Journal of Pharmaceutical and Biomedical Analysis 128 (2016) 89–97

Contents lists available at ScienceDirect

Journal of Pharmaceutical and Biomedical Analysis

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

Near infrared spectroscopy for counterfeit detection using a large database of pharmaceutical tablets

Klara Dégardin[a,*], Aurélie Guillemain[a], Nicole Viegas Guerreiro[b], Yves Roggo[a]

[a] F. Hoffmann-La Roche Ltd., Bldg 250 Room 3.504.01, Wurmisweg, 4303 Kaiseraugst, Switzerland
[b] ETH Zurich, Vladimir-Prelog-Weg 1-5/10, 8093 Zurich, Switzerland

ARTICLE INFO

Article history:
Received 7 April 2016
Received in revised form 29 April 2016
Accepted 1 May 2016
Available online 6 May 2016

Keywords:
Counterfeit medicines
Near infrared spectroscopy
Pharmaceutical tablets
Supervised classification methods
Validation

ABSTRACT

Medicine counterfeiting is one of the current burdens of the pharmaceutical world. Reliable technologies have become available for the chemical analysis of suspect medicines. Near infrared spectroscopy (NIRS) allows for instance fast, specific and non-destructive authentication of pharmaceutical products. In this paper, a NIRS method is presented for the identification of 29 different pharmaceutical product families of tablets, one family containing one or more formulation (s), e.g. different dosages. This selection represents the whole tablet portfolio of our firm. The high number of product families constituted a challenge, given that the measurement of the samples, made on two similar instruments, generated a dataset of 7120 spectra. Several chemometric tools proved efficient for the identification of these medicines. The dataset was first investigated with a Principal Component Analysis (PCA) in order to provide an overview of the distribution of the samples. The K-Nearest Neighbors (KNN), the Support Vector Machines (SVM) and the Discriminant Analysis (DA) supervised classification tools were successfully applied and generated an outstanding classification rate of 100% of correct answer. The methods were then fully validated with an independent set of spectra. The DA was selected as the method for the routine analysis of suspect tablets with the Mahalanobis distance as acceptance criterion for identification. Counterfeits, generics and placebos samples, constituting a second validation set, were tested and rejected by the method. NIRS has thus been demonstrated as an efficient tool for the quick identification of a large dataset of pharmaceutical tablets and the detection of counterfeit medicines.

1. Introduction

Counterfeiting is a crime with dreadful consequences, especially in the case of medicines. Estimates are difficult to quantify, since the figures only reflect the cases reported by the patients and/or the authorities. However the provided number of counterfeits has been increasing for the last years, making them one of the worst current pharmaceutical burdens [1–4]. The risk for the patients is consequent. All types of counterfeits can indeed be found, like products devoid of active pharmaceutical ingredient (API), or under dosed, medicines with wrong APIs or even with toxic compounds [5]. Even the so-called "placebo counterfeits" – incorrectly designating products made of excipients – are harmful, since they are not active and are usually found to be produced in frightening conditions [6,7]. Besides, the effect on the economy and on the image of the healthcare system and the pharmaceutical companies is appalling. This trade is even considered more rentable than illicit drugs [8]. Organised criminal networks have been demonstrated to be behind the production and the selling of counterfeit medicines [9–11]. An appropriate response is consequently necessary and as much as possible, prevention, e.g. by the use of anti-counterfeiting features and a better control of the legal supply chain [12,13]. Fast and reliable analyses are necessary to confirm the cases and evaluate the risk encountered by the patients. Many techniques have proved efficient for counterfeit analysis, and can generally be classified into chromatographic or spectroscopic methods [14–20]. Near Infrared spectroscopy (NIRS) presents many advantages for that purpose. It is indeed a fast, non-destructive method, that provides chemical and physical information about the analysed samples, even through plastic or glass [21]. NIRS has been widely used in the pharmaceutical industry for instance for the identification of raw materials or API, or the determination of water content [22]. In the last years, NIRS has also been evaluated as a reliable tool for counterfeit identification. Thanks to chemometric tools, the chemical and physical signature of a suspect sample can be rapidly compared to the genuine references, providing a fast yes/no answer. The great

* Corresponding author.
E-mail address: klara.degardin@roche.com (K. Dégardin).

http://dx.doi.org/10.1016/j.jpba.2016.05.004
0731-7085/

advantage of NIRS is that the whole formulation can be checked. Indeed the presence of the right excipients must also be controlled, since counterfeits can still present the right API – even in the right amount – but the wrong excipients [23]. NIRS also presents the advantage of providing information about the chemical composition of the confirmed counterfeits thanks to spectral analysis [24]. The difficulty in the NIRS analyses resides in the computation of the right calibration methods for a correct and easy authentication. Unsupervised techniques like Principal Component Analysis (PCA) or Cluster analysis present the advantage of providing an overview of the distribution of the sample set and are often used in a first step prior to other chemometric tools [25–28]. Supervised techniques are usually privileged for counterfeit authentication since they allow the attribution of a class to the unknown sample and also a clear identification. Supervised clustering techniques are also available, that can be divided into linear, like the Discriminant Analysis (DA) or Partial Least Squares – DA (PLS – DA, and non-linear techniques, like the Support Vector Machines (SVM)) [29,30]. Several techniques have been published for counterfeit identification of tablets and capsules, and among them the PLS – DA analysis for mid infrared spectroscopy (MIRS), NIRS and Raman spectroscopy analysis of 94 counterfeits of two different product families [31], or for the NIR detection of artesunate antimalarial tablets [32]. The most published chemometric tool for counterfeit authentication with NIR is probably the Soft Independent Modeling of Class Analogy (SIMCA), employed for the identification of paracetamol tablets [33], or in another paper of thirteen different types of tablets [34]. Two types of generic products were analysed by further teams, who also used algorithms based on the SIMCA [35,36]. O'Neil et al. proposed another method based on Discriminant Analysis (DA) called Unequal Class modeling (UNEQ) associated with the Mahalanobis distance for the identification of two types of tablet with a handheld device [37].

The method presented in this paper consists in the NIR identification of all the tablets produced by the firm, which represents 29 pharmaceutical product families, each of them having at least one formulation, e.g. one dosage. This constitutes the main novelty of this work since to our knowledge, much more molecules have been introduced in the presented database than previously published. This huge dataset, composed of more than 7000 spectra, was first investigated by PCA and then different supervised techniques, namely the SVM (SVC Kernel linear and Radial Basis Function), the K-Nearest Neighbors (KNN) and the DA. The different chemometric methods were calibrated and validated with an independent set of spectra. The chemometric approach constitutes another novelty of this work. Indeed advanced methods computed with the Python software have first been used while in a second phase, a method based on discriminant analysis has been presented for the quick and automatic analysis of suspected counterfeits in the routine.

2. Material and methods

2.1. Samples for the development and the validation of the methods

2.1.1. Genuine samples

29 Different pharmaceutical products have been measured. Each of these products consisted of one to four dosage(s). In total, 53 formulations have been included in the dataset.

At least five batches per product were measured. The resulting spectra were split into three sets: the calibration set, the test set and the validation set, constituted of independent spectra. The test set was used for the optimization of the SVM and the KNN methods. For the products manufactured at several production sites, more than five batches were selected in order to include this variability in the dataset. Additionally five tablets per batch were measured.

In the end, the measurement were performed for 53 formulations, at least 5 batches (from all possible production sites), 5 tablets per batch, 2 repetitions per tablet and 2 instruments, which led to a total of 7120 spectra. 1060 spectra were extracted for the test set, representing spectra of one batch per formulation. Then the resting spectra were divided into two thirds for the calibration, and one third for the validation. To sum up, 4040 spectra were integrated in the calibration set, 1060 spectra in the test set, and 2020 spectra constituted the validation set.

2.1.2. Dataset for the second validation: challenging samples

An additional dataset was added for the second validation, constituted of 350 spectra of counterfeits, generics and placebos. These spectra were acquired with one repetition and one instrument, since the samples were measured during routine analyses. The spectra were then reprocessed for the purpose of this study.

2.1.3. Calibration and validation concept of the screening method

The SVM and KNN identification methods were computed with calibration spectra and validated with an independent set of spectra. A test set was created in order to optimize the methods. The methods were then validated with the validation set. The specificity was checked by the rate of correct identification of the calibration, test and validation spectra.

The NIRS methods were developed and validated according to the European Medicine Agency (EMA) [38] and ICH guidelines Q2 (R1) "Validation of Analytical Procedures: Text and Methodology" [39].

2.1.4. Calibration and validation concept of the DA method

The DA was then computed with calibration spectra, and then tested with the validation spectra. The DA was selected as the method of choice for the routine analysis of counterfeits and therefore finally tested and validated against a second validation set of spectra. This set consisted of 94 samples, split into 65 counterfeits, 25 generics and 4 placebos. These samples were tested to challenge the method with high quality counterfeits that could be received in the future. The specificity of the method was checked by the rate of correct identification of the calibration, validation and second validation spectra.

2.2. Analytical method

2.2.1. Near infrared spectroscopy

More recent than MIRS, NIRS first distinguishes by its range going from 700 to 2500 nm. The NIR spectra are generated by the vibration of the —CH, —OH, —NH and —SH bonds. The absorption bands result from the overtones and the combinations of the MIR fundamental bands [40]. The shape of the NIR spectra is consequently very different from the one of the MIR spectra, and their interpretation quite difficult. For this reason NIRS is often combined with chemometric tools gathering mathematical pretreatments, classification methods and regression techniques that allow to enhance variations between the spectra.

2.2.2. NIR measurements

Two identical Thermo® Antaris II Fourier Transform (FT)-NIR spectrometers have been used for the measurements. The instruments are built with a high-sensitivity InGaAs detector, a long-life Michelson interferometer and a high intensity halogen NIR source. Their wavenumber accuracy is $\pm 0.03\ cm^{-1}$.

The spectra were acquired in reflection mode on an integrating sphere using 32 scans, a resolution of $16\ cm^{-1}$, a gain of 1 and no attenuator. The whole available spectral range was used, going from

4000 to 10,000 cm^{-1} (or 1000 to 2500 nm). The tablets were measured directly on the measurement window without any sample holder. Two repetitions were performed, the tablets being flipped between each measurement. The measurements were performed on both instruments.

2.3. Chemometric tools

2.3.1. Data pretreatments

Different pretreatments were used depending on the chemometric tool chosen and the best results obtained for each method.

The PCA was calculated with spectra pretreated with a Standard Normal Variate (SNV) and with a Savitzky–Golay second derivative. The SVM and the KNN were computed using spectra first pretreated with a SNV (SNV) and then with a Savitzky–Golay second derivative. The spectra used for the DA were either pretreated with SNV, or with SNV coupled with a Savitzky–Golay first or second derivative.

2.3.2. Principal Component Analysis (PCA)

The PCA is a common unsupervised technique allowing the exploration of data through the reduction of its dimensionality. Linear combinations of the initial variables are computed, called Principal Components (PCs). The data, in the present case spectra, are then represented in the newly defined referential. Differences and similarities between the spectra are enhanced, allowing the detection of underlying clusters [29,41–43]. This technique was used in this context to have a first idea of the distribution of the sample set.

2.3.3. Support Vector Machines (SVM)

The SVM classification algorithm is non-linear and based on maximal margin hyperplanes. The dimension of the data is increased in order to maximize the distance between the samples and separate them in a new hyperplane. This is made possible by the use of a kernel function. Different types of kernel function can be applied, like the linear, the polynomial, the sigmoid and the radial basis functions (RBF). In the present case, the Support Vector Classification (SVC) kernel linear and the RBF have been used. Different parameters have to be optimized to increase the performance of the algorithm, like the regularization parameter (c) and the gamma parameter for the RBF [29,44–46]. C and gamma have been optimized by a grid search and chosen to obtain the lowest error rate on the test set. Due to its non-linear algorithm, the SVM is particularly adapted to separating samples with similar profiles, and was consequently also tested in this study.

2.3.4. K-Nearest Neighbors (KNN)

Another non-linear classification method presented in this paper is the KNN algorithm [47]. When applied to a dataset of spectra, the KNN classifies an unknown spectrum in function of the K closest spectra available, it means the K-Nearest Neighbors of the tested spectrum. The distance between unknown and dataset spectra is indeed calculated, for instance using the Euclidean distance. The unknown spectrum is then attributed the class of its closest neighbors. The important parameter to optimize when using the KNN algorighm is the K number, usually a small odd number [29]. As a non-linear tool, the KNN provides a strong algorithm for the identification of products that might present similar spectra between one another.

2.3.5. Discriminant analysis (DA)

The last classification tool tested in this study was the discriminant analysis, a traditional classifier that maximizes the between-class variance by computing linear combinations of the original variables. The algorithm calculates the distance of a sample from the mean of a set of samples.

A criterion used for acceptance is for instance the Mahalanobis distance (MD), which defines the distance between a sample and the center of a class [29,48]. The discriminant analysis was tested here for the identification of the tablets since it has proved its efficiency as a traditional and simple-to-use classifier. The intra-class distance was defined as the value between each calibration spectrum and the center of the class. A maximum intra-class distance was then calculated for each class. The inter-class distance was determined as the distance between the center of a class to the closest spectrum of the other product families. The minimum inter-class MD was also calculated for each class. The MD limit for each product family was then defined as the average value between the maximum intra-class and the minimum inter-class values, as described in the following equation:

$$\text{Limit} = \frac{\min(\text{inter}) + \max(\text{intra})}{2} \text{and } \text{Limit}_{\max} = 5$$

with min(inter) the minimum inter-class, and max(intra) the maximum intra-class.

The limit was calculated for each product family and was determined as the criterion for identification. Over this value the analysed product family had to be rejected, and thus considered as unidentified or “counterfeit”.

2.3.6. Software

The “Result NIR Suite 3” software (Thermo Fisher Scientific®, Madison, USA) with Result Operation, Result Integration and TQ Analyst software was used for the measurement and the computation of the spectra. The NIR spectra were recorded with the Result Operation and Integration software, while the DA methods were developed with TQ Analyst.

Additionally the other chemometric methods – the PCA, the SVM and the KNN – were computed with Python 3.4 (Anaconda Package, Continuum Analytics) [49,50] using the Spyder (The Scientific PYthon Development EnviRonment) 2.3.5.2 interface [51]. The data structures and analysis tools were provided by Pandas 0.16.2 while the fundamental package for scientific computing with Python was included in Numpy 1.9.2. The Scikit-lear 0.16.1 toolbox [52] enabled the computation of the PCA, the SVM and the KNN. The graphics were computed with Matplotlib 1.4.3 [53] and the derivative pretreatments with Scipy 0.15.1.

3. Results and interpretation

3.1. Investigation of the dataset

3.1.1. Spectra interpretation

The dataset obtained consisted of 7120 spectra of 29 different product families. The amount of product families represented a challenge for the identification. Indeed a genuine product family can present different formulations – and thus different spectra – depending on the dosage. The excipients may have been chosen different between two dosages of a same product family (Fig. 1). On the contrary, two different product families can present formulations that are very similar, like it is the case for the product families 7 and 21 (Fig. 2). Low dosage products, like the product family 21, can have a very close spectrum to its placebo (Fig. 3a). In Fig. 3b, one can see that a generic of the product family 12 has a very close composition to the product family 12, and might then be very difficult to reject. The example of counterfeit chosen for the product family 12 is in the present case quite easy to differentiate.

Consequently the complexity of the classification first resulted in the heterogeneity of the classes, with one product family having different spectra signatures. The second problem for the classification was that different product families, and thus classes, had close formulations and therefore similar spectra.

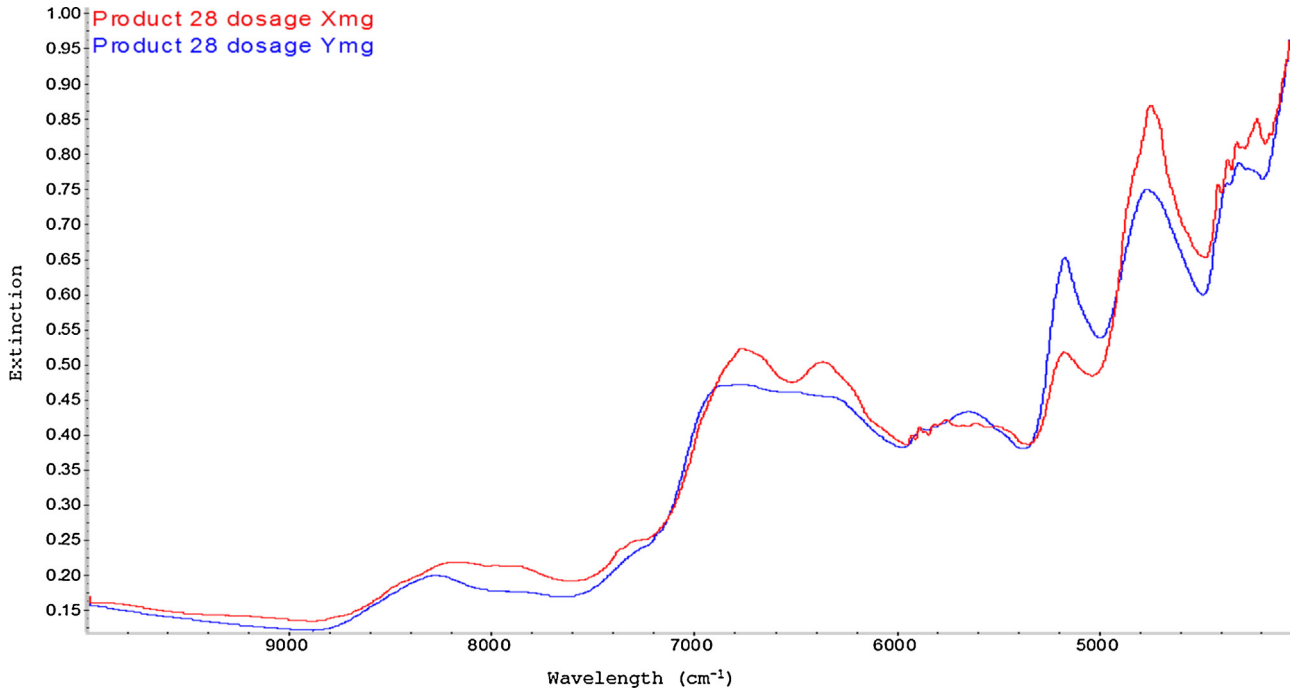

Fig. 1. NIR raw spectra of two dosages of the same product family. Differences are observed between the spectra, representative of differences in the formulation. Indeed the composition of the excipients is different.

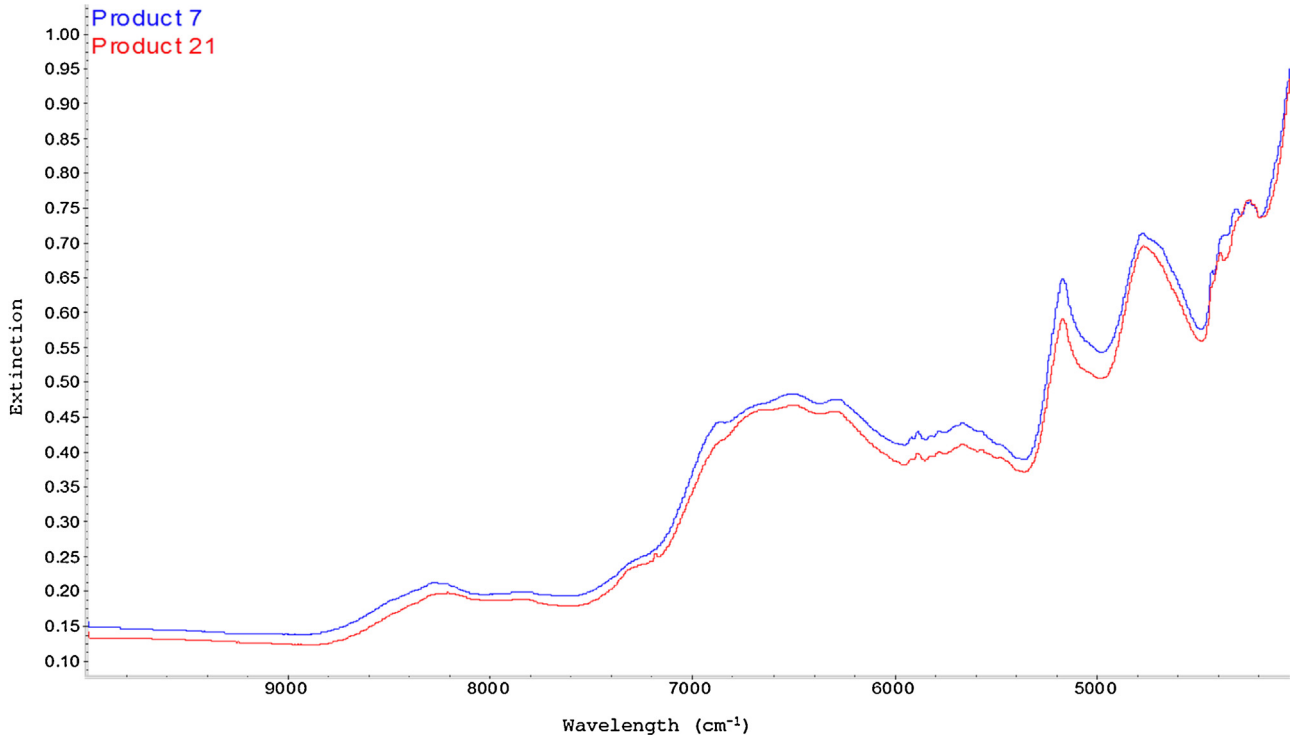

Fig. 2. NIR raw spectra of two different product families. Only small differences can be observed between the spectra, which reveals the closeness of formulation between both products.

3.1.2. Results of the PCA

The PCA applied to the calibration spectra is presented in Fig. 4. The spectra were pretreated with a SNV and a second derivative. Ten PCs have been used for the computation, describing 94% of variability.

While the first two PCs enable only the distinction of two product families (product family 4 and 29) from the dataset, the third and fourth PCs allow the separation of at least eight other product families. When considering the plots generated with the next PCs, one can quite easily differentiate most of the product families. This preliminary work suggests that the spectra of the 29 different product families should be sufficiently different from one another to enable the distinction of the 29 product families by supervised classification. However strong supervised tools should be used to

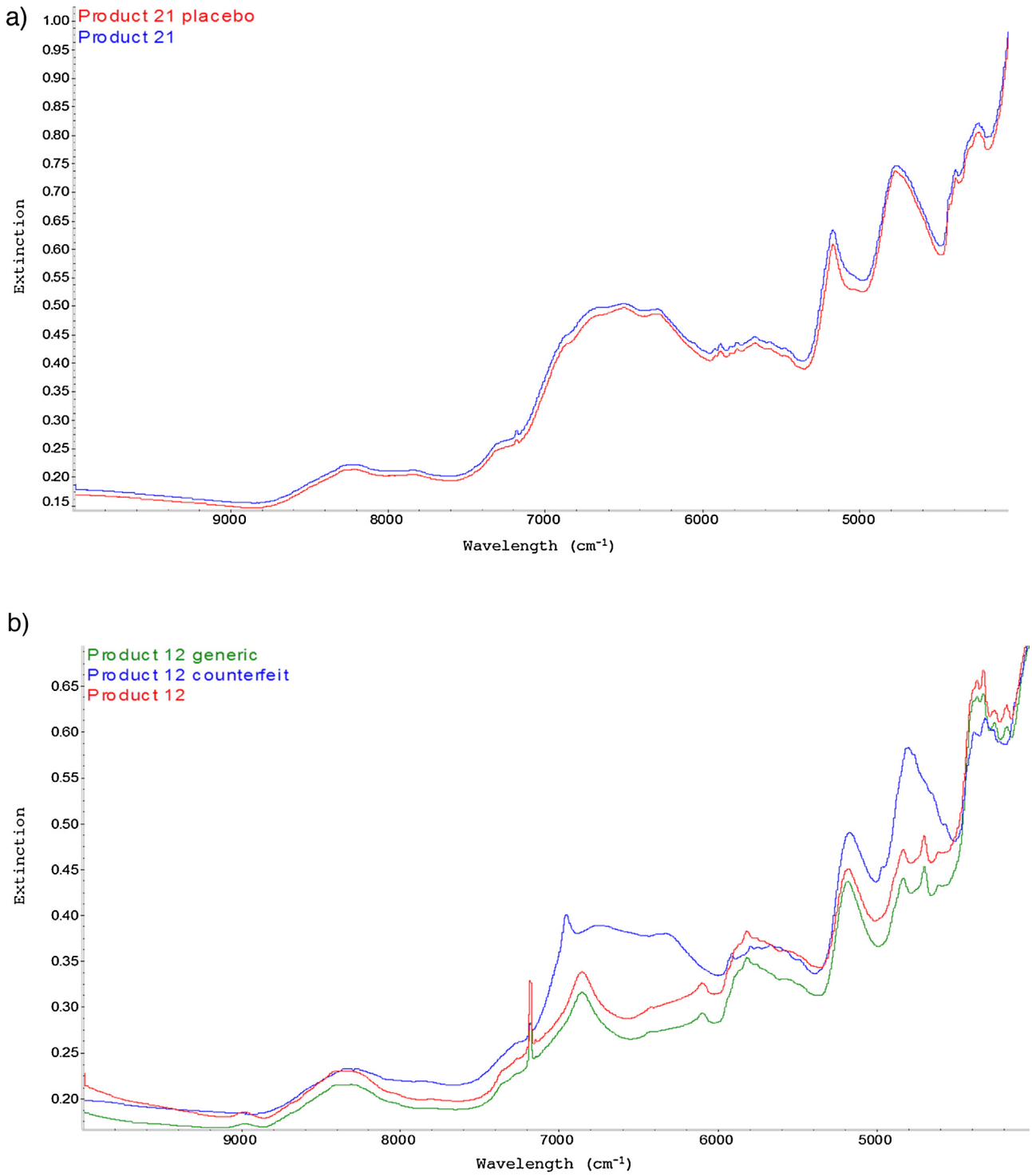

Fig. 3. NIR raw spectra of one genuine product with a placebo, a generic and a counterfeit. (a) The product 21 was compared to its placebo. Very little differences can be observed between the spectra. (b) The product 12 was compared to one of its generics and one of its counterfeits. While the differences with the counterfeit are quite obvious, the spectrum of the generics is very close to the spectrum of the product 12.

be able to distinguish product families with similar patterns with as little number of mismatches as possible.

3.2. Classification of the tablets with the SVM

Based on the results obtained with the PCA, the SVM algorithm was tested in order to dissociate product families that would be close in formulation. Two Kernel functions were tested: the SVC Kernel linear and the RBF.

In order to obtain reliable results, some parameters of the SVM functions had to be optimized: the C parameter for the SVC Kernel linear, and the C and the gamma parameters for the RBF. The parameters were optimized with a grid search on the test set. Only the selected results are displayed in this paper.

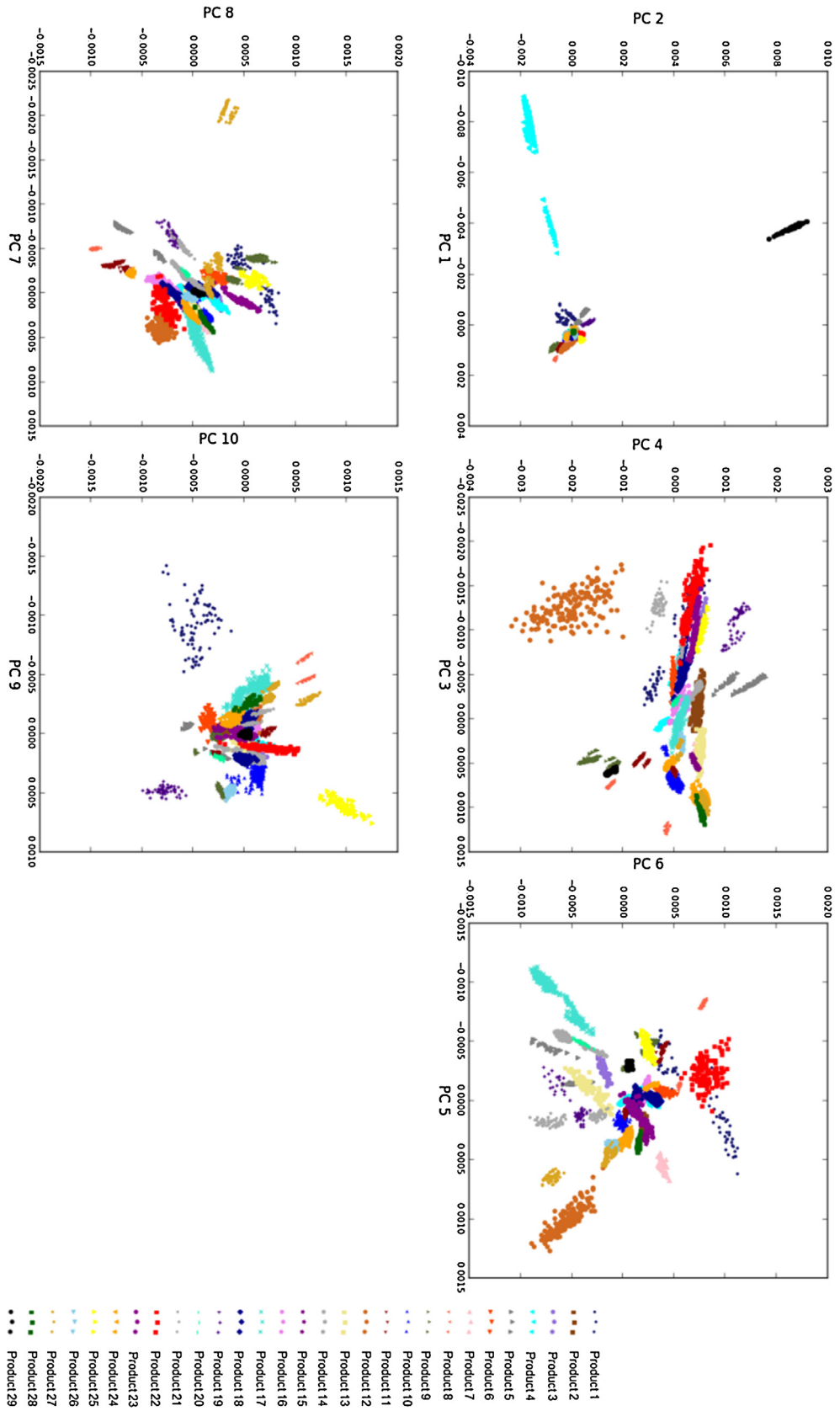

Fig. 4. Results of the PCA applied to the calibration spectra pretreated with SNV. Each color represents one of the 29 product families.

The classification was performed on the product families, resulting in 29 classes. The SVM methods were computed with the calibration spectra of the 29 product families and tested against the test set and the validation set. The results of the SVC Kernel linear and the RBF are displayed in Table 1. All the 29 product families were correctly identified with both functions. 100% of correct

Table 1
Results of the SVM classification obtained after optimization of the C and Gamma parameters. The identification was first computed with spectra pretreated with an SNV, and then with an SNV and a 2nd derivative.

SVM function used	SVC Kernel		RBF	
Pretreatments	SNV	SNV + 2nd derivative	SNV	SNV + 2nd derivative
C parameter	1	100	10	10
Gamma parameter	N.A.	N.A.	15	15
Calibration set results (% of correct identification)	100	100	100	100
Test set results (% of correct identification)	100	100	78.4	100
Validation set results (% of correct identification)	100	100	75.9	100

Table 2
Results obtained for the KNN classification with spectra pretreated with an SNV, and with an SNV coupled with a 2nd derivative.

Pretreatments	SNV	SNV + 2nd derivative
K parameter	3	3
Calibration set results (% of correct identification)	100	100
Test set results (% of correct identification)	100	100
Validation set results (% of correct identification)	100	100

identification was obtained using spectra pretreated with either an SNV or an SNV coupled with a second derivative for the SVC Kernel linear, and using an SNV coupled with a second derivative for the RBF. All the validation spectra were also correctly recognized by the methods. The SVM methods using either the SVC Kernel linear or the RBF can be used for the authentication of the 29 product families. The optimization of the parameters is an important step for the correct identification of the spectra. The SVM is a forced classification, meaning that unknown spectra will inevitably be attributed a class name. In order to use the SVM methods in the routine for the identification of suspected counterfeits, additional parameters should be used to define the limit between genuine and counterfeit samples, e.g. a distance or a correlation [44].

3.3. Classification with the KNN

The KNN algorithm was tested on the 29 product families of the database, in order to obtain 29 classes. The use of this algorithm requires the optimization of the number of nearest neighbors, represented by the K parameter. After optimization, 3 was retained as the K parameter that would provide the most accurate classification. The results obtained for the identification of the spectra with the SNV and the SNV coupled with a second derivative is provided in Table 2. The KNN methods computed with either the SNV or the SNV coupled with a second derivative provided a correct identification for all the calibration, test and validation spectra. The KNN algorithm is consequently also suitable for the identification of the studied pharmaceutical tablets. Such as for the SVM, the KNN algorithm forces the classification, always attributing a class to unknown samples. An additional test, like a correlation, should then also be used in order to distinguish the genuine samples from the counterfeits and allow an easy pass/fail answer for the analysis of unknown samples.

3.4. Results and interpretation of the DA

The classification was performed on the product families, in order to obtain 29 classes. The calibration spectra were then used to compute the DA. The models were then tested against the validation spectra. Additionally a second validation was performed with the available counterfeits, generics and placebos of the 29 product families of the database.

The DA algorithm is also one that forces the classification. Here it was used in combination with MD, in order to define a limit above which a suspect sample would be considered a counterfeit. 29 DA models have been computed, using 10 principal components. The results of the DA are presented in Table 3. 100% of the spectra from the calibration and validation sets have been correctly identified. Moreover, 100% of all other product families were rejected by each individual model, meaning that there was no confusion between one product family and all the other product families. Also 100% of the counterfeits, generics and placebo have been rejected by the method.

4. Discussion

The study presented in this paper deals with a huge database of medicines. This is partly the novelty of this work and constituted a challenge for the identification. 29 different pharmaceutical product families were analysed, with a total of 53 different formulations. Different manufacturing sites were also included in the database, representing potential differences in the formulation. In case of addition of a new production site, and also in the case of a new product, the database would then be completed accordingly. Two instruments were also used for the measurement of the samples. The first aim was to build a common database to install on both spectrometers. In case of failure or maintenance, a support instrument would therefore be available. Furthermore measuring on two spectrometers means that a larger set of data was available and more variability could that way be introduced in the dataset. Another reason is that the transfer on new instruments is that way also simplified.

The identification method was successfully calibrated and validated by all the tested tools – the SVM, the KNN and the DA –, it means all the product families were correctly recognized. These methods are all supervised and provided the same outstanding results. The chemometric tool chosen for the routine analysis was the DA because an acceptance limit, the Mahalanobis distance, was introduced for an easy yes/no answer. Such a limit might be added on almost all the supervised methods. The DA was chosen in this study for its simplicity of use and implementation on the dedicated instrumentation. For each DA model of a product family, all the other product families were also tested in order to avoid any possible mismatch. Additionally all the counterfeits and generics of the second validation tests were tested and rejected by the method.

The advantages of NIRS for the identification of tablets are numerous. It is first non-destructive, which is of advantage when for instance only one suspect tablet is available for analysis. NIRS is then chemically specific. Very low dosage tablets might however need to be analysed with another analytical method in order to confirm the results. The combination with Raman can for instance be advised, since other information can be obtained like the presence of API peaks in the suspect sample. NIRS can then be used as a reliable first screening technique. A counterfeit can indeed be confirmed within a minute, which is of great advantage compared to classical methods like chromatography. Additionally NIR spectra provide information that give a first idea of the composition of the counterfeits, especially for tablets made of only a few components. APIs in high amount, sugar or talc are for example quite easy to detect with NIRS. Moreover using chemometric tools, a profiling of the counterfeits can be quickly made with NIRS. The grouping of counterfeits of similar chemical composition can then be observed,

Table 3
Results obtained for the DA classification. The pretreatments used for each product family are indicated in the table.

Product family	Pretreatments	Selected Mahalanobis distance	Calibration set results (% of correct identification)	Validation set results (% of correct identification)	2nd validation set results (% of correct identification = samples rejected)
1	SNV	4.7	100	100	n.a.[a]
2	SNV + 1st derivative	5.0	100	100	n.a.
3	SNV	5.0	100	100	n.a.
4	SNV + 1st derivative	5.0	100	100	100
5	SNV	5.0	100	100	n.a.
6	SNV	5.0	100	100	100
7	SNV + 1st derivative	3.6	100	100	n.a.
8	SNV	4.7	100	100	n.a.
9	SNV + 1st derivative	5.0	100	100	n.a.
10	SNV	3.6	100	100	n.a.
11	SNV + 1st derivative	5.0	100	100	n.a.
12	SNV	5.0	100	100	n.a.
13	SNV + 1st derivative	4.5	100	100	n.a.
14	SNV	5.0	100	100	n.a.
15	SNV	5.0	100	100	n.a.
16	SNV + 1st derivative	3.7	100	100	n.a.
17	SNV	5.0	100	100	n.a.
18	SNV	5.0	100	100	n.a.
19	SNV	3.2	100	100	100
20	SNV	5.0	100	100	100
21	SNV + 1st derivative	4.0	100	100	n.a.
22	SNV	4.3	100	100	100
23	SNV + 1st derivative	3.5	100	100	100
24	SNV	3.0	100	100	100
25	SNV	5.0	100	100	n.a.
26	SNV	5.0	100	100	n.a.
27	SNV	5.0	100	100	n.a.
28	SNV + 2nd derivative	2.8	100	100	100
29	SNV	5.0	100	100	100

[a] n.a.: no sample available for testing.

providing valuable data for forensic intelligence and law enforcement [27,54,55]. Furthermore NIR devices can be miniaturized and the market has recently been spread with new portable and handheld spectrometers that allow an even faster identification on the field [21,37]. A drawback that might be seen in these devices is that the spectra and the software available might be of lower quality compared to lab instruments. Methods developed on lab instruments might also be tricky to import on such devices, which would then need a customized calibration. However the use of such instruments, through their use on the field, would definitely help fasten the investigations.

5. Conclusion

Methods for the identification of pharmaceutical tablets by the means of NIRS have been presented in this paper in order to provide a quick investigation tool for counterfeit detection. 29 different pharmaceutical product families have been measured, with a total of 53 formulations, resulting in a huge database of spectra. The classification consisted in several challenges. First of all, some product families consist of different excipients depending on the dosage. Then the formulation of the excipients might be very close between two different product families. The complexity of the classification thus resulted in the heterogeneity of the classes and the similarity of spectra between different product families.

The spectra were investigated with several chemometric tools, the PCA, the KNN, the SVM and the DA. The applied PCA first provided interesting insights into the dataset, enabling an overview of underlying clusters of product families. Then the KNN, SVM and DA supervised methods allowed the calibration and validation of the spectra with an outstanding rate of 100% of correct answer. The example of the DA results was taken as a simple method for implementation in routine with a simple yes/no answer attributed thanks to a Mahalanobis distance acceptance criterion. The available counterfeits, generics and placebos were tested and rejected by the method, which confirmed the reliable use of the NIRS method for the analysis of suspected counterfeits of tablets.

Medicine counterfeiting is a crime that requires quick and reliable investigation tools. The faster the analysis, the more chances the investigators have to pursue the criminals. In this context, NIR spectroscopy is a helpful technique that has now long demonstrated to be quick, specific and non-destructive. Besides, more and more portable and handheld NIRS devices have emerged on the market, enabling fast and reliable on site analysis of suspect samples. Another advantage of NIRS is that the generated data can also enable the profiling of counterfeits, and consequently allow detect common sources of counterfeits. Such information could be included in forensic databases and definitely help make a step forward in the fight against medicine counterfeiting.

References

[1] Pharmaceutical Security Institute, Counterfeit Situation PSI Report, 2015 (accessed 11.04.16) http://www.psi-inc.org/,.
[2] SafeMedicines, Counterfeit Drug Incident Encyclopedia Partnership for Safe Medicines, 2009 (accessed 11.04.16) http://www.safemedicines.org/counterfeit-drug-incident-encyclopedia.html.
[3] WHO, Growing threat from counterfeit medicines, Bull. World Health Organ. 88 (2010) 241–320.
[4] O. Obi-Eyisi, A.I. Wertheimer, The background and history of counterfeit medicines, in: A.I. Wertheimer, P.G. Wang (Eds.), Counterfeit Medicines Volume I: Policy, Economics and Countermeasures, ILM, Publications, UK, 2012, pp. 1–17.
[5] R. Bate, Historical examples of dangerous medicines, in: R. Bate (Ed.), Phake—The Deadly World of Falsified and Substandard Medicines, AEI for Public Policy Research, Washington, D.C, 2012, pp. 11–24.
[6] IMPACT, Counterfeit Drugs Kill! WHO Brochure, 2008 (accessed 11.04.16) http://www.who.int/impact/FinalBrochureWHA2008a.pdf.

[7] F. Vanderdonck, Counterfeit medicines: a threat to public health and safety, Louvain Med. 126 (2007) 138–143.
[8] G.H. Schweim, Zytostatika-, Arzneimittelfälschungen und Internethandel, 2009 (accessed 11.04.16) http://www.dgim.de/portals/pdf/Schweim-F%C3%A4lschungen-08.pdf.
[9] K. Dégardin, Y. Roggo, P. Margot, Understanding and fighting the medicine counterfeit market, J. Pharm. Biomed. Anal. 87 (2014) 167–175.
[10] E. Przyswa, Counterfeit Medicines and Criminal Organisations, IRACM, Study Report, 2013 (accessed 11.04.16) http://www.iracm.comhttp://www.iracm.com/flip2/english/IRACM.html.
[11] Union des Fabricants, Counterfeiting and Terrorism Unifab Report, 2006 (accessed 11.04.16) http://www.unifab.com/images/Rapport-A-Terrorisme-2015_GB.pdf,.
[12] EAASM, Packaging Patient Protection—Recommendations For New Legislation to Combat Counterfeit Medicines, EAASM Report, 2009 (accessed 11.04.16) http://www.eaasm.eu/Counterfeit_medicines.
[13] G. Power, Anti-counterfeit Technologies for the Protection of Medicines WHO Report, 2008 (accessed 11.0416) http://www.who.int/impact/events/IMPACT-ACTechnologiesv3LIS.pdf.
[14] M. De Veij, A. Deneckere, P. Vandenabeele, D. de Kaste, L. Moens, Detection of counterfeit viagra® with raman spectroscopy, J. Pharm. Biomed. Anal. 46 (2008) 303–309.
[15] E. Deconinck, M. Canfyn, P.Y. Sacré, S. Baudewyns, P. Courselle, J.O. De Beer, A validated GC–MS method for the determination and quantification of residual solvents in counterfeit tablets and capsules, J. Pharm. Biomed. Anal. 70 (2012) 64–70.
[16] F.M. Fernandez, M.D. Green, P.N. Newton, Prevalence and detection of counterfeit pharmaceuticals: a mini review, Ind. Eng. Chem. Res. 47 (2008) 585–590.
[17] M.C. Gaudiano, A. Di Maggio, E. Antoniella, L. Valvo, P. Bertocchi, L. Manna, M. Bartolomei, S. Alimonti, A.L. Rodomonte, An LC method for the simultaneous screening of some common counterfeit and sub-standard antibiotics: validation and uncertainty estimation, J. Pharm. Biomed. Anal. 48 (2008) 303–309.
[18] J.E. Polli, S.W. Hoag, S. Flank, Comparison of authentic and suspect pharmaceuticals, Pharm. Technol. 33 (2009) 46–52.
[19] K. Dégardin, Y. Roggo, F. Been, P. Margot, Detection and chemical profiling of medicine counterfeits by raman spectroscopy and chemometrics, Anal. Chim. Acta 705 (2011) 334–341.
[20] W.C. Samms, Y.J. Jiang, M.D. Dixon, S.S. Houck, M. Mozayani, Analysis of alprazolam by DART-TOF mass spectrometry in counterfeit and routine drug identification cases, J. Forensic Sci. 56 (2011) 993–998.
[21] R. Bate, R. Tren, K. Hess, L. Mooney, K. Porter, Pilot study comparing technologies to test for substandard drugs in field settings, Afr. J. Pharm. Pharmacol. 3 (2009) 165–170.
[22] Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond, N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, J. Pharm. Biomed. Anal. 44 (2007) 683–700.
[23] R. Bate, Understanding the problem, in: R. Bate (Ed.), Phake—The Deadly World of Falsified and Substandard Medicines, AEI for Public Policy Research, Washington, D.C, 2012, pp. 11–24.
[24] K. Dégardin, Y. Roggo, Counterfeit analysis strategy illustrated by a case study, Drug Test. Anal. 8 (2016) 388–397.
[25] M.J. Vredenbregt, L. Blok-Tip, R. Hoogerbrugge, D.M. Barends, D. de Kaste, Screening suspected counterfeit viagra® and imitations of vagra® with near-infrared spectroscopy, J. Pharm. Biomed. Anal. 40 (2005) 840–849.
[26] J. Dubois, J.-C. Wolff, J.K. Warrack, J. Schoppelrei, E.N. Lewis, NIR chemical imaging for counterfeit pharmaceutical products analysis, Spectroscopy 22 (2007) 36–41.
[27] M.B. Lopes, J.C. Wolff, Investigation into Classification/Sourcing of suspect counterfeit heptodin™ tablets by near infrared chemical imaging, Anal. Chim. Acta 633 (2009) 149–155.
[28] C. Hu, Y. Feng, L. Ying, Review of the characteristics and prospects of near infrared spectroscopy for rapid drug-screening systems in China, J. Near Infrared Spectrosc. 23 (2015) 271–283.
[29] D. Ballabio, R. Todeschini, Multivariate classification for qualitative analysis, in: D.-W. Sun (Ed.), Infrared Spectroscopy for Food Quality Analysis and Control, Elsevier, London, UK, 2009, pp. 83–104.
[30] C. Gendrin, Y. Roggo, C. Collet, Pharmaceutical applications of vibrational chemical imaging and chemometrics: a review, J. Pharm. Biomed. Anal. 48 (2008) 533–553.
[31] P.Y. Sacré, E. Deconinck, T. De Beer, P. Courselle, R. Vancauwenberghe, P. Chiap, J. Crommen, J.O. De Beer, Comparison and combination of spectroscopic techniques for the detection of counterfeit medicines, J. Pharm. Biomed. Anal. 53 (2010) 445–453.
[32] F.E. Dowell, E.B. Maghirang, F.M. Fernandez, P.N. Newton, M.D. Green, Detecting counterfeit antimalarial tablets by near-infrared spectroscopy, J. Pharm. Biomed. Anal. 48 (2008) 1011–1014.
[33] M.M. Said, S. Gibbons, A.C. Moffat, M. Zloh, Near-infrared spectroscopy (NIRS) and chemometric analysis of Malaysian and UK paracetamol tablets: a spectral database study, Int. J. Pharm. (Amsterdam, Netherlands) 415 (2011) 102–109.
[34] S.H.F. Scafi, C. Pasquini, Identification of counterfeit drugs using near infrared spectroscopy, Analyst 126 (2001) 2218–2224.
[35] I. Storme-Paris, H. Rebiere, M. Matoga, C. Civade, P.A. Bonnet, M.H. Tissier, P. Chaminade, Challenging near Infrared spectroscopy discriminating ability for counterfeit pharmaceuticals detection, Anal. Chim. Acta 658 (2009) 163–174.
[36] O.Y. Rodionova, K.S. Balyklova, A.V. Titova, A.L. Pomerantsev, Quantitative risk assessment in classification of drugs with identical API content, J. Pharm. Biomed. Anal. 98 (2014) 186–192.
[37] A.J. O'Neil, R.D. Jee, G. Lee, A. Charvill, A.C. Moffat, Use of a portable near infrared spectrometer for the authentication of tablets and the detection of counterfeit versions, J. Near Infrared Spectrosc. 16 (2008) 327–333.
[38] EMEA, Guideline on the use of near infrared spectroscopy by the pharmaceutical industry and the data requirements for new submissions and variations, 2014 (accessed 11.04.16) http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2014/06/WC500167967.pdf.
[39] ICH, Validation of Analytical Procedures: Text and Methodology, Guideline Q2 (R1), 1996 (accessed 11.04.16) http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073384.pdf.
[40] H.W. Siesler, Y. Ozaki, S. Kawata, H.M. Heise, Near-Infrared Spectroscopy Principles Instruments Applications, Wiley-VCH, Weinheim, 2002.
[41] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley and Sons, Inc., USA, 2001.
[42] H. Martens, T. Naes, Multivariate Calibration, Biddles Ltd., Guildford and King's Lynn, Norway, 1989.
[43] H. Martens, M. Martens, Multivariate Analysis of Quality—An Introduction, John Wiley & Sons, Chichester, 2001.
[44] Y. Roggo, K. Dégardin, P. Margot, Identification of pharmaceutical tablets by raman spectroscopy and chemometrics, Talanta 81 (2010) 988–995.
[45] C.-W. Hsu, C.-C. Chang, C.-J. Lin, A Practical Guide to Support Vector Classification, 2010 (accessed 11.04.16) http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf.
[46] R. Burbidge, M. Trotter, B. Buxton, S. Holden, Drug design by machine learning: support vector machines for pharmaceutical data analysis, Comput. Chem. (Oxford) 26 (2001) 5–14.
[47] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE 13 (1967) 21–27.
[48] Thermo Nicolet, 2015. TQ Analyst Algorithms: Valpro for Antaris, in: Thermo Nicolet Industrial Solutions User Guide, USA, pp. 44–46.
[49] D. Ascher, P.F. Dubois, K. Hinsen, J. Hugunin, T. Oliphant, Numerical Python, 1999 (accessed 11.04.16) http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.476.2204&rep=rep1&type=pdf.
[50] T.E. Oliphant, Python for scientific computing, Comput. Sci. Eng. 1521–9615 (2007) 10–20.
[51] S. Van der Walt, S.C. Colbert, G. Varoquaux, The NumPy array: a structure for efficient numerical computation, Comput. Sci. Eng. 1521–9615 (2011) 22–30.
[52] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[53] J.D. Hunter, Matplotlib: a 2D graphics environment, Comput. Sci. Eng. 9 (2007) 90–95.
[54] F. Been, Y. Roggo, K. Dégardin, P. Esseiva, P. Margot, Profiling of counterfeit medicines by vibrational spectroscopy, Forensic Sci. Int. 211 (2011) 83–100.
[55] K. Dégardin, Y. Roggo, P. Margot, Forensic intelligence for medicine anti-counterfeiting, Forensic Sci. Int. 248 (2015) 15–32.