Forensic Science International 235 (2014) 1–7

Contents lists available at ScienceDirect

Forensic Science International

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

A framework for selecting analytical techniques in profiling authentic and counterfeit Viagra and Cialis

Michel J. Anzanello [a,*], Rafael S. Ortiz [b], Renata Limberger [c], Kristiane Mariotti [c]

[a] Department of Industrial Engineering, Federal University of Rio Grande do Sul, Av. Osvaldo Aranha, 99 – 5° andar, Porto Alegre, RS, Brazil
[b] Rio Grande do Sul Technical and Scientifical Division, Brazilian Federal Police, Avenida Ipiranga 1365, 90160-093 Porto Alegre, RS, Brazil
[c] Department of Pharmacy, Universidade Federal do Rio Grande do Sul, Av. Ipiranga, 2752, 90610-000 Porto Alegre, RS, Brazil

ARTICLE INFO

Article history:
Received 5 August 2013
Received in revised form 8 November 2013
Accepted 7 December 2013
Available online 17 December 2013

Keywords:
Analytical technique
Counterfeit drugs
PCA
k-Nearest Neighbor
Support Vector Machine

ABSTRACT

Several analytical techniques aimed at profiling drugs are deemed costly and time consuming, and may not be promptly available for analysis when required. This paper proposes a method for identifying the analytical techniques providing the most relevant data for classification of drug samples into authentic and unauthentic categories. For that matter, we integrate principal components analysis (PCA) to *k*-Nearest Neighbor (KNN) and Support Vector Machine (SVM) classification tools. PCA is first applied to data from five techniques, i.e., physical profile, X-ray fluorescence (XRF), direct infusion electrospray ionization mass spectrometry (ESI-MS), active pharmacological ingredients profile (ultra performance liquid chromatography, UPLC–MS), and infrared spectroscopic profile (ATR-FTIR). Subsets of PCA scores are then combined with a "leave one subset out at a time" approach, and the classification accuracy using KNN and SVM evaluated after each subset is omitted. Subsets yielding the maximum accuracy indicate the techniques to be prioritized in profiling applications. When applied to data from Viagra and Cialis, the proposed method recommended using the data from UPLC–MS, physical profile and ATR-FTIR techniques, which increased the categorization accuracy. In addition, the SVM classification tool is suggested as more accurate when compared to the KNN.

1. Introduction

The production of counterfeit medicines is a criminal problem that carries serious risks to public health [1], since there is no certainty about pharmaceutical dosage forms, active pharmacological ingredient, and origin of raw materials or manufacturing conditions. Such problem has motivated the development and application of a large number of analytical techniques and multivariate tools in forensic analysis tailored at discriminating between authentic and unauthentic samples of seized drugs, and finding similar properties in unauthentic samples [2,3].

In recent years, our group has applied profiling approaches and image processing in counterfeit Viagra® (sildenafil citrate, SLD, Pfizer) and Cialis® (tadalafil, TAD, Eli Lilly) samples seized by the Brazilian Federal Police in the state of Rio Grande do Sul, Southern Brazil. Such studies have relied on five analytical techniques, which were later integrated to multivariate techniques. The first analytical technique is the assessment of tablets physical profile [4] as suggested in [5,6], which provided 4 post-tabletting variables: mass (mg), thickness (mm), shorter length (mm) and longer length (mm). Such variables were analyzed through the *F*-test (ANOVA), showing that counterfeiters cannot mimic the mass variable in unauthentic drugs, but can often mimic genuine products length. The second analytical technique provided inorganic fingerprinting data obtained by X-ray fluorescence (XRF), originating 2048 variables (energy, keV) [7]. XRF is a nondestructive technique for characterization of metal presence [8] featured by multielemental capability, good detectivity, and short analysis time. In our propositions, XRF was aligned with PCA and hierarchical cluster analysis to enable the semi-quantitative determination of sildenafil citrate and excipients such as calcium phosphate, titanium oxide and iron oxide. That allowed us to classify authentic and counterfeit Cialis and Viagra samples.

The third analytical technique, organic profile achieved by direct infusion electrospray ionization mass spectrometry (ESI-MS), highlights the polar composition of seized tablets [9] without the need for isolation and prior chromatographic separation. Such analysis generated 99 variables corresponding to ions (m/z) in

* Corresponding author. Tel.: +55 51 3308 4423.
E-mail addresses: michel.anzanello@gmail.com (M.J. Anzanello), rafaelortiz.rso@dpf.gov.br (R.S. Ortiz), renata@ufrgs.br (R. Limberger), krismariotti@gmail.com (K. Mariotti).

0379-0738/$ – see front matter © 2013 Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.forsciint.2013.12.005

methanol extracts samples. Spectra for authentic Viagra® depicted ions exclusively corresponding to the SLD molecule: $[SLD + H]^+$ of *m/z* 475; $[SLD + Na]^+$ of *m/z* 497; and $[2SLD + H]^+$ of *m/z* 949; authentic Cialis® showed ions of *m/z* 343, 365 and 707 from the lactose molecule (excipient). PCA was applied to ESI-MS fingerprint data, allowing the classification of samples into categories based on their contents of active ingredients. The fourth analytical technique, UPLC–MS [10], enabled qualitative and quantitative determination of the active pharmaceutical ingredients (APIs), and yielded 3 variables: SLD content, HSD (Homosildenafil) content and TAD content (in mcg/mg). Such variables enabled assessing products authenticity, presence of the API indicated on the label, API concentration and presence of other contaminants. Finally, the fifth analytical technique relied on attenuated total reflection Fourier transform infrared (ATR-FTIR) [11]. This experiment generated 661 variables (wavenumbers, cm^{-1}) on mid-infrared region (1800–525 cm^{-1}), which includes the absorption region that highlights major differences in PDE-5 inhibitors [12]. PCA was applied to ATR-FTIR aimed at grouping samples according to chemical profiles, distinguishing successfully between authentic and counterfeits samples and suggesting a common illicit source for different seized batches. Although not assessed in this paper, other relevant analytical techniques and multivariate tools have been employed to detect counterfeit drug samples. Degardin et al. [13] applied Support Vector Machine and PCA to classify and profile data from Raman spectra, while Deconinck et al. [14] performed an exploratory analysis based on projection pursuit and clustering aimed at discriminate between authentic and counterfeit samples of Viagra and Cialis. Similarly, Been et al. [15] applied several supervised and unsupervised techniques to near-infrared (NIR) and Raman spectroscopy data, including KNN, partial least squares discriminant analysis and probabilistic neural network, while Deconinck et al. [16] suggested a framework based on classification trees to categorize authentic and unauthentic medicine samples. Finally, Anzanello et al. [3] proposed a wavenumber selection approach to identify the most relevant FTIR bands for inserting Viagra and Cialis samples into two classes, and Sacré et al. [17] used partial least squares to identify the most effective FITR bands for detecting forged medicines.

Among the several benefits provided by the five aforementioned analytical techniques in forensic applications, two are noteworthy: (i) they enable detecting authentic and unauthentic samples of seized medicines; and (ii) they allow finding similar properties in unauthentic samples, making it possible for police forces to unveil identical sources from different drug seizures. The availability and cost of such techniques, however, may limit the ability to run the entire set of analyses, since many laboratories and forensic institutes do not have the necessary equipment to perform all techniques for a consolidated result. In addition, reduced time for obtaining conclusive results is typically verified in practical cases, prohibiting the running of all experiments. Thus, it seems reasonable developing methods able to identify the analytical techniques responsible for providing the most relevant data aimed at inserting seized samples into proper categories.

This paper proposes a method for selecting the analytical techniques providing the most conclusive data for categorizing seized drugs into authentic and unauthentic classes. For that matter, we integrate PCA to two data mining tools with classification purposes, *k*-Nearest Neighbor (KNN) and Support Vector Machine (SVM). In our propositions, PCA is applied to the data provided by the five analytical techniques, and the subsets of PCA scores deriving from each analytical technique are compiled into a single matrix consisting of new classificatory variables. The use of PCA is justified by its ability to merge information from analytical techniques described by a different number of variables into a new set of variables (the scores) similar in number and magnitude. Next, the most efficient subsets of PCA scores (each subset related to an analytical technique) are identified combining a "leave one subset out at a time" procedure with the KNN and SVM. In such procedure, each subset of scores is momentarily omitted from the dataset, a categorization using the remaining scores subsets is performed, and the classification accuracy, i.e., the proportion of correct categorizations, is evaluated. When all subsets were omitted once, the subset yielding the highest accuracy is removed from the dataset since it is the one that contributes the least in separating samples into categories. This iterative procedure is performed in the remaining subsets until there is only one subset left. The maximum accuracy obtained during the elimination process indicates the analytical techniques that provide the most relevant data for accurate classifications. Such techniques are then recommended in time and budget limited scenarios, and may be used as a guide to acquire new equipments for future analyses.

When applied to Viagra and Cialis data derived from the five aforementioned techniques, the proposed method recommended the use of data provided by UPLC–MS, physical profile and ATR FTIR techniques, since such data increased the classification accuracy of samples into authentic and counterfeit classes. Further, the SVM classification tool was suggested as a more accurate technique when compared to the KNN algorithm.

2. Materials and methods

2.1. Samples

Fifteen counterfeit Viagra tablets from four different seizures, 36 counterfeit Cialis tablets from five different seizures, 4 authentic tablets of authentic Viagra®, and 4 tablets of authentic Cialis® were sent to the Technical and Scientific Division for Forensic Analysis of Rio Grande do Sul State. Sildenafil citrate (99.9%) and Viagra® tablets containing 50 mg of SLD were supplied by Pfizer Ltda Laboratories. Tadalafil (99.8%) and Cialis® tablets containing 20 mg of TAD were supplied by Eli Lilly do Brasil Ltda Laboratories.

2.2. Analytical techniques

2.2.1. Physical profile

The post-tabletting characteristics, i.e., mass (mg), thickness (mm), shorter length (mm) and longer length (mm), were used as variables in the study. Measurements were performed using an analytical balance (Mettler Toledo XP205, Brazil), and a micrometer (Mitutoyo, Japan). No data pre-processing was performed.

2.2.2. Inorganic profile – XRF analysis

The ED-XRF experiments were performed using an X-ray spectrometer Shimadzu® model EDX 700 (Kyoto, Japan). Measurements were performed under air with a beam collimation of 3 mm, 25% of detector dead time, with the current automatically adjusted during spectrum acquisition to keep the detector dead time of 25%. The Shimadzu EDX 700 features (1) a Rh X-ray generator, with tube voltage ranging from 5 to 50 kV, and tube current from 1 to 1000 A, and (2) a semiconductor detector, Si(Li), with detection area of 10 mm^2 and resolution < 155 eV. Sildenafil and tadalafil tablets were crushed using a mortar and placed into XRF cells on Mylar™ film (3 mm thickness). Measurement time was 250 s. In all cases, spectra were recorded from 0 to 40 keV with an energy step of 0.02 keV, resulting in 2048 points for each spectrum. The spectral data were mean centered.