Iranian Food Science and Technology
Research Journal
Homepage: https://ifstrj.um.ac.ir

Research Article
Vol. 21, No. 3, Aug.-Sep. 2025, p. 261-269

Classification of Iranian Wheat Flour by FT-MIR Spectroscopy based on Max-Relevance Min-Redundancy Wavelength Selection Coupled with SVM

A. Kazemi[1*], A. Mahmoudi[1], S.H. Fattahi[2]

1- Department of Biosystems Engineering, University of Tabriz, Tabriz, Iran
(*- Corresponding Author Email: amirkazemi422@gmail.com)
2- Department of Biosystems Engineering, University of Maragheh, Maragheh, Iran

Received: 22.06.2024 Revised: 01.11.2024 Accepted: 20.11.2024 Available Online: 17.06.2025	**How to cite this article:** Kazemi A., Mahmoudi, A., & Fattahi, S.H. (2025). Classification of Iranian wheat flour by FT-MIR spectroscopy based on max-relevance min-redundancy wavelength selection coupled with SVM. *Iranian Food Science and Technology Research Journal, 21*(3), 261-269. https://doi.org/10.22067/ifstrj. 2024.88624.1342

Abstract

Different varieties of wheat as one of the strategic crops are cultivated in Iran based on the specific geographical and climatic conditions of each area. Classification of wheat varieties is important in order to guarantee the final products acquired from wheat flour. Fourier Transform-Mid Infrared (FT-MIR) spectroscopy as a nondestructive approach combined with chemometrics was employed to classify four varieties of Iranian wheat. 160 samples were analyzed and various preprocessing algorithms were used to correct unwanted information. Then, Principal Component Analysis (PCA) as unsupervised and Support Vector Machine (SVM) as supervised models with Max-Relevance Min-Redundancy (MRMR) feature selection algorithm were applied to investigate the classification of these varieties. The best result of SVM model without feature selection was with S-G+D2+MSC preprocessing with 99.4% of accuracy. The output of 100% with SVM model and MRMR feature selection algorithm confirmed the capability of FT-MIR spectroscopy method for classification of Iranian wheat flour varieties.

Keywords: Classification, FT-MIR spectroscopy, PCA, Preprocessing, Wheat flour

Introduction

Wheat stands out as the most commonly used cereal due to its high carbohydrate content, providing enough calories to satisfy both human and animal needs on a daily basis. Several factors contribute to the differences in bread wheat flour, such as the type of wheat variety, the growing conditions, and the environmental conditions. These factors vary between regions and from year to year, impacting the quality of flour. Therefore, if various types of wheat are intentionally combined, variations and inconsistencies can lead to variations in the quality of the resulting wheat flour and it is important to assess these factors before and during the trading process to ensure the quality of the wheat flour. Unfortunately, mixing varieties from different wheat classes can occasionally occur during storage, which cause potentially lower quality and is the primary concern for both buyers and sellers due to the fact that the product may not meet the specifications and expectations when delivered. Therefore, it is important to evaluate the uniformity and consistency of the wheat varieties before they are utilized in food processing facilities. Achieving the correct wheat quality will minimize the low crop output and boost production, that is the primary

https://doi.org/10.22067/ifstrj. 2024.88624.1342

objective for profitable wheat production. In Iran, the authentication of various wheat flour varieties, such as Baran, Homa, Sardari, Hashtrood, and Sadra, is a significant concern due to their popularity and productivity. The yield of such varieties ranging from 1200-1600 kg per hectare in the northwest region of the country (Khojastehnazhand & Roostaei, 2022).

Currently, various traditional techniques, including chemical and mechanical methods, were utilized for classification. However, these approaches had several drawbacks, such as being destructive, expensive, and time-consuming. As a result, there is a need for alternative methods that are non-destructive, cost-effective, and efficient. Vibrational spectroscopies such as Fourier Transform-Infrared (FT-IR) spectroscopy are commonly chosen for authenticating flour varieties due to their non-destructive nature, affordability, and user-friendly application. The FTIR method is a reliable tool for identifying biochemical fingerprints, allowing for the detection of trace compounds in intricate food compositions with subtle differences between closely related samples. Additionally, this technique provides precise measurement capabilities and efficient screening processes (Deniz *et al*., 2018; Ellis, Muhamadali, Haughey, Elliott, & Goodacre, 2015). In FTIR spectra, the positions of bands are associated with structural information, whereas the areas of the bands indicate the concentration of specific molecules. This information can be valuable for examining food samples based on their contents of proteins, carbohydrates, lipids, and nucleic acids (Deniz *et al*., 2018).

After data acquisition, it is essential to establish a link between the targeted sample characteristics and the IR absorption or transmittance measurements. Chemometrics is the keystone that utilizes spectral data alongside pattern recognition methods in order to efficiently tackle authentication challenges in products. The combination of FT-IR and chemometrics has been applied in numerous studies. For example, Wadood et al. (2019) utilized NIR spectroscopy combined with chemometric methods to classify wheat based on its production year, geographical origin, and genotypes. The results of LDA model indicated classification accuracies of 69% for production year, between 72.2% and 100% for geographical origin, and 69% for genotypes (Wadood, Guo, Zhang, & Wei, 2019).

In another study, De Girolma et al. used FT-NIR and FT-MIR techniques to evaluate identification of durum wheat pasta adulterated with common wheat. The results of LDA and PLS-DA models to classify samples had acceptable classification accuracy ranged from 80% to 95% for three groups, and from 91% to 97% for two groups (De Girolamo *et al*., 2020).

In another research, FT-IR spectroscopy combined with chemometrics were applied to detect the adulteration of blankit in wheat flour. By evaluating various preprocessing techniques with different models, the SVM model achieved the highest classification accuracy, attaining 100% for two classes (pure and adulterated) and 79.62% for five classes (pure and adulterated at various adulteration levels). This demonstrates the effectiveness of FT-MIR spectroscopy combined with chemometric methods for identifying fraud in wheat (Kazemi, Mahmoudi, & Khojastehnazhand, 2023).

The aim of this research is to classify four major varieties of Iranian wheat flour through Fourier Transform-Mid Infrared (FT-MIR) spectroscopy combined with chemometric analysis. In this study, the Max-Relevance Min-Redundancy (MRMR) algorithm was used to assess the performance of feature selection with SVM model to develop a robust classification method for Iranian wheat flour varieties. This methodology can enhance quality control, prevent mislabeling, and address food fraud, thus supporting the Iranian wheat industry in meeting stringent quality standards and reinforcing consumer confidence.

Materials and Methods

After preparing the sample classes and getting spectral data from each sample, various data preprocessing methods were applied to remove unwanted effects on acquired spectral data. Then firstly, PCA analysis was performed for clustering and SVM model without application of feature selection method was created. MRMR feature selection algorithms were then applied to compare the outcomes. The flowchart of steps for classification of Iranian wheat flour is shown in Fig. 1.

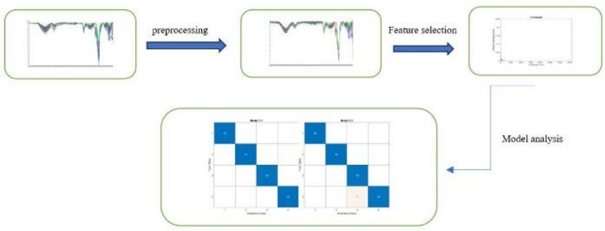

Fig. 1. The flowchart of wheat flour classification using FT-MIR spectroscopy approach

Data acquisition

After getting wheat seed varieties including (Baran, Hashtrood, Homa, Sadra) from dryland agricultural research institute in Maragheh, Iran, some initial processes including removing impurities from samples, milling the seeds with benchtop mill, and passing from sieve (420µm) were done to acquire pure flour samples of wheat. Then 40 samples of each variety were transferred to the laboratory in microtubes. All flour samples were then scanned by an FT-MIR spectrometer (Perkin Elmer, USA) equipped with an Attenuated Total Reflectance (ATR). Approximately 50mg of each sample was loaded on the ATR crystal and pressed until the signal intensity could be measured. The acquired spectra were in transmission mode and in the range of 400-4000 cm^{-1} with a resolution of 0.5 cm^{-1}.

Preprocessing of Spectral Data

Data were preprocessed before being explored or classified. Dealing with uninformative spectra due to light scattering or system noise is a key aspect of spectral data analysis (López-Maestresalas *et al.*, 2019). There are 2 primary categories of preprocessing methods in NIR spectroscopy: spectral normalization and spectral derivatives. Multiplicative Scattering Correction (MSC) and Standard Normal Variate (SNV) are the most common used algorithms used to correct the scatter effects from spectral data. Savitzky-Golay derivative is one of the most common algorithms which is used to smooth the spectra and eliminate the useless changes in signals (Holden, Wolfe, Ogejo, & Cummins, 2021; Sacré *et al.*, 2014). In this study, the combination of S-G smoothing (window size of 15 points), first and second derivatives, and MSC algorithms were applied. All of preprocessing methods were applied by Unscrambler X10.4 (Camo software, Oslo, Norway).

MRMR Feature Selection

Using all the different characteristics of a spectrum to create the model can lead to overfitting, which can make it harder for the model to apply to new data. Furthermore, in datasets with many variables, some of the features may be duplicate or provide redundant information. This indicates that some characteristics are closely related to other characteristics, leading to redundant information being supplied. This significantly increases the complexity of training the model and can lead to overfitting problems. To enhance the model's effectiveness, it is crucial to decrease the dimensionality of the data. Feature selection is one of the techniques that involves selecting a specific subset of unique and independent features to improve the modeling process by removing irrelevant elements from the original set. Maximum-Relevance Minimum-Redundancy (MRMR) is an algorithm that selects features by minimizing the correlation among variables to reduce redundancy in information. The main goal of MRMR is to find the most fitting group of attributes through maximizing correlation with target values and minimizing redundancy among variables (Ma, Chen, & Liu, 2024).

The MRMR feature selection method takes into account both the relationship between the input variable features and the target variable, as well as the redundancy between features, in order to select the most optimal subset of representative features. Before applying the learning algorithm, this method employs a filtering algorithm, like the MRMR algorithm, to choose a subset of features (Ma *et al*., 2024). By minimizing overlap between the new feature and the chosen subset, this method enhances the relationship between the new feature and the target variable (Ramírez-Gallego *et al*., 2017). The ultimate target is to screen and determine the most appropriate feature subset.

Mutual information (MI) is a core concept in information theory, that measures the dependency level between two random variables. After picking a specific subset S of features, we can measure the correlation between these features and the target value y using MI as:

$$I(S,y) = \frac{1}{|s|} \sum_{x_i \in S} I\ (x_i\,;y)$$

In the same way, within the unique subset S, the redundancy concept is defined as R by employing MI.

$$R(S) = \frac{1}{|s|^2} \sum_{x_i, x_j \in S} I(x_i\,, x_j)$$

The i_{th} feature variable within the feature subset S is represented by the variable x_i.

In selecting wavelengths for NIR spectral characteristics, the chosen wavelengths must adhere to two fundamental principles simultaneously. The goal is to increase the mutual information between the chosen feature subset S and the target value y, making sure that each feature in S has a meaningful impact on the target. Additionally, the aim is to decrease redundancy among the features in subset S in order to decrease correlation between model variables and improve the model's efficiency and performance. In order to simultaneously follow the constraints noted, it is possible to enhance the difference in importance between MI and redundancy R (Ma *et al*., 2024).

Modelling

The output data obtained from different preprocessing methods were applied to build classification models firstly without application of feature selection and then with application of feature selection algorithm. The applied classification models were unsupervised (PCA) and supervised (SVM). Among the unsupervised clustering methods available, including PCA, hierarchical cluster analysis (HCA), and K-means, PCA is often preferred for detecting potential data patterns through examination of variable similarities and differences (Granato, Santos, Escher, Ferreira, & Maggio, 2018). In the present investigation, PCA model was utilized on the FTIR-S-G-D2-MSC spectra to enhance data visualization within a transformed space. By utilizing PCA, the fundamental data is summarized into a score plot and loading plot derived from the original dataset. The score plot is considered as a useful tool for identifying patterns in highly correlated data, allowing us to observe how samples from similar groups tend to cluster together. Furthermore, the loading plot can uncover valuable insights about the original dataset by identifying which wavenumbers show the highest data variance and contribute most significantly to group clustering and separation. Therefore, by implementing a PCA model to reduce the original number of variables, the decision-making process for selecting classification algorithms in machine learning tests is streamlined and made more intuitive.

Support vector machine as one of the best techniques for classification, unlike clustering algorithms, is considered in the category of supervised learning and has two phases of training and testing. The result of this model is a representation of data in a multi-dimensional space where the data is divided in classes and the ranges between the data are determined by a hyperplane. The SVM utilizes a method known as kernel trick, including Linear, Polynomial, Radial Basis function, and Sigmoid for data conversion. In simpler terms, SVM uses intricate calculations to figure out how to differentiate the data according to specified labels or results (Khojastehnazhand & Roostaei, 2022). In the present research, all of modelling algorithms were applied by MATLAB R2017b (The MathWorks, MA, Natick, USA).

Results and Discussion

PCA

The acquired spectra of flour samples is shown in Fig. 2.

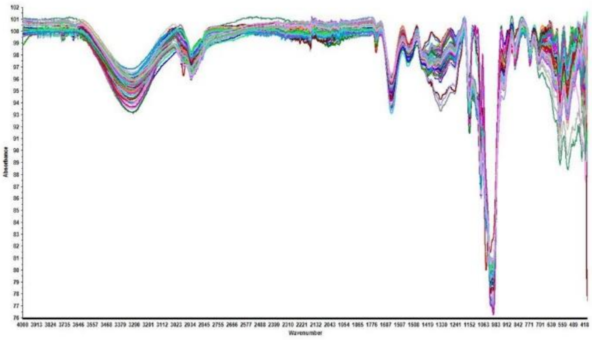

Fig. 2.The raw acquired spectra for flour varieties

PCA model was applied on the spectral data to observe any possible similarities and differences among four samples. PCA categorizes vast datasets with common identifying features in to clusters using an unsupervised approach. In PCA interpretation, it is typical for each separate cluster to contain samples with the highest spectral similarities that mirror composition similarities (Keshavarzi, Barzegari Banadkoki, Faizi, Zolghadri, & Shirazi, 2020). Fig.3.a, illustrates the score plot of flour samples. PC1, PC2, and PC3 account for 53%, 21%, and 7% of the overall variances, respectively. The explained-variance plot Fig. 3.b, gives an indication of how much of the variation in the data is described by the different components. According to Fig. 3.b, the optimal value of the factor for the model is 7, and factors number first and second had more ability to describe independence variables, among which the first factor had the largest contribution in describing the data. From the fourth factor onwards, the increase in the number of main components was ineffective and may decrease the validity of the model. The classification of flour classes was acceptable according to the obtained score plot. Various classes were identified separately and clustered together. However, some samples from Homa and Baran varieties showed overlap most likely due to the similarities of their composition. In addition, the loading plot of NIR data were obtained to determine optimal wavelengths. Fig. 3.c, shows the obtained loading plot for the first three PCs. The regression coefficients of each wavelength at each PC in PCA loadings provide insight in to which wavelengths have the greatest impact on the discrimination (Barbin, Badaro, Honorato, Ida, & Shimokomaki, 2020). When there is a larger absolute loading value for a certain wavelength region, the connection to PC becomes stronger. The peaks at 3350, 950, and 530 cm^{-1} related to the fats, unsaturated bonds C=C connected to the oxygen atoms (Coţovanu & Mironeasa, 2022).

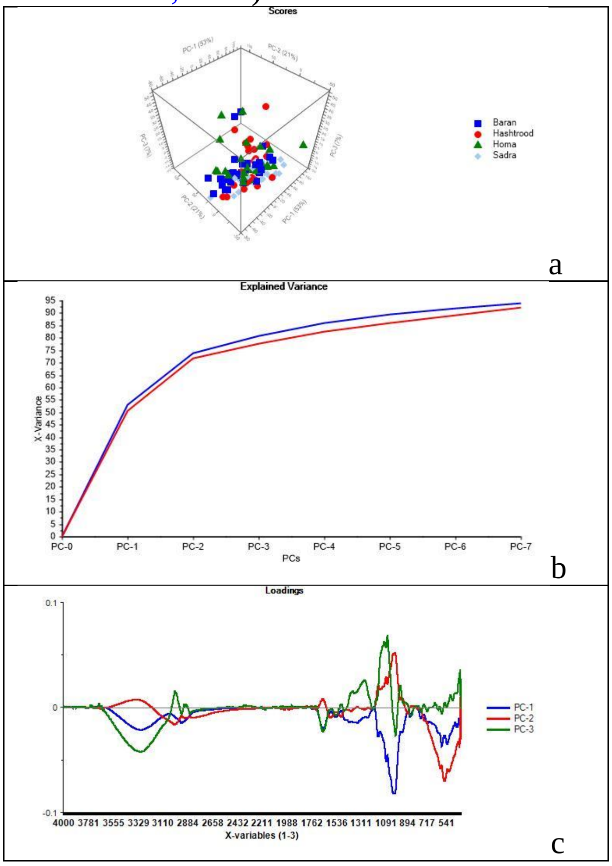

Fig. 3. a) The score plot of PCA model b) The explained-variance plot of PCA model c) The loading plot of PCA model

SVM Model

SVM, an algorithm rooted in statistical learning theory, is recognized for its ability to classify samples with its nonlinear computational approach. SVM model with 70% of data for training and 30% for test was employed for the obtained data with various preprocessing methods. The application of SVM model on the spectral data for flour variety classification yielded promising results, particularly after the preprocessed data with S-G + D2 + MSC method. The combination of pretreatment algorithms played a significant role in enhancing the discriminative capabilities of SVM model, S-G filtering helps in reducing noise and smoothing the spectral data, thereby improving the signal-to-noise ratio. The application of the second or first derivative enhances the spectral features by highlighting variations in the spectral curve, making subtle differences between flour varieties more pronounced. Additionally, MSC preprocessing addresses multiplicative effects such as baseline shifts and scaling differences, thereby normalizing the spectral data and mitigating variability related to instrumental differences or sample presentation. As it is observable from Table.1, the result of MSC preprocessing was weaker than D1 or D2. The choice of preprocessing method can affect the performance of the model trained on the preprocessed data. If the model is better suited to the features generated by derivative preprocessing, it might perform better compared to when trained on features generated by MSC preprocessing. Furthermore, certain preprocessing techniques may be better suited to handling specific types of data. In the spectral data of present research, the multiplicative effect was weaker than noises and the peaks of data had to be clearer by derivative algorithms. According to Table.1, the application of SVM model on the spectral data for classification of flour varieties had yielded acceptable results, particularly with the application of various kernel functions. In this study, we explored the performance of SVM model applying linear, quadratic, and cubic kernels. The outcomes indicated that the linear kernel outperformed other kernels, achieving the accuracy of 99.4%. Linear kernels establish decision boundaries that are linear hyperplanes in the feature space. The reason for better performance of linear kernel might be the less complexity of dataset. Furthermore, quadratic and cubic kernels introduce higher degrees of non-linearity in to the decision boundary, which can lead to overfitting, particularly when dealing with high-dimensional data and the risk of overfitting is less with a linear kernel, as it imposes a simpler, more constrained decision boundary among classes. In another study, Mohamed *et al.* used combined handheld spectrometer and SVM model to classify three different types of flour (whole wheat, organic wheat, and rice flour). The outcome of 100% for accuracy was acquired with the proper preprocessing algorithm (Mohamed, Solihin, Astuti, Ang, & Zailah, 2019). Therefore, the result of present study was similar to the result of that study, however, the classification in our study was between 4 flour varieties and a little more difficult than 3 classes. In another research, Sampaio *et al.* applied near-infrared (NIR) spectroscopy associated to SVM model created after MSC preprocessing showed the result of 91% accuracy for prediction dataset which was weaker than the outcomes of present study (Sampaio, Castanho, Almeida, Oliveira, & Brites, 2020). In order to decrease and select important features in the dataset, MRMR feature selection algorithm was also applied to illustrate the effects of each feature on the classification results. Removing of irrelevant features can lead to improved classification results. Furthermore, with utilizing just a fraction of the available features will greatly enhance the systems speed (Karimi, Kondrood, & Alizadeh, 2017; Khojastehnazhand & Roostaei, 2022). The acquired plot of MRMR feature selection algorithm is shown in fig.4, the threshold of important feature was selected 250 cm^{-1} by trial and error.

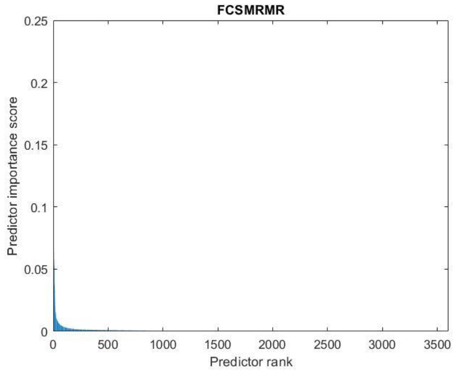

Fig. 4. The plot of MRMR feature selection algorithm

Table 1, shows the result of SVM model without application of feature selection and with feature selection algorithm selecting 250 important features. As shown in the table, SVM model was improved in S-G and S-G+MSC preprocessing methods. The accuracy of S-G+D2+MSC was improved from 99.4% to 100% and was selected as the best model. However, the feature selection algorithm did not have especial effect of S-G+D1, S-G+D2 models. In a similar research Khojastehnezhand and Roostaei applied this algorithm to the dataset of wheat seeds classification by computer vision technique. The SVM model had not improved in any of the number of features used (Khojastehnazhand & Roostaei, 2022). For further studies it is suggested to explore the effects of other feature selection algorithms on the machine learning algorithms like SVM.

Conclusion

The classification of wheat flour was explored with the combination of Ft-MIR spectroscopy and chemometrics techniques. PCA model as unsupervised method was used to reduce dimensionality of data and explore similarities of samples of classes. SVM algorithm as supervised model was applied to investigate classification. The best outcome was with dataset with S-G+D2+MSC preprocessing with 99.4% of accuracy. Then with the application of MRMR feature selection method the accuracy improved to 100%. Therefore, the applicability of FT-MIR spectroscopy and machine learning algorithms for classification of Iranian wheat flour as one of the strategic crops was approved.

Table 1- The results of SVM model for different preprocessing methods

Model	SVM			MRMR + SVM		
Kernel function	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
S-G	78.6	78	74.8	83	81.1	80.5
S-G+MSC	75.5	73.6	73	78.6	76.7	76.7
S-G+D1	99.4	99.4	99.4	99.4	99.4	99.4
S-G+D2	99.4	99.4	99.4	99.4	99.4	99.4
S-G+D1+MSC	99.4	99.4	99.4	99.4	99.4	99.4
S-G+D2+MSC	99.4	99.4	99.4	100	100	100

Author Contributions

Amir Kazemi: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Data curation, Writing–review & editing, Project administration, Writing-Original Draf, **Asghar Mahmoudi:** Software, supervision, validation, writing-reviewing, and editing.

Founding Source

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

References

1. Barbin, D.F., Badaro, A.T., Honorato, D.C., Ida, E.Y., & Shimokomaki, M. (2020). Identification of turkey meat and processed products using near infrared spectroscopy. *Food Control, 107*, 106816. https://doi.org/10.1016/j.foodcont.2019.106816
2. Coţovanu, I., & Mironeasa, S. (2022). Influence of buckwheat seed fractions on dough and baking performance of wheat bread. *Agronomy, 12*(1), 137. https://doi.org/10.3390/agronomy12010137
3. De Girolamo, A., Arroyo, M.C., Cervellieri, S., Cortese, M., Pascale, M., Logrieco, A.F., & Lippolis, V. (2020). Detection of durum wheat pasta adulteration with common wheat by infrared spectroscopy and chemometrics: A case study. *LWT, 127*, 109368. https://doi.org/10.1016/j.lwt.2020.109368
4. Deniz, E., Güneş Altuntaş, E., Ayhan, B., İğci, N., Özel Demiralp, D., & Candoğan, K. (2018). Differentiation of beef mixtures adulterated with chicken or turkey meat using FTIR spectroscopy. *Journal of Food Processing and Preservation, 42*(10), e13767. https://doi.org/10.1111/jfpp.13767
5. Ellis, D.I., Muhamadali, H., Haughey, S.A., Elliott, C.T., & Goodacre, R. (2015). Point-and-shoot: rapid quantitative detection methods for on-site food fraud analysis–moving out of the laboratory and into the food supply chain. *Analytical Methods, 7*(22), 9401-9414.

https://doi.org/10.1039/C5AY02048D
6. Granato, D., Santos, J.S., Escher, G.B., Ferreira, B.L., & Maggio, R.M. (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. https://doi.org/10.1016/j.tifs.2017.12.006
7. Holden, N.M., Wolfe, M.L., Ogejo, J.A., & Cummins, E.J. (2021). Introduction to biosystems engineering *Introduction to Biosystems Engineering* (pp. 0): American Society of Agricultural and Biological Engineers. https://doi.org/10.21061/intro2biosystemsengineering
8. Karimi, N., Kondrood, R.R., & Alizadeh, T. (2017). An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms. *Measurement*, *107*, 68-76. https://doi.org/10.1016/j.measurement.2017.05.009
9. Kazemi, A., Mahmoudi, A., & Khojastehnazhand, M. (2023). Detection of sodium hydrosulfite adulteration in wheat flour by FT-MIR spectroscopy. *Journal of Food Measurement and Characterization*, *17*(2), 1932-1939. https://doi.org/10.1007/s11694-022-01763-x
10. Keshavarzi, Z., Barzegari Banadkoki, S., Faizi, M., Zolghadri, Y., & Shirazi, F.H. (2020). Comparison of transmission FTIR and ATR spectra for discrimination between beef and chicken meat and quantification of chicken in beef meat mixture using ATR-FTIR combined with chemometrics. *Journal of Food Science and Technology*, *57*, 1430-1438. https://doi.org/10.1007/s13197-019-04178-7
11. Khojastehnazhand, M., & Roostaei, M. (2022). Classification of seven Iranian wheat varieties using texture features. *Expert Systems with Applications*, *199*, 117014. https://doi.org/10.1016/j.eswa.2022.117014
12. López-Maestresalas, A., Insausti, K., Jarén, C., Pérez-Roncal, C., Urrutia, O., Beriain, M.J., & Arazuri, S. (2019). Detection of minced lamb and beef fraud using NIR spectroscopy. *Food Control*, *98*, 465-473. https://doi.org/10.1016/j.foodcont.2018.12.003
13. Ma, X.-H., Chen, Z.-G., & Liu, J.-M. (2024). Wavelength selection method for near-infrared spectroscopy based on Max-Relevance Min-Redundancy. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *310*, 123933. https://doi.org/10.1016/j.saa.2024.123933
14. Mohamed, M.Y., Solihin, M.I., Astuti, W., Ang, C.K., & Zailah, W. (2019). *Food powders classification using handheld near-infrared spectroscopy and support vector machine*. Paper presented at the Journal of Physics: Conference Series. https://doi.org/10.1088/1742-6596/1367/1/012029
15. Ramírez-Gallego, S., Lastra, I., Martínez-Rego, D., Bolón-Canedo, V., Benítez, J.M., Herrera, F., & Alonso-Betanzos, A. (2017). Fast-mRMR: Fast minimum redundancy maximum relevance algorithm for high-dimensional big data. *International Journal of Intelligent Systems*, *32*(2), 134-152. https://doi.org/10.1002/int.21833
16. Sacré, P.-Y., De Bleye, C., Chavez, P.-F., Netchacovitch, L., Hubert, P., & Ziemons, E. (2014). Data processing of vibrational chemical imaging for pharmaceutical applications. *Journal of Pharmaceutical and Biomedical Analysis*, *101*, 123-140. https://doi.org/10.1016/j.jpba.2014.04.012
17. 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. https://doi.org/10.1007/s00217-019-03419-5
18. Wadood, S.A., Guo, B., Zhang, X., & Wei, Y. (2019). Geographical origin discrimination of wheat kernel and white flour using near-infrared reflectance spectroscopy fingerprinting coupled with chemometrics. *International Journal of Food Science & Technology*, *54*(6), 2045-2054. https://doi.org/10.1111/ijfs.14105

نشریه پژوهشهای علوم و صنایع غذایی ایران
https://ifstrj.um.ac.ir

مقاله پژوهشی
جلد ۲۱، شماره ۳، مرداد- شهریور ۱٤۰٤، ص. ۲٦۹-۲٦۱

طبقه‌بندی آرد گندم ایرانی با استفاده از طیف‌سنجی FT-MIR بر پایه‌ی انتخاب طول موج با بیشینه‌ی ارتباط و کمینه‌ی افزونگی، همراه با ماشین بردار پشتیبان SVM

امیر کاظمی[1*] - اصغر محمودی[1] - سید حسین فتاحی[2]

تاریخ دریافت: ۱۴۰۳/۰۴/۰۲
تاریخ پذیرش: ۱۴۰۳/۰۸/۳۰

چکیده

انواع واریته گندم، به‌عنوان یکی از محصولات راهبردی، در ایران بر اساس شرایط خاص جغرافیایی و اقلیمی هر منطقه کشت می‌شوند. طبقه‌بندی این واریته‌های گندم برای تضمین کیفیت محصولات نهایی حاصل از آرد گندم اهمیت دارد. در این پژوهش، از طیف‌سنجی مادون قرمز میان‌ناحیه با تبدیل فوریه (FT-MIR) به‌عنوان روشی غیرمخرب، همراه با شیمی‌سنجی، برای طبقه‌بندی چهار رقم از گندم ایرانی استفاده شد. در مجموع ۱۶۰ نمونه مورد تحلیل قرار گرفت و از الگوریتم‌های مختلف پیش‌پردازش برای حذف اطلاعات ناخواسته بهره گرفته شد. سپس، از تحلیل مؤلفه‌های اصلی (PCA) به‌عنوان مدل بدون ناظر و ماشین بردار پشتیبان (SVM) به‌عنوان مدل با ناظر، همراه با الگوریتم انتخاب ویژگی با بیشینه‌ی ارتباط و کمینه‌ی افزونگی (MRMR)، برای بررسی رده‌بندی این گونه‌ها استفاده شد. بهترین نتیجه مدل SVM بدون انتخاب ویژگی، با پیش‌پردازش ترکیبیS-G+D2+MSC ، دقتی برابر با ۹۹/۴ درصد به‌دست آورد. خروجی ۱۰۰ درصد حاصل از مدل SVM همراه با الگوریتم انتخاب ویژگی MRMR، توانمندی روش طیف‌سنجی FT-MIRرا در رده‌بندی گونه‌های آرد گندم ایرانی تأیید کرد.

واژه‌های کلیدی: آرد گندم، پیش‌پردازش، طبقه بندی، طیف‌سنجی تبدیل فوریه

۱- گروه مهندسی بیوسیستم، دانشگاه تبریز، تبریز، ایران
(نویسنده مسئول: Email: a.kazemi@tabrizu.ac.ir)
۲- گروه مهندسی بیوسیستم، دانشکده کشاورزی، دانشگاه مراغه، مراغه، ایران

https://doi.org/10.22067/ifstrj. 2024.88624.1342