Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 171 (2017) 72–79

Contents lists available at ScienceDirect

Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy

journal homepage: www.elsevier.com/locate/saa

Data fusion of near-infrared and mid-infrared spectra for identification of rhubarb

Wenjuan Sun, Xin Zhang *, Zhuoyong Zhang *, Ruohua Zhu

Department of Chemistry, Capital Normal University, Beijing 100048, China

ARTICLE INFO

Article history:
Received 17 March 2016
Received in revised form 26 July 2016
Accepted 26 July 2016
Available online 27 July 2016

Keywords:
Rhubarb
Near-infrared
Mid-infrared
Data fusion

ABSTRACT

Rhubarb has different medicinal efficacy to official rhubarb and may affect the clinical medication safety. In order to guarantee the quality of rhubarb, we established a method to distinguish unofficial rhubarbs. 52 official and unofficial rhubarb samples were analyzed using near-infrared (NIR) spectroscopy and mid-infrared (MIR) spectroscopy for classification. The feature vectors, which were selected by wavelet compression (WC) and interval partial least squares (iPLS) from NIR, MIR spectra, were fused together for identifying rhubarb samples. Partial least squares-discriminant analysis (PLS-DA), soft independent modeling of class analogies (SIMCA), support vector machine (SVM) and artificial neural network (ANN) were compared for classifying rhubarb. The use of data fusion strategies improved the classification model and allowed correct classification of all the samples.

1. Introduction

Rhubarb grows wild in the mountains of western and northwestern of China [1,2]. As a traditional herb used in Chinese medicine prescriptions for thousands of years, it has biological properties, including analgesic, anti-inflammatory, liver protective, choleretic, purgative, and antimicrobial [3–5]. In the Chinese Pharmacopoeia (PPRC, 2010a), three species are assigned as official rhubarb, i.e., *Rheum palmatum* L., *Rheum tanguticum* Maxim. ex Balf., and *Rheum officeinale* Baill [6,7].

Due to the increasing demands, the output of wild rhubarbs gradually decreases and official cultivated rhubarb cannot meet the demand of the medical industry in recent years. Some unofficial rhubarb was found in medicine market. Unofficial rhubarb has lower evacuative activity and higher antidiarrheal activity compare to the official rhubarb. The different medicinal efficacy between official rhubarb and unofficial rhubarb would negatively affect the clinical medication safety. Therefore, in order to guarantee the safety of rhubarb drugs, it is necessary to develop an efficient method to identify unofficial rhubarbs.

The identification of rhubarbs depends on its appearance, microscopic characteristic and physical-chemical properties. The published methods for identification relay on the experience of the inspector to a certain extent, and they are difficult to identify the difference between official rhubarbs and unofficial rhubarbs root, especially after the rhubarb samples have been processed into powder [8].

To develop novel fast and nondestructive methods, Zhang et al. [9] used NIR spectrometry and Takagi-Sugeno Fuzzy system to identify rhubarb samples. Zhao et al. [10] used wavelet package entropy and fisher discriminant in near-infrared spectrum to identify medicinal rhubarbs. Feng et al. [11] focused their attention on the HPLC method to identify rhubarbs and some prescriptions containing rhubarb. Wang et al. [12] used near-infrared spectrometry and temperature-constrained cascade correlation networks to identify the rhubarbs. In general, for rhubarbs, all these instrumental techniques coupled with chemometrics could provide an authentication which was only conducted on single spectrum instrument. The truth is different instruments may provide different chemical information, and data fusion method gives a chance to full use of those information and combine the advantages of various kinds of instrument technology.

The aim of present study was try to investigate the spectroscopy technique combine with chemometrics methods to identify official rhubarbs. In order to fully use the advantages of both near-infrared (NIR) and mid-infrared (MIR) spectra instrument, different data fusion methods were adopted in this study. Chemometrics classification methods like partial least squares-discriminant analysis (PLS-DA), soft independent modeling of class analogies (SIMCA), support vector machine (SVM) and artificial neural network (ANN) were used to distinguish official rhubarbs from unofficial ones. Then, in order to take advantage of all kinds of complementary collected information, data fusion methods selected feature vectors from NIR and MIR spectra by wavelet compression (WC) and interval partial least squares (iPLS), were carried out to integrate the data from different instruments. Random Forest is a relatively novel method gaining a lot of attention in infrared spectra analysis [13–15]. But consider in this case, we should focused on the introduction of data fusion method, we only compared the most commonly used classifiers. Model evaluation parameters, such as train set accuracy rate (the correct classification number of

* Corresponding authors.
E-mail addresses: xinkevinzhang@outlook.com (X. Zhang), gusto2008@vip.sina.com (Z. Zhang).

http://dx.doi.org/10.1016/j.saa.2016.07.039
1386-1425/

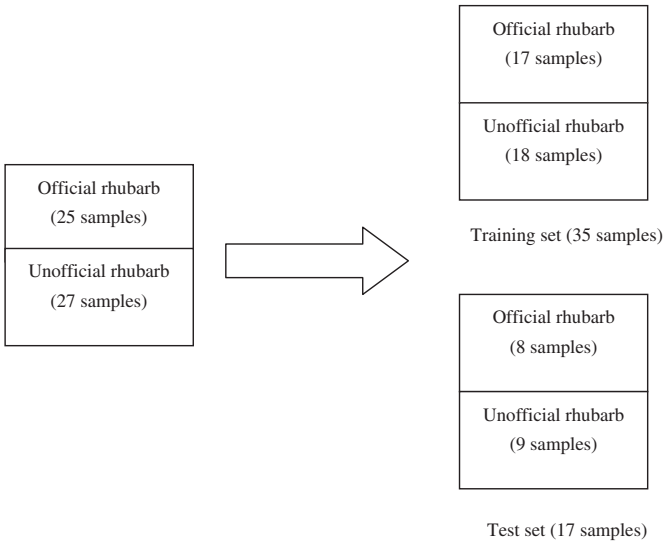

Fig. 1. Graphical representation of the data set from official and unofficial rhubarbs.

train set divided by the total number of train set), cross validation accuracy rate and test accuracy rate (the correct classification number of test set divided by the total number of test set) were used to evaluate the accuracy of models.

2. Experimental

2.1. Samples

25 official and 27 unofficial rhubarb samples were collected from different regions of China for study. A graphical representation of the composition of the data set is displayed in Fig. 1.

All of the samples were dried under 50°C for 2 h and ground into 60 mesh powder. The NIR spectra were collected from prepared rhubarb samples directly. The MIR spectra were measured from tablets pressed by rhubarb samples powder blended with potassium bromide powder [16,17].

Savitzky-Golay (SG) smoothing [18,19], 1st derivative, 2nd derivative and multiplicative scatter correction (MSC) [20] were used for data pretreatment.

2.2. Instrumental

NIR spectra were acquired using a Foss 6500 NIR spectroscopy equipped with a PbS detector. The spectra were acquired between 4000 and 8893 cm^{-1} at a resolution of 7 cm^{-1}. The experimental temperature during the whole experiment was maintained at room temperature.

MIR spectra were acquired using a Perkin Elmer 1730 FT-IR spectroscopy equipped with a DTGS detector. The spectra were acquired between 400 and 1798 cm^{-1} at a resolution of 2 cm^{-1}. The experimental temperature during the whole experiment was maintained at room temperature.

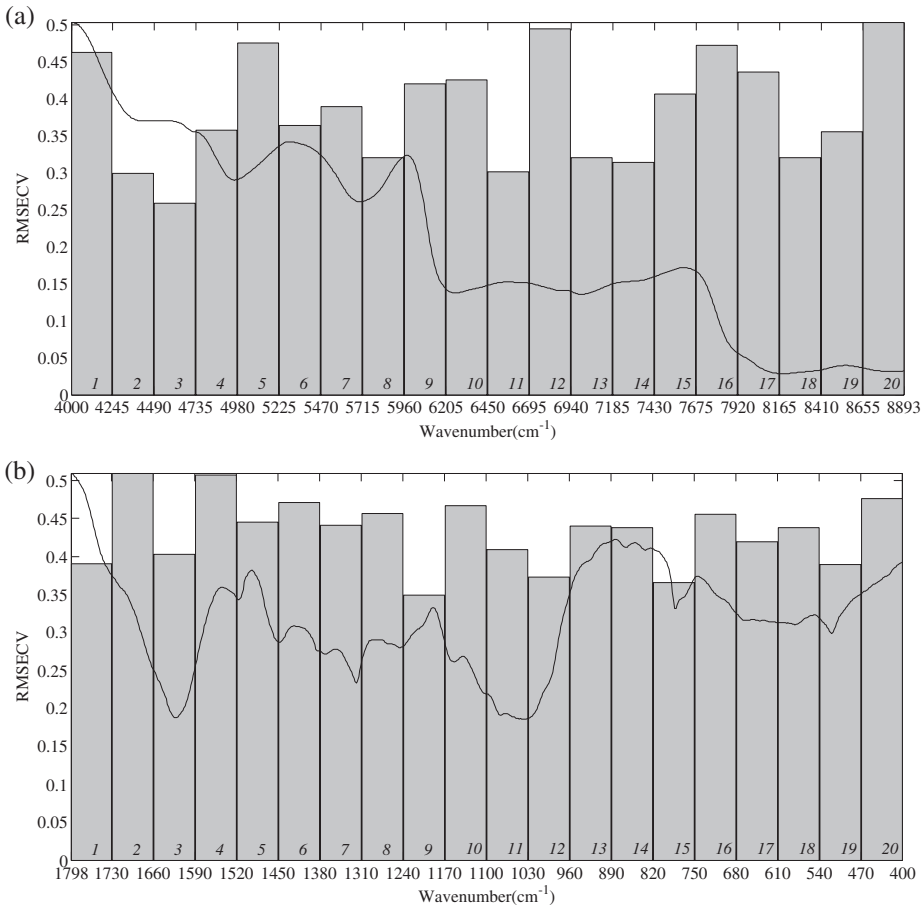

Fig. 2. The RMSECV values of iPLS models with interval number 20 on different spectroscopy: (a) NIR; (b) MIR.