Random Forest (RF) based identification of rice powder mixture using terahertz spectroscopy

J Friska¹, M Navaneetha Velammal², A Rajeshwari³, and P Hannah Blessy⁴

¹²Associate professor, Francis Xavier Engineering College-Tirunelveli 627003, India,
³⁴Student (PG-Scholar), Francis Xavier Engineering College-Tirunelveli 627003, India.

friskajoseph@gmail.com, navaneethavelammal.m@francisxavier.ac.in,
rajarjun271997@gmail.com, hannahblessyp@gmail.com

Abstract. Adulteration is a severe problem in agriculture field. It may cause some dishonest traders in the marketing side. For their own wish, dishonest traders make adulteration products for higher profit. In consumer side, agro product adulterations provide some unhealthy situation. It may cause severe health problem and also some incurable disease. Its right time, need to find an adulteration various level of agro products. For detecting rice adulteration, to collect spectral data for mixed rice powder (low quality rice powder and high quality rice powder), and a Fast Independent Component Analysis (FICA) algorithm is used to extract valuable information. Then, apply Random Forest Classifier algorithm for classifying purpose, whether the mixture is low quality or high quality. It achieves high accuracy with a single model and also easy to implement.

Keywords - Adulteration, Fast Independent Component Analysis, Random Forest Classifier, terahertz spectroscopy

1. Introduction

Rice is one of the important agro products in the world. Most of the peoples are consuming rice without known their quality of rice. Some peoples are consuming rice as a raw form and other peoples are consuming as a rice as powder form for preparing own food. Here the problem is, where the peoples are consuming rice in powder form really don’t know about the quality of rice powder. All the rice powders are in white color i.e. low quality rice powder and high quality rice powder. It’s very hard to finding adulteration rice in powder form really don’t know about the quality of rice powder. All the rice powders are in white color i.e. low quality rice powder and high quality rice powder. It’s very hard to finding adulteration rice in powder form, these kinds of adulterations not visible in eyes. Contamination in food, both commercially and extra income, is unhealthy and life threatening. Rice is an unspeakable ingredient in everyday foods. This creates an environment where many people get sick. Thus, consumers are failing in many places for finding adulterated product. Product managers and sellers of various products accomplish this by deceiving people into selling their substandard, counterfeit food by giving what they owe to authorities. Some dishonest traders are knowingly selling their adulterated product for harm the physical health of the people. In our daily lives, a variety of additives are added to the food items buy in stores that cannot find.

This admixture is performed in such a way that cannot even tell the difference between the good product and the adulterated product. Unknowingly pay for it, eat it, and suffer from various form of disease. Currently mixing ‘plastic’ rice with rice and fake eggs is painful. The fact is that the culture of the mixture does not change, except that the substance changes with the passage of time. Can food be
adulterated like this? That is shockingly adulterated without changing the appearance of the products. In this methodology really helpful for the rice powder mixed with their low quality rice powder. Foods that rely on to buy in stores contain a variety of additives that cannot find with our eyes. How do these traders, who play with people's lives to make money in the wrong way. In this way, it is painful for a few to see money being stuffed into products for profit. Food adulteration did not come yesterday, today; it has been going on for many years. Starting with rice they mix and sell substandard rice, small grains, soil, paddy and bran with rice. The advent of plastic foods in this age of technology has also caused great fear. These plastics, even if they are in the soil, can cause insomnia, disrupting the movement of the esophagus and up to the digestive organs, causing life-threatening atrocities. This can lead to abdominal pain, indigestion, vomiting and, if left untreated, can lead to cancer, kidney disease, heart attack, and vascular rupture.

In early stage, adulteration of rice powder identification is based on color and texture. This method is one of the traditional methods for identification of rice mixture. Based on this methodology couldn't find exact adulteration ratio of rice powder. This method, find whether the rice to be adulterated or not. The environmental factors also affect the results of adulteration in rice powder. After so many years, chemical methods were used to find adulteration of rice powder. This method involves complex pre-treatment and also expensive equipment. It also couldn't find exact adulteration ratio of mixed rice powder. It's very hard to implement, and also equipment setup cost is too high. Chemical method is very complex method and lengthy process for finding adulteration ratio of mixed rice powder.

Now a day, Spectral technology used to find adulterated agro products. This methodology is very popular for finding adulterated products. This spectroscopy method predicts the adulterated level of mixed rice powder based on surface liquid content, along with near-infrared (NIR) spectroscopy [1]. Providing better accuracy NIR combined with a Back propagation neural network was used. However, this method doesn't suits for long wavelength band spectrum for rice powder. In this method, long wavelength spectrum is still unknown. In this spectroscopy method the information found from spectral response in terahertz band. This terahertz band information can't obtain from other traditional method. Thus, the terahertz time-domain spectroscopy method is effective tool for finding adulteration in various levels.

2. Related work
Wenwen, et.al. (2013) has proposed to identify the rice seed cultivars using NIR spectroscopy. In this study[16], NIR spectral data are used as input. Partial Least Squares Discriminant Analysis (PLS-DA) algorithm used to take out the information for classification purpose. Here, three classifier are used and provide the comparable result one with another. PLS-DA algorithm with KNN classifier gives 80% accuracy. In another side, SIMCA with Random forest provides the 100% accuracy. Anyway it’s not suitable for longer wavelength spectrum.

Jianjun, et.al. (2014) has proposed to identify the transgenic cotton seed based on GA-SVM. In this study, [17]Terahertz spectral data are used as an input. Principle component analysis is used for extracting corresponding feature. Here, SVM classifier is used for classification purpose. By using this method the rate of accuracy will be 96.67%.For improving accuracy, PCA is replaced by FICA.
Saritha, et al. (2017) has proposed to identify the adulteration of papaya seed in black pepper. In this study [6] the real-time images are captured by camera. After that, pre-processing is used for removing unwanted noise from input images. Some features are extracted for finding accurate classification results. In this study, for the feature selection purpose, Contrast, texture, shape features are used. Finally, KNN classifier is used for classifying the result of adulteration in black pepper. In this method, the accuracy rate will be 90%.

3. System implementation

In this study, there are four steps [1] involved for finding various levels of adulteration in mixed rice powder.

Step 1 - Collecting spectral data
- Different levels of spectral data to be collected for mixed rice powder.

Step 2 – Spectral Pretreatment & FICA
- Spectral pretreatment used to reduce the dimensions of THz spectral data
- FICA (Fast Independent Component Analysis) used for grouping the component values where the values are close to requirements for adulterating various levels of spectral samples.

Step 3 – Feature extraction
- In this module, two features are used. That is spectral and temporal features
  - In spectral feature analysis based on frequency domain
  - In temporal feature analysis based on time domain
  - Both the feature values are calculated for finding various levels of adulteration in rice samples in different ratios.

Step 4 – Random Forest Classifier
- In this study, random forest classifier is used for classification purposes. Based on the above calculated values, the samples are already well trained.
- After that, the testing spectral samples are compared with their trained samples.

These four steps are important for calculating the various levels of adulteration in rice powder. It’s a very effective method for finding adulterating different kinds of mixed rice powder.

3.1 Spectral data collection

In this study, standard database images are used under some standard criteria [1]. Before capturing terahertz data, various rice samples are ground into powder form.

![Figure 1. Architecture of proposed system.](image)

Then different samples of rice powder are mixed with each other. In this study, there are five adulterated samples used in various mixed ratios. That adulterated rice powders are in small circular slices under some pressure on it. After the setup arrangement, terahertz spectral data are captured. In this study, standard database images (Figure 2) are used for the analysis of finding the adulteration of rice powder.
3.2 Fast independent component analysis

In existing system Principle compound analysis used for extracting feature value with clustering based on their requirement. Using PCA doesn’t provide the exact principle compound value. So, in this study Fast Independent compound analysis used for extracting feature. It provides the exact component value for each clustering component. And also it’s quite faster than PCA.

3.3 Feature extraction

The feature extraction step function to discover the various features that represent best adulteration level of mixed rice powder. Here, various feature techniques used such as temporal and spectral.

3.3.1 Temporal feature

In feature extraction, temporal feature is one of the important features. Most of the real time images are taken at the different time. Correlations among the images are used to calculate the every change of images in dataset. Various temporal features are there for analyzing images in dataset. Temporal features are used to extract the valuable information such as amplitude, energy based on time domain analysis. This feature calculates the various images in dataset whether it’s associate with time or changes over the time. In this study, the temporal features such as mean, Variance, Skewness are considered.

3.3.2 Spectral Feature

Spectral feature also very important feature for extracting valuable information from dataset. This feature provides the frequency domain metrics for each image in dataset. It’s easy to establish whether the data value in power spectrum value or spectral value. Here some spectral features are spectral centroid, spectral spread, spectral flux, spectral flatness measure and Chroma. In this Study, the Spectral features such as centroid, Spectral Variance considered.

3.4 Random forest classifier

In this study, random forest classifier is used for classifying various level of adulteration in rice mixture. Random forest, works under tree like structure.

Figure 2. Sample rice powder mixture in 2:0.

Figure 3. Random forest tree structure.
In this classifier, multiple trees form a forest. Here, each tree of forest provides the votes for corresponding input criteria. Random forest classifier method is a fast and also provides accurate result based on their training sample data.

3.4.1 Steps involved in Random forest (Refer Figure 3)
- Let the number of already trained sample A.
- Let the number of testing sample B.
- Each and every iteration, the testing samples are compared with trained samples of classifier.
- Each tree structure of the forest gives a vote for every iteration $K_1, K_2, K_3$.
- Finally, All the iteration value ($K_1$, $K_2$, $K_3$...) makes accurate result of classification value $K$
- Repeat the step from 2 to step 5 for getting result for various testing samples.

4. Simulation results & discussion
In this study, the rice spectrum images are collected from the standard database. The implementation starts with acquiring the database images. Next, FICA was done for clustering with their requirement samples. Features are subsequently extracted from the image which is used for training and testing. In this study, the temporal and spectral features are extracted for finding the adulteration of rice in various levels.

The correct grading rate of 30 spectrum images was tested by using random forest classifier. After testing 30 rice spectrum images, in which 29 rice spectrum gave a positive result but remaining one image gave a negative result for finding the adulteration level of rice. The rice spectrum images are collected from the standard database. In Figure 4 represents the THz spectral data for mixed rice powder in 3:0 ratios.

![Figure 4. THz Spectra of selected Rice band.](image)

In Figure 5 shows the results of reference spectrum for various mixed level of rice powder mixture, based on these reference spectra found adulterated level of various rice powder mixtures in ratio form like 1:1, 1:2, 2:1, 0:3, 3:0. In these reference rice spectra says about classifying main categories of five different ratios. In Figure 6 shows the estimated graph analysis for FICA with their exact cluster value. This estimated value used for future purpose of feature extraction of given spectrum. In feature extraction, Spectral and temporal feature values are calculated for classifying different kind of mixed rice powder in ratios. In FICA the values are started from $100^{th}$ to $350^{th}$. 
Fast Independent component analysis provides the estimated cluster values with doesn’t provide any noise when compared to PCA. Adulteration samples are grouped based on their reference spectral ratio. The samples are grouped under five category of different mixed ratio.

In Figure 7 shows the Random Forest classifier snapshot for the adulteration of mixed rice powder. Below Figure 7 represents the value where the rice mixed with the combo of 3:0 ratios. This is very easy method for finding adulterated product by using image processing. Here there is no chemical used and also there is no persons doesn’t involved for finding adulterated product. In this study, automatically found weather product is adulterated or good product. This method little bit consuming time only for trained different kind of adulterated product. Once trained process done, it will provide the adulteration result very quickly. Here, performance metrics based on error rate. So many spectral images were trained, from that 30 images were tested for different kind of adulteration in mixed rice powder. Accuracy was found by using this error rate based performance metrics which means it provides only one negative result for finding different kind of adulteration in mixed rice powder.

5. Conclusion
The aim of this study, used to found a different kind of adulterated rice mixture. In this project, initially spectrum images were collected from standard database [1]. The valuable information is gathered by using FICA. The result of FICA image was applied to extract the features such as temporal and spectral image to determine the effective data and classify the samples. These features are used to grade the various adulteration levels of rice mixture. Here, different kinds of samples are trained and 30 spectrum images were tested. Then the images were classified by using the random forest classifier, Based on random classifier, instead of the entire rice spectrum provide true positive except only one spectrum image. The accuracy of the adulteration of rice was calculated with the help of yielded test results. The total accuracy of using Random forest classifier was 97.66%. Random forest classifier provided the better results on finding each category on various adulteration levels of...
rice mixture. The proposed, the accuracy based on random forest classifier produces the better result than the PCA cluster along with Random forest classifier.

Future scope

- To increase the speed to speed up the process, the training of convolutional neural network (CNN) could be used.
- In future, the adulterated system can be connected to make it available for users to test the adulteration level of rice powder mixture.

References

[1] Chao li Bin li and Dapeng ye 2020 Analysis and Identification of Rice Adulteration Using Terahertz Spectroscopy and Pattern Recognition Algorithms IEEE 10.1109/2020.2970868
[2] Vlaic Durasevic M and Jakobovic D 2019 Improving genetic algorithm performance by population initialization with dispatching rules Com- put. Ind. Eng vol. 137 Art. no. 106030
[3] MihaiDanciu ,TeodoraAlexa-Stratulat 2019 Terahertz Spectroscopy and Imaging: A Cutting EdgeMethod for Diagnosing Digestive Cancers MDPI2019 12,doi:10.3390/ma12091519
[4] Basavaraj S. Anami, Naveen N. Malvade, SurendraPalaiah 2018 Automated recognition and classification of adulteration levels from bulk paddy grain samples Information processing in agriculture chapter 6 pp 47–60
[5] Neha S Dhande Rupesh D Sushir (2018) Detection and Estimation of Adulteration in Oil Sample Using Digital Image Processing IJSDR 4(2)
[6] Saritha B,Gokila G, Gokulpriya A,Mahalakshmi M 2017 Adulteration Identification of Papaya Seeds in Black Pepper Using Digital Image Processing IJIRCC 20(47)
[7] NikhadePratibha, More Hemlata, ManekarKrunaliand Prof Khot S. T. 2017 Analysis And Identificationof Rice Granules Using Image Processing and Neural Network International Journal of Electronics and Communication Engineering. ISSN 0974-2166 Volume 10, Number 1 pp. 25-33
[8] Deng W Zhao H Zou L Li G Yang X and Wu D 2017 A novel collaborative optimization algorithm in solving complex optimization problems Soft. Comput vol. 21, no. 15, pp. 43874398
[9] Liu Zhang and Han 2016 Feasibility study of determination of high-fructose syrup content of Acacia honey by terahertz technique ProcSPIE vol. 10030
[10] KasunHerath 2016 Rice grains classification using image processing technics researchgate-11373270
[11] Liu Hu Yang and Zheng 2016 Application of terahertz spectroscopy imaging for discrimination of transgenic Rice seeds with chemo metrics Food Chem., vol. 210, pp. 415421
[12] Liu W Liu C Chen F 2016 Discrimination of transgenic soybean seeds by terahertz Spectroscopy Sci.Research
[13] Mendes R Rocha Da 2015 Quantification of extra-virgin olive oil adulteration with Soybean oil: study of NIR, MIR, and Raman spectroscopy associated with chemo metric approaches Food Anal. Methods
[14] Xu W Xie L Ye Z Gao WY Yao M Chen and Y Ying 2015 Discrimination of transgenic Rice containing the Cry1Ab protein using terahertz spectroscopy and chemometrics Sci. Rep vol. 5, Art. no. 11115
[15] Kong Zhang Liu 2013 Rice seed cultivar identification using near-infrared hyperspectral imaging and multivariatedata analysis Sensors
[16] Saritha B Gokila G Gokulpriya A Mahalakshmi K 2017 Adulteration Identification of Papaya Seeds in Black Pepper Using Digital Image Processing Vol. 5, Issue 3
[17] Lakshmi Narayanan K Santhana Krishnan R Golden Julie E Harold Robinson Raghvendra Kumar Le Hoang Son Trinh Xuan Hung Pijush Samui
Phuong Thao Thi Ngo Dieu Tien Bui 2020 *A New Integrated Approach Based on the Iterative Super-Resolution Algorithm and Expectation Maximization for Face Hallucination* Applied Sciences, Vol. 10, No. 2, pp. 718-740.

[18] K. Lakshmi Narayanan, Dr. G. P. Ramesh 2017 *VLSI Architecture for Multi-Band Wavelet Transform based Image Compression and Image Reconstruction* Journal of Engineering and Applied Science

[19] K. Lakshmi Narayanan Dr. G. P. Ramesh 2017 *VLSI Architecture for 2D-Discrete Wavelet Transform DWT Based Lifting Method* Indian Journal of Public Health Research and Development, 08(03)276-281

[20] Redo-Sanchez Laman Infr. Milli 2013 “Review of Terahertz technology readiness assessment and applications”, IEEE Trans. THz Sci. Technology

[21] Feng X Zhan Q Cong P 2013 *Preliminary study on classification of Rice and detection of paraffin in the adulterated samples by Raman spectroscopy combined with multivariate analysis* Talent

[22] Gowen A O'Sullivan 2012 *Terahertz time domain spectroscopy and imaging: Emerging techniques for food process monitoring and quality control* Trends Food Sci Technology

[23] Alvarado-Diaz Witman Meneses-Claudio Brian and Roman-Gonzalez Avid 2011 *Milk Purity Recognition Software through Image Processing* IJACSA pp 2(3):27:1–27:27

[24] De Macedo Neto J Jefersson A dos Santos and William Robson Schwartz 2010 *Meat adulteration detection through digital image analysis of histological cuts using LBP* Journal of emerging trends in computing and information sciences-EISS 2218-6301: volume no-2; pp (90-94)

[25] Ioannis Ganopoulos, Anagnostis Argiriou, Athanasios Tsaftaris 2010 *Adulterations in Basmati rice detected quantitatively by combined use of microsatellite and fragrance typing with High Resolution Melting (HRM) analysis* Food Chemistry