردانی می‌کنند.

منابع

فتحی، ناهید، نبی پور، علیرضا. (۱۳۹۸). روش‌های تعیین خلوص و کیفیت ارقام برنج. مجله تولیدات گیاهی، انتشارات پژوهشکده برنج کشور.

REFERENCES

- Abdi, H., & Williams, L. J. (2010). Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4), 433-459.
- Adán, J. G., & Wilfrido, G. F. (2015). Automatic clustering using natureinspired metaheuristics: A survey. *Appl. Soft Comput.* [http://dx. doi. org/10.1016/j. asoc, 1](http://dx.doi.org/10.1016/j.asoc.1).
- Aznan, A., Gonzalez Viejo, C., Pang, A., & Fuentes, S. (2022). Rapid detection of fraudulent rice using low-cost digital sensing devices and machine learning. *Sensors*, 22(22), 8655.
- Edris, M., Ghasemi-Varnamkhashti, M., Kiani, S., Yazdanpanah, H., & Izadi, Z. (2024). Identifying the authenticity and geographical origin of rice by analyzing hyperspectral images using unsupervised clustering algorithms. *Journal of Food Composition and Analysis*, 125, 105737.
- ElMasry, G., & Sun, D.-W. (2010). Principles of hyperspectral imaging technology. In *Hyperspectral imaging for food quality analysis and control* (pp. 3-43). Elsevier.
- Faqeerzada, M. A., Akter, T., Aline, U., Pahlawan, M. F. R., & Cho, B. K. (2023). Application of Hyperspectral Imaging for Rapid and Nondestructive Detection of Paraffine-Contaminated Rice. In *BIO Web of Conferences* (Vol. 80, p. 01001). EDP Sciences.
- Fathi, N., Nabipour, A. (2019). Methods of determining the purity and quality of rice varieties. *Journal of crop production, publications of the country's rice research institute* (in Persian).
- Ghasemi-Varnamkhashti, M., Mohtasebi, S. S., Rodriguez-Mendez, M. L., Lozano, J., Razavi, S. H., & Ahmadi, H. (2011). Potential application of electronic nose technology in brewery. *Trends in Food Science & Technology*, 22(4), 165-174.
- He, X., Feng, X., Sun, D., Liu, F., Bao, Y., & He, Y. (2019). Rapid and nondestructive measurement of rice seed vitality of different years using near-infrared hyperspectral imaging. *Molecules*, 24(12), 2227.
- Izadi, Z., & Kiani, S. Authenticity Identification of Pomegranate Molasses Using Hyperspectral Imaging System Coupled with Automatic Clustering by Artificial Bee Colony. *Available at SSRN 4423331*.
- Kheiralipour, K., & Jayas, D. S. (2024). Current and future applications of hyperspectral imaging in agriculture, nature and food. *Trends in Technical & Scientific Research*, 7(2), 1-9.
- Kiani, S., Azimifar, Z., & Kamgar, S. (2010). Wavelet-based crop detection and classification. In *2010 18th Iranian Conference on Electrical Engineering* (pp. 587-591). IEEE.
- Kiani, S., Van Ruth, S.M., Minaei, S. (2018). Hyperspectral imaging, a non-contact and non-destructive technique in aromatic/medicinal plant products industry: current status and potential future applications. *Computers and Electronics in Agriculture*, 152: 9-18.
- Le Nguyen Doan, D., Nguyen, Q. C., Marini, F., & Biancolillo, A. 2021. Authentication of rice (*Oryza sativa* L.) using near-infrared spectroscopy combined with different chemometric classification strategies. *Applied Sciences*, 11(1), 362.
- Li, T., Zhan, Z. H., Xu, J. C., Yang, Q., & Ma, Y. Y. (2022). A binary individual search strategy-based bi-objective evolutionary algorithm for high-dimensional feature selection. *Information Sciences*, 610, 651-673.



- Liew, K. T., Pui, L. P., & Solihin, M. I. (2020, December). Feasibility of fraud detection in rice using a handheld near-infrared spectroscopy. In *AIP Conference Proceedings* (Vol. 2306, No. 1). AIP Publishing.
- Liu, Y., Li, Y., Peng, Y., Yang, Y., & Wang, Q. (2020). Detection of fraud in high-quality rice by near-infrared spectroscopy. *Journal of food science*, 85(9), 2773-2782.
- McGrath, T. F., Shannon, M., Chevallier, O. P., Ch, R., Xu, F., Kong, F., ... & Elliott, C. T. (2021). Food Fingerprinting: Using a two-tiered approach to monitor and mitigate food fraud in rice. *Journal of AOAC International*, 104(1), 16-28.
- Miao, A., Zhuang, J., Tang, Y., He, Y., Chu, X., & Luo, S. (2018). Hyperspectral image based variety classification of waxy maize seeds by the t-SNE model and procrustes analysis. *Sensors*, 18(12), 4391.
- Nargesi, M. H., Kheiralipour, K., & Jayas, D. S. (2024). Classification of different wheat flour types using hyperspectral imaging and machine learning techniques. *Infrared Physics & Technology*, 142, 105520.
- Pedrycz, W. (2021). Fuzzy clustering. An Introduction to Computing with Fuzzy Sets: *Analysis, Design, and Applications*, 125-145.
- Rahimzadeh, H., Sadeghi, M., Mireei, S. A., & Ghasemi-Varnamkhasti, M. (2022). Unsupervised modelling of rice aroma change during ageing based on electronic nose coupled with bio-inspired algorithms. *Biosystems Engineering*, 216, 132-146.
- Sampaio, P. S., Soares, A., Castanho, A., Almeida, A. S., Oliveira, J., & Brites, C. (2018). Optimization of rice amylose determination by NIR-spectroscopy using PLS chemometrics algorithms. *Food Chemistry*, 242, 196-204.
- Sampaio, P. S., Castanho, A., Almeida, A. S., Oliveira, J., & Brites, C. (2020). Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods. *European food research and technology*, 246, 527-537.
- Savitzky, A., & Golay, M. J. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8), 1627-1639.
- Seo, Y., Lee, A., Kim, B., & Lim, J. (2020). Classification of rice and starch flours by using multiple hyperspectral imaging systems and chemometric methods. *Applied Sciences*, 10(19), 6724.
- Shannon, M., Ratnasekhar, C.H., McGrath, T.F., Kapil, A.P. and Elliott, C.T., (2021). A two-tiered system of analysis to tackle rice fraud: The Indian Basmati study. *Talanta*, 225, p.122038.
- Shaw, G., & Manolakis, D. (2002). Signal processing for hyperspectral image exploitation. *IEEE Signal processing magazine*, 19(1), 12-16.
- Siripatrawan, U., Makino, Y., Kawagoe, Y., & Oshita, S. (2011). Rapid detection of Escherichia coli contamination in packaged fresh spinach using hyperspectral imaging. *Talanta*, 85(1), 276-281.
- Teye, E., Amuah, C. L., McGrath, T., & Elliott, C. (2019). Innovative and rapid analysis for rice authenticity using hand-held NIR spectrometry and chemometrics. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 217, 147-154.
- Theodoridis, S., Pikrakis, A., Koutroumbas, K., & Cavouras, D. (2010). *Introduction to pattern recognition: a matlab approach*. Academic Press.
- Van Haute, Sam., Nikkhah, A., Malavi, D., Kiani, S. 2023. Prediction of essential oil content in spearmint (*Mentha spicata*) via near-infrared hyperspectral imaging and chemometrics. *Scientific Reports*, 13(1), 4261.
- Weng, S., Tang, P., Yuan, H., Guo, B., Yu, S., Huang, L., & Xu, C. (2020). Hyperspectral imaging for accurate determination of rice variety using a deep learning network with multi-feature fusion. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 234, 118237.