Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Anal. Methods*, 2024, **16**, 1611

Mitigating the impact of gelatin capsule variability on detection of substandard and falsified pharmaceuticals with near-IR spectroscopy†

Olatunde Awotunde, Jiaqi Lu, Jin Cai, Nicholas Roseboom, Sarah Honegger, Ornella Joseph, Alyssa Wicks, Kathleen Hayes and Marya Lieberman *

Portable NIR spectrometers are effective in detecting authentic pharmaceutical products in intact capsule formulations, which can be used to screen for substandard or falsified versions of those authentic products. However, the chemometric models are trained on libraries of authentic products, and are generally unreliable for detection of quality problems in products from outside their training set, even for products that are nominally the same active pharmaceutical ingredient and same dosage as products in the training set. As part of our research directed at developing better non-brand-specific strategies for pharmaceutical screening, we investigated the impact of capsule composition on NIR modeling. We found that capsule features like gelatin type, color, or thickness, give rise to a similar amount of variance in the NIR spectra as the type of API stored within the capsules. Our results highlight the efficacy of orthogonal projection to latent structures in mitigating the impacts of different types of capsules on the accuracy of NIR chemometric models for classification and regression analysis of lab-made samples. The models showed good performance for classification of field-collected doxycycline capsules as good or bad quality when an NIR-based % w/w metric was used, identifying five samples that were adulterated with talc. However, the % w/w was systematically underestimated, so when evaluating the capsules based on their absolute API content according to the monograph standard, the classification accuracy decreased from 100% to 70%. The underestimation was attributed to an unforeseen variability in the quantities and types of excipients present in the capsules.

Received 1st January 2024
Accepted 22nd February 2024

DOI: 10.1039/d4ay00001c

rsc.li/methods

1. Introduction

Near infra-red spectroscopy (NIR) is useful in analytical applications outside laboratory settings because it is cheap, non-destructive, fast, and versatile.[1] The affordability, portability and ease of use of small NIR spectrometers has further broadened the acceptance of this technology for post-market surveillance (PMS) of pharmaceutical dosage forms.[2–4] Regulatory standards for assessing the actual content of the API in a pill generally rely on high performance liquid chromatography (HPLC) assay, which is a costly laboratory technique.[5] NIR has the potential to quickly assess the identity and quantity of active ingredients in solid dosage forms, which could serve to flag suspicious pills for HPLC assay, enabling more effective use of HPLC resources. Because over 80% of pharmaceuticals are dispensed as solid dosage forms such as tablets or capsules,[6] NIR has been explored extensively to characterize the moisture content, particle size, polymorphisms, blend uniformity, concentration, porosity and packing density of solid forms of active pharmaceutical ingredients (APIs) and pharmaceutical excipients.[7–11] NIR can read through thin containers such as plastic bags, capsules, or tablet coatings.[12] However, the resulting NIR spectrum contains chemical information from the API, excipient(s) and the coating or packing material, and this affects data analysis.[13–16]

NIR spectra cannot be directly interpreted like IR spectra as evidence for the presence or absence of expected functional groups; instead, various chemometric models must be trained and validated to achieve qualitative identification or quantitative analysis of specific target analytes.[17–19] The training of these chemometric models usually requires access to libraries of authentic products, although some alternative approaches are feasible.[20,21] When authentic libraries of products are used, NIR has high specificity to discriminate samples based on minute differences in spectra.[14] Developing a unified database of NIR spectra for pharmaceutical dosage forms is a huge challenge because there are thousands of pharmaceutical manufacturers, each with products that are unique formulations for what are

Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA. E-mail: mlieberm@nd.edu; jlu22@nd.edu; jcai@nd.edu; nroseboo@nd.edu; shonegge@nd.edu; ojoseph2@nd.edu; awicks@nd.edu; khayes5@nd.edu

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ay00001c

nominally the same API and dose of a drug. While countries with strong regulatory systems are able to limit the influx of unregistered dosage forms into their markets, in other countries, inadequate monitoring and strong networks of informal or grey markets make this impossible.[22–25]

Gelatin capsules and tablet coatings are known sources of variability between products that are nominally the same API and dose, but are produced by different manufacturers. Mainka *et al.* studied captopril, hydrochlorothiazide, and sildenafil citrate mixed with various excipients in closed hard shell capsules made from gelatin or hypromellose *via* NIR; the different capsule materials were a major source of variability in the NIR spectra.[26] In a study of generic fluoxetine capsules and ciprofloxacin tablets, Storme-Paris *et al.* reported that NIRs picked up even slight differences in formulation, such as variations of 2.5% (w/w) in API or 1.0% (w/w) in excipient; even coating variations of less than 1% (w/w) of the pill mass were detected for tablets that otherwise had identical contents.[14] Rodionovaa *et al.* recognized the importance of investigating pill or capsule coatings before developing statistical models.[12] In a study by Caillet *et al.* various portable spectrometers, including NIR devices, were tested on both intact and crushed tablets in a simulated Laotian pharmacy environment. Results consistently showed distinct spectra between the two tablet states. Such spectral variations, influenced by the coatings, can introduce errors during field analysis.[27–30] For instance, an NIR device in that study misclassified augmentin (an amoxicillin and clavulanate combination) as roxithroxyl tablets even though they represent different active ingredients, probably because both types of tablets had the same type of coating.[31]

Capsules are made from solutions of gelatin (porcine, bovine) or vegetarian equivalents such as hypromellose with plasticizers such as glycerine or D-sorbitol, coloring agents (iron oxide pigments, dyes) and opacifying agents such as titanium dioxide.[32] The resulting liquid is formed into capsules and dried. In addition to variations in the composition of the capsules, the thickness and moisture content of the dried capsules can vary from batch to batch, and capsules may be printed or stamped with product names and logos. These manufacturing variations are all expected to affect the appearance of the NIR spectrum of the resulting capsules.[33] If NIR chemometric models were trained to recognize the authentic product in one type of capsule, subsequent batches of product packaged in other types of capsules could be falsely identified as substandard and falsified pharmaceuticals (SFPs).

Our overall goal is to develop a robust NIR approach that probes the quality of pharmaceutical products such as anti-infectives in field settings. Anti-infectives are among the most commonly reported substandard and falsified medical products.[34] We were interested in differentiating conforming formulations from substandard or falsified (non-conforming) ones *in situ* in capsules, simulating use of NIR in field settings. The chemical and physical variations in the NIR signal arising from different capsule materials and dilution with different types of excipients can be seen as a kind of experimental noise, which we sought to minimize through data pretreatment. NIR chemometric models commonly start with data pretreatment such as standard normal variate (SNV) and Savitszky–Golay (SG) methods to optimize prediction accuracy of machine learning algorithms.[10,35–37] We explored the unique strengths of these data pretreatments in addition to orthogonal projection to latent structures (O-PLS) in reducing interference from noise.[35,38] We generated data from a diffuse reflectance NIR device to explore the performance of different data pretreatments, then applied the pre-treatment methods to capsules filled with APIs and different diluents. Finally, we applied the regression algorithm to a set of doxycycline capsules collected in Kenya and Liberia, then compared the NIR predictions to the HPLC assay results.

2. Experimental

2.1 Materials

Transparent and uncolored size 00 gelatin capsules were purchased from NOW, Bloomingdale, IL. Transparent and uncolored size 3 gelatin capsules were purchased from **https://pharmacapsules.com**, Mound House, NV. These uncolored capsules are 100% bovine gelatin with no additives. Size 00 capsules of varying opacity and colors (Fig. S3†) were purchased from XPRS Nutra NV, West Jordan, UT. One of the Nutra capsules (the coral-colored capsule – third from left in Fig. S3†) was a vegetarian product made from hypromellose vegetable cellulose; the others were all manufactured from gelatin.[39] Isoniazid (INH) with purity >99%, microcrystalline cellulose, and α-lactose with purity >99% were obtained from Sigma Aldrich (St. Louis, MO). Doxycycline hydrate (DOX) with purity 98% was obtained from Alfa Aesar (Haverhill, MA).

2.2 Lab formulated mixtures and field collected dosage samples

Binary mixtures of isoniazid (INH) and microcrystalline cellulose were formulated as shown in Table S1† and binary mixtures of doxycycline x' and lactose were formulated as shown in Table S2.† To ensure homogeneity, the components were ground and mixed together for ~10 min in a clean mortar and pestle, then placed in scintillation vials and vortex-mixed for another 5 min.

23 samples of doxycycline dosage forms were collected in Western Kenya and Liberia from 2016 to 2021. The products were 100 mg oral dosage forms, mostly in the form of capsules packaged in blister packs, some with outer cardboard boxes. Samples were stored at 4 °C. Institutional review board (IRB) approval for sample collection *via* covert shoppers was provided through the University of Notre Dame protocols 17-11-4224 (exp. 2019), and 18-02-4442 (exp. 2026). The samples were analyzed by HPLC and NIR spectrometer.

2.3 High performance liquid chromatography analysis of the branded drugs

Assay of doxycycline dosage forms was performed using high-performance liquid chromatography (HPLC) based on a previously published method using a waters 2695 separations module equipped with a waters 2487 UV-vis detector.[40,41] Analysis was conducted isocratically at ambient temperature on an

XTerra C8 column (150 × 4.6 mm, 3.5 μm particle size) with 70 : 30 of 0.1% trifluoroacetic acid to acetonitrile at a flow rate of 1.0 mL min^{-1}. The injection volume was set to 20 μL and the analytical wavelength was 360 nm. Doxycycline eluted at 3.5 min. After completing method validation and establishing system suitability, finished doxycycline dosage forms were accurately weighed, crushed, and prepared at a nominal concentration of 0.5 mg mL^{-1} in 18 MΩ water for analysis.[41] All samples were analyzed before expiration. The assay results showed doxycycline content for single capsules ranged from 48% to 118% of the stated dose (Table S3†).

2.4 SEM and XRF analysis of empty capsules

Scanning electron microscope (SEM) images of capsules were acquired using a FESEM-Magellan 400 (FEI Company, Hillsboro, OR). Samples were examined at 5 kV and a working distance of 4 mm. In preparation for imaging, capsule samples were cut into flat strips and sputter coated with 2.5 nm of Ir. Sputter coating was performed with an 8004 Desktop High Resolution Coating System (Ted Pella, Redding, CA).

The gel capsules were tested using an X-200 handheld X-ray fluorescence (XRF) analyzer (SciAps, Woburn, MA). Prior to analysis, the gel capsules were cut open, flattened, and secured directly above the XRF window. On the XRF instrument, the 'soil' analysis mode was used, and the amount of each element was displayed in units of ppm.

2.5 NIR data acquisition

NIR spectra were acquired on a NIR-M-R2 USB-powered portable spectrometer using NIRScan Winform software (InnoSpectra Corporation, Hsinchu, Taiwan). Spectra were acquired by averaging 20 scans from 900 nm to 1700 nm with 228 data points (pattern width of 3.51 nm/data point, 20 second total acquisition time). Ten capsules of each color were each scanned ten times, with each spectrum generated after shaking and repositioning in the instrument cuvette holder, generating 100 NIR spectra of each type were generated. For the authentic doxycycline samples, single capsules from 9 different blister packs each scanned 10 times, each time, the capsule shaken and represented to the spectrometer.

We used a previously described capsule holder to house the size 00 capsules.[20] A new sample holder was designed using Solidworks® software (see file named 'New NIR Cuvette Model (5 mm Capsule).SLDPRT' in our GitHub page)[42] to accommodate the size 3 capsules used for the 100 mg doxycycline dosage forms. The opening aperture of the designed holder was aligned with the beam size of the spectrometer to minimize stray light as discussed in our earlier study (Fig. S9†).[43]

2.6 Processing and analysis of NIR spectra

NIR data files from NIRScan Winform with header sections containing sample metadata and 228 wavelength/absorption data points were merged and formatted for analysis using excel. The collated data files (ISCE_data_mg.csv, isce_conc.csv,DE_Studies_mg_.csv, DE_Studies_conc_.csv, external_set_doxy_conc.csv, external_set_doxy.csv, ALL_EXTERNAL_DE_CAPSULES.csv and ALL_EXTERNAL_DE_CAPSULES_cap.csv in our GitHub repository)[42] were uploaded to The Unscrambler X version 10.4 (Camo Software, Oslo, Norway) or JupyterLab for data preprocessing and analysis.

Standard normal variate (SNV) preprocessing ensures that each spectrum has a standard deviation of one and a mean of zero.[10,43,44] Savitzky–Golay filtering (SG) was employed for data smoothing.[46,47]

Orthogonal projection to latent structures (O-PLS) was used to eliminate unrelated orthogonal variations (noise) from the data. O-PLS explores orthogonal projections to suppress or delete uncorrelated variability in the spectra.[35,36] O-PLS was employed to filter out unrelated variations (or noise) from the dataset. O-PLS accentuates spectral features that correlate linearly with the target variable. In contrast, the orthogonal component captures the uncorrelated variation. The training spectral data underwent O-PLS transformation using its respective Y vector. For processing an unfamiliar spectrum, we utilized the orthogonal projection function from the pyopls library in Python. Specifically, the 'opls.transform' function facilitated orthogonal data transformation without requiring a Y vector.

Multi-variate data analysis—either support vector machine regression (SVM-R, (kernel = 'poly', gamma = 0.2, $C = 1$)) or partial least squares regression (PLS-R, number of latent variables = 7)—was next applied to model the data. For each data model, we initiated the process by randomly partitioning the dataset. In order to determine the optimal parameters, we varied the number of latent variables (nLVs) and determine which of these nLVs gave the best performance. Specifically, for the Python code 70% of the spectra were designated for model training, while the remaining 30% served as a test set to evaluate the model's performance post-training. For The UnscramblerX, cross-validation (leave-one-out) was employed. Manual or grid search was explored for optimization of the parameters.

To further ensure robustness in our evaluation, we incorporated K-fold cross-validation techniques for both support vector machine (SVM) and partial least squares regression (PLS-R) models. For K-fold cross-validation, the dataset is divided into 'n_splits' distinct subsets or "folds". The models are then trained on a combination of these folds and tested on the remaining fold, iteratively. This iterative process aids in gauging the model's consistency and reliability across diverse data subsets, presenting a more comprehensive evaluation than a singular train/test partition.

To optimize the performance of our models, we utilized the 'GridSearchCV' function from the scikit-learn library in Python. This approach allowed us to systematically explore a range of latent variables for both SVM and PLS-R models. The optimal model was chosen based on its performance metrics, particularly the cross-validated mean squared error.

All pivotal parameters, configurations, and code implementations pertinent to the SVM and PLS-R models have been made publicly accessible in our GitHub repository.[42]

The key parameters used for SG in UnscramblerX were second derivative, second order polynomial and 25 smoothing

points. The key parameters used for SVM-R in UnscramblerX were radial basis function as kernel type, *C*-value of 1, epsilon value of 0.1 and gamma function of 0.00438. The key parameters used for PLS-R in UnscramblerX includes algorithm – kernel, validation method – cross validation, and optimal number of latent variables of 7.

2.7 Data and code access

The open-source code as well as the spectra used in this study are available in our GitHub repository,[42] along with a guided workflow in the Jupyter notebook named 'minimizing capsule interference on NIR.ipynb'. Spectra of the formulations and the performance of the developed models before and after data pretreatments are displayed in our GitHub repository.[42]

The pairwise Euclidean distance served as a pivotal metric to quantify the dissimilarity between individual data points within our dataset. Originating from the classical Euclidean geometry, this distance measure calculates the straight-line distance between two points in a multidimensional space, offering a straightforward and intuitive method to assess the proximity or divergence of observations.[47] By computing the pairwise distances, we were able to discern patterns, identify clusters, and elucidate relationships among our samples, thereby facilitating a deeper understanding of the underlying structure and variability inherent in the data.

3. Results and discussion

This study was directed at the use of NIR for screening the quality of pharmaceutical dosage forms without the use of training data derived from authentic products. The goal is to quantify the API within any brand of the pharmaceutical. To accomplish this, we systematically formulated binary mixtures of the API and an excipient such as lactose or cellulose. These mixtures were then enclosed in different capsule types, and predictive models were developed from the resulting NIR spectra. We used these predictive NIR models to evaluate field collected samples, comparing the NIR predictions with HPLC quantification of the API in the field collected samples.

Initially, we generated NIR spectra from multiple samples of each distinctive empty capsule type, and then projected the spectra to principal component space to understand the spectral features of the diverse capsule compositions. Second, we prepared binary mixtures of APIs with excipients that were packaged in capsules of different colors. Pairwise Euclidean distance measurement was used to probe the significance of capsule color and composition. We evaluated the ability of the NIR models to classify the APIs correctly and to identify substandard lab-made formulations of isoniazid or doxycycline cut with multiple types of excipients and packaged in multiple types of gelatin capsules. Finally, we tested the models on authentic samples of doxycycline in multiple types of capsules.

3.1 NIR, SEM, and XRF characterization of capsules of varying opacities, compositions, and colors

The spectra generated from doxycycline in five different colored capsules showed distinct features, including peak shifts and changes in intensity, which suggested that capsule color would interfere with chemometric data analysis. Data pre-processing techniques including moving average smoothing, standard normal variate (SNV) and SNV followed by Savitzky–Golay filtering, SG (Fig. 1B–D respectively) were explored to minimize the variations in the spectra. Initially, SNV appeared to remove more variability than SNV + SG, but it was not clear if this would be a general result so the effects of a wider range of capsule colors on NIR spectra was explored using empty capsules.

NIR spectra were generated from fourteen empty size 00 capsules. For bicolor capsules, which include different dyes and opacifying agent compositions, spectra were generated from both the body (larger piece) and cap (smaller piece). Principal component (PC) analysis of the unprocessed NIR spectra revealed distinct clusters in PC1/PC2 space for each capsule color (Fig. 2A). NIR is known to be sensitive to sample morphology and composition (gelatin *vs.* hypromellose, presence of TiO_2, dyes, inks), so we next investigated capsule morphology and elemental composition.[48] Pharmaceutical capsules contain opaquing agents, such as TiO_2 powder, designed to scatter light, which can impact the signals from the NIR instrument.[2,7,9,16] Scanning electron microscopy with energy dispersive analysis (SEM EDX) and X-ray fluorescence revealed uneven elemental distribution on a scale of millimeters in the capsules (Fig. S2 and S3†). This could contribute to the observed variability in the NIR spectra, based on how the capsule is presented to the NIR instrument. However, analysis of the distance metric shows that this source of variability is insignificant compared to the variability from the different capsule compositions. X-ray fluorescence has been used to identify heavier elements in samples including pharmaceuticals,[49] and we found that all the pharmaceutical capsules contained titanium (0.5–2.0% w/w) along with small amounts of sulfur, calcium, and potassium (Fig. S3†). These varying elemental concentrations result from addition of different amounts of opaqueing agent needed to provide the desired capsule color and transparency, and from varying degrees of mineralization of collagen, the raw material that gelatin is derived from.[50] Assi *et al.* identified titanium(IV) oxide as one of the most common excipients in antibiotic products (generic and branded tablets as well as capsules) obtained from 11 different countries.[51] Kauffman *et al.* and Romero-Torres *et al.* found that capsules and coatings containing titanium(IV) oxide interfered with vibrational spectroscopic measurements of API content.[30,52,53] However, the clusters in PC1/PC2 space seen in Fig. 2 had no correlation with the TiO_2 content of the corresponding capsules.

3.2 Orthogonal projection to latent structures (O-PLS) minimizes undesired variability caused by different colored capsules

NIR can analyze samples inside some packaging such as vials that have relatively thick walls, so the impact of a thin gelatin capsule on the NIR spectrum might be small compared to the effect of the pharmaceuticals inside the gel capsule. The average capsule thickness was just 0.125 ± 0.015 mm. However, we

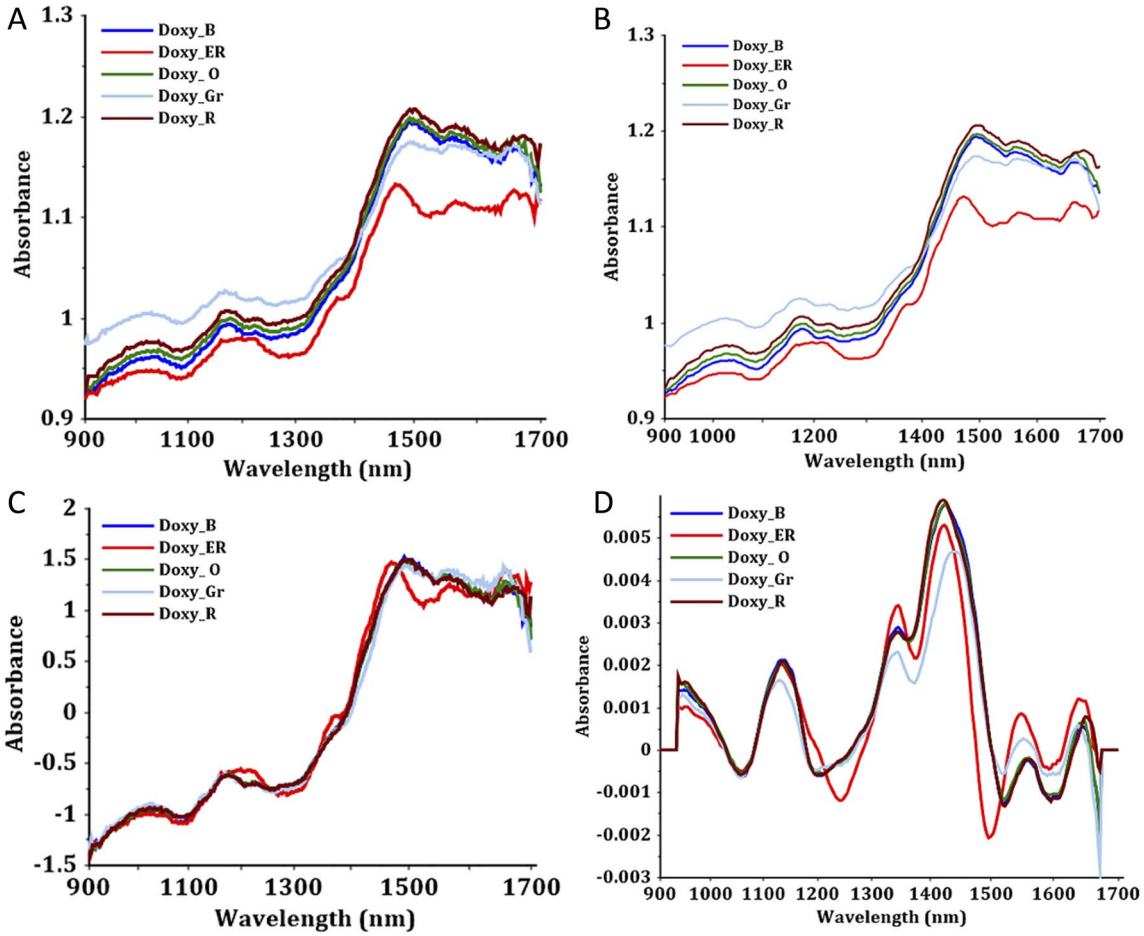

Fig. 1 NIR spectra of doxycycline in five colored capsules (blue gelatin (B), vegetable cellulose (ER), orange gelatin (O), green gelatin (Gr), and red gelatin (R)), (A) raw NIR spectra (B) moving average (C) standard normal variate (SNV), (D) SNV + Savitzky–Golay (SG).

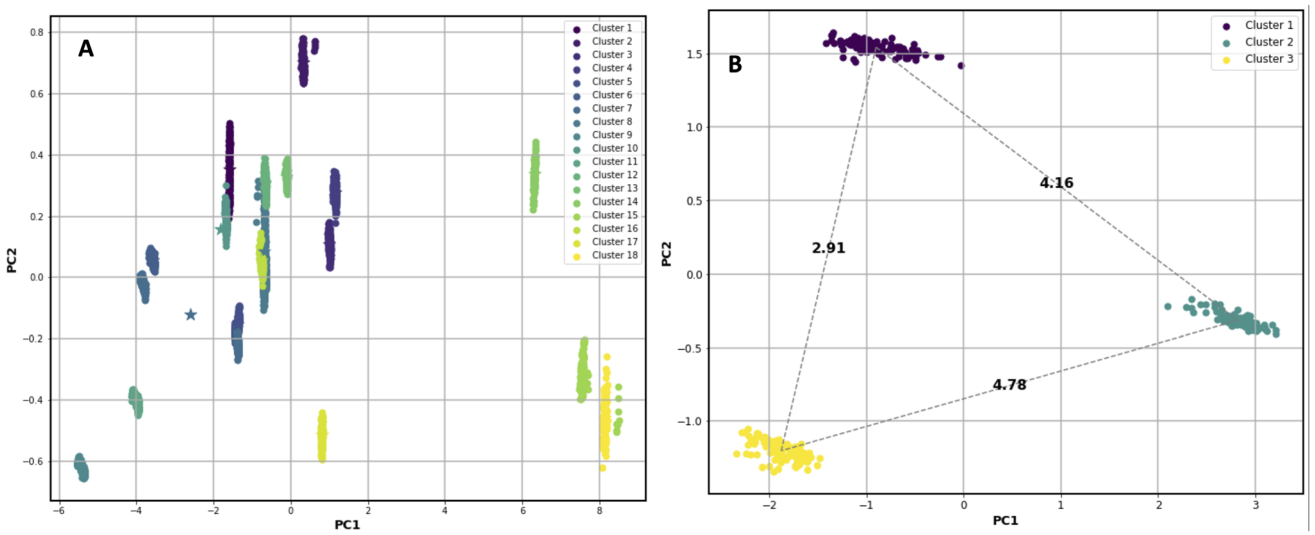

Fig. 2 (A) PCA scatter plot of the treated spectra of 18 empty capsule types showing some selected inter-cluster pairwise Euclidean distances (B) PCA scatter plot and pairwise Euclidean distances of NIR spectra from transparent gelatin capsules holding different APIs (10% acetaminophen/90% lactose denoted as cluster 1, and 10% acetaminophen/90% ascorbic acid, denoted as cluster 2, and 100% isoniazid denoted as cluster 3).

found that the capsule-related clustering was still detectable when the capsules were filled with various pharmaceutical ingredients. Fig. S4–S6† show PCA analysis of samples of isoniazid prepared in different colored capsules; the resulting spectra revealed distinct clusters for samples prepared in different capsules, even if the samples were of the same chemical composition, and even after SNV or SNV + SG transformation. This finding reflects the sensitivity of NIR not only to the actual content of the capsule but also to the container (capsule/pill coating) housing it.

We compared a common NIR data pre-treatment method – standard normal variate (SNV), followed by SG smoothing – with orthogonal projection to latent variables, which is less commonly used for pharmaceutical analysis.[45,54] O-PLS is an orthogonal projection approach that removes unwanted sources of variability from spectra,[35,36] by removing any features that are not correlated with the variation in the targets' concentrations. We compared the effects of SNV, SNV-SG, and O-PLS data pretreatment methods; in each case, the processed spectra were evaluated by PC cluster analysis. Minimization of the differences between the capsule compositions would result in a single cluster (aggregation of individual clusters). The raw spectra show nearly complete distinction of each type of capsule; SNV, SNV-SG and O-PLS each reduced the variability in the NIR data from different capsule compositions (Fig. S1(iii)†). The average distance metrics for the clusters were 2.40, 0.017, and 0.31, showing that SNV-SG and O-PLS were better pre-processing option.

3.3 O-PLS pre-treatment removes the interference of capsule type on classification of API type within the capsules

We employed the O-PLS pre-treatment technique on NIR spectra of several APIs enclosed within capsules of various colors. Through exploratory data analysis utilizing PCA, we observed that O-PLS data pre-processing effectively eliminated undesired sources of variability present in the spectra, allowing us to focus on the contribution stemming from the API itself (Fig. 3A and C). Given the differences in excipients and other formulation factors that can result in overlapping peaks, we refrained from employing a wavelength selection approach throughout this study.

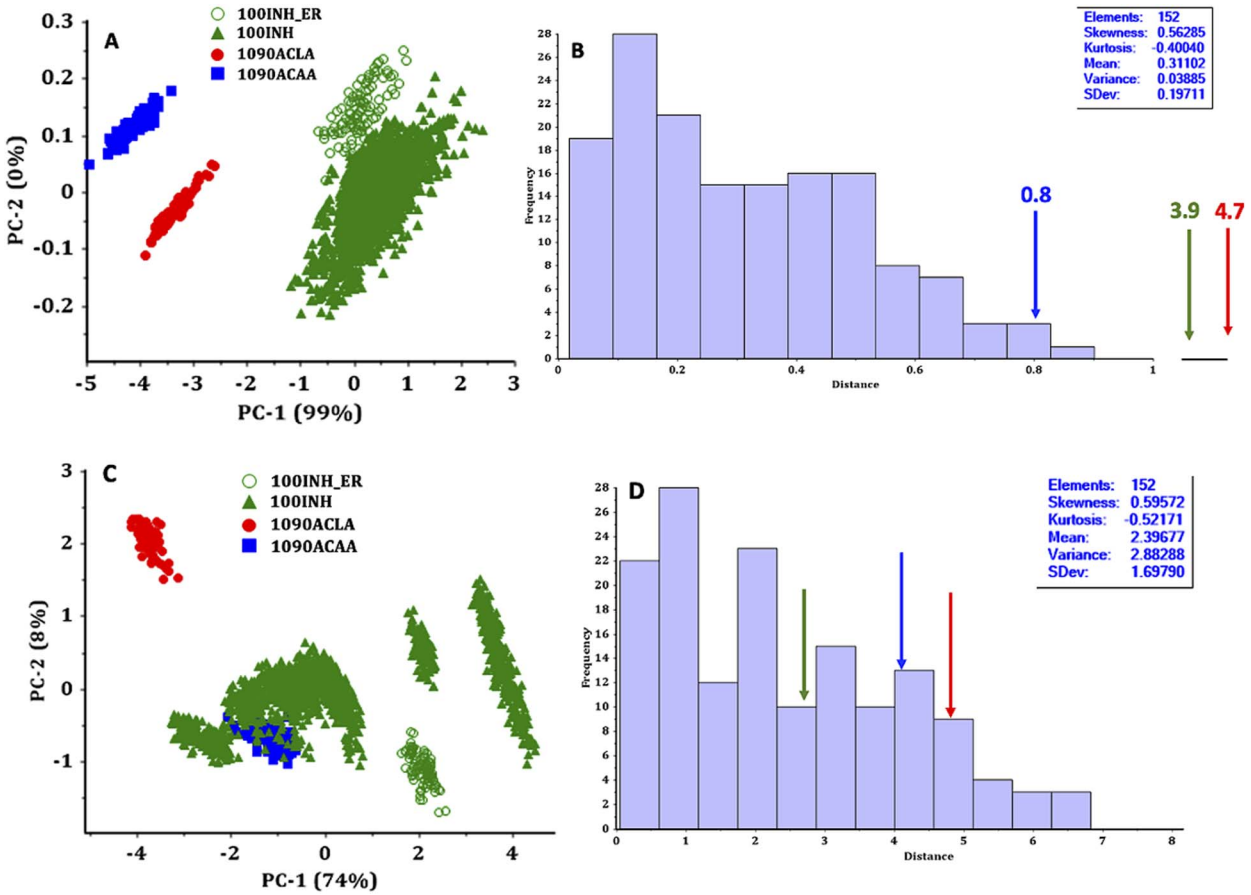

Fig. 3 PCA plots and Euclidian distances for lab formulations containing different APIs (green = pure isoniazid, red = 10% acetaminophen/90% lactose, blue = 10% acetaminophen/90% ascorbic acid). The isoniazid was housed in 18 different capsule types/colors; the other formulations were housed in transparent gelatin capsules. (A) PCA plot showing O-PLS treated data (B) histogram of the inter-cluster pairwise Euclidean distances for all the 18 capsule types (blue bars) after O-PLS treatment; arrows indicate pairwise Euclidean distances for the different APIs housed in transparent capsules. (C) PCA plot showing SNV treated spectra (D) histogram of the inter-cluster pairwise Euclidean distances for all the 18 capsule types (blue bars) after SNV treatment; arrows indicate pairwise Euclidean distances for the different APIs housed in transparent capsules.

In order to highlight the impact of capsule color on the NIR data, in contrast to variations in the API and excipient, we conducted two specific evaluations. We examined pure isoniazid samples in capsules of diverse colors, and we examined a single type of transparent capsule housing pure isoniazid or binary mixtures of acetaminophen with either lactose or ascorbic acid. Both O-PLS and SNV pre-processing successfully clustered the samples based on variations in chemical composition. After O-PLS pre-processing, the isoniazid NIR spectra were all well resolved in PC space from the acetaminophen/lactose and acetaminophen/ascorbic acid NIR spectra. The isoniazid samples exhibited two primary clusters: one corresponding to gelatin capsules of various colors, and the other associated with vegetable cellulose capsules (Fig. 3A). However, SNV pre-processing resulted in distinct clusters within the isoniazid samples (Fig. 3C) due to variations in the capsule composition, and the acetaminophen/ascorbic acid samples overlapped with several of the isoniazid samples.

The pairwise Euclidean distance probes how much the capsule color and composition affect the clustering of the NIR spectra. It quantifies the distance between two data points (or centroids of the respective clusters) by finding the square root of the sum of their squared differences in PC coordinates; it should be noted that the distance is calculated in the first two components.[55] Fig. 3 shows histogram plots of these distances for all the different capsule color types housing isoniazid; after O-PLS pre-treatment, the average distance metric ranges from 0.0126 to 4.97, much smaller (Fig. 3B) than the range of 0.019–7.82 seen after SNV pre-treatment (Fig. 3D). In the SNV pre-treated data, the Euclidian distance metric between the isoniazid, acetaminophen/lactose and acetaminophen/ascorbic acid clusters averaged 3.95, which is in the range of the distance metrics seen for different types of capsules; this means that for SNV pre-treated data, the capsule type is as important a source of variability as the contents of the capsules. In the O-PLS pre-treated data, the Euclidian distance metric between the isoniazid, acetaminophen/lactose and acetaminophen/ascorbic acid clusters averaged 3.12, while the distance metric between all the clusters from different capsule types was below 1.0. This shows that O-PLS pre-treatment removes most of the variability due to capsule type without impacting variability due to capsule contents. The efficacy of the O-PLS pre-treatment here is consistent with the results for the empty capsules (Fig. S5–S7†).

3.4 O-PLS data pre-treatment removes interference from capsule type on quantification of the API

Pharmaceutical dosage forms must contain a specified quantity of API, so we developed an NIR regression model for use in

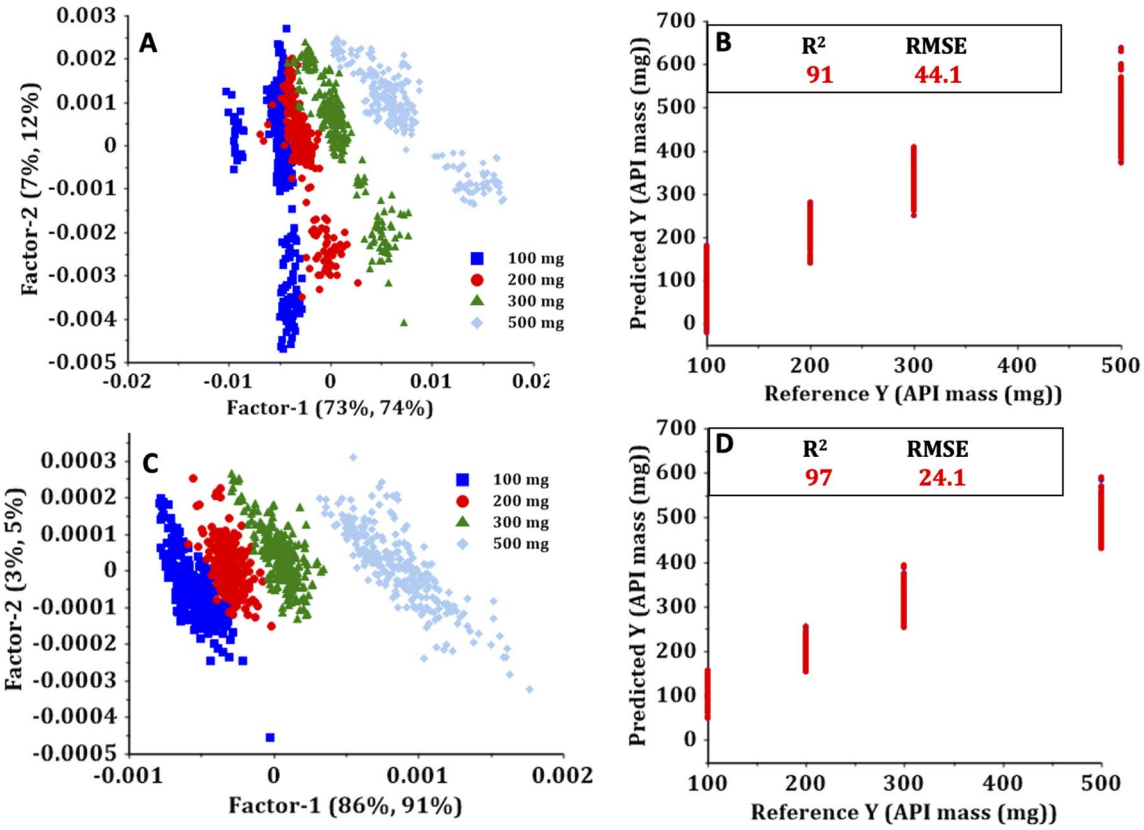

Fig. 4 Partial least squares regression model trained with lab-formulated mixtures of isoniazid and microcrystalline cellulose in multicolor capsules can predict API concentration of powders presented in different colored capsules; all NIR data was pre-treated by either SNV-SG (A and B) or O-PLS (C and D). INH_100, INH_200, INH_300 & INH_500 are 14%, 28%, 42% and 71% isoniazid respectively. Panels (A) and (C) show principal component projections of the NIR data, panels (B) and (D) show both calibration and validation performance.

detection of substandard dosage forms. The regression model was trained on mixtures of isoniazid and crystalline cellulose at different % w/w ratios (Table S1†). Mirroring the findings observed in the pure isoniazid samples displayed in Fig. 3, distinctive clusters corresponding to different capsules as well as different % w/w ratios were evident in the PC1/PC2 projection of the data (Fig. S7†). A regression model based on the untreated data yielded an R^2 value of 0.89 (data not shown) which was deemed low.

Data pre-treatment improved the regression models. Both the standard normal variate followed by Savitzky–Golay (SNV-SG) and orthogonal projection to latent structures (O-PLS) data pre-treatments mitigated the heterogeneity within the data. Consequently, the resulting clusters became primarily separated by the % API w/w ratio rather than by capsule color (Fig. 4A and C). Partial least squares regression (PLS-R) and support vector machine regression (SVM-R) models were trained on the SNV-SG or O-PLS treated data and cross-validated (Fig. 4B and D). The O-PLS pre-treated data exhibited higher correlation coefficient values for validation (0.97) compared to the SNV-SG treated data (0.91). The root mean square error (RMSE) for validation, a metric gauging the accuracy of quantitative predictions, also indicated a significant improvement for the O-PLS pre-treated data. The RMSE of the PLS-R model was twice as high for SNV-SG pre-treated data as for O-PLS pre-treated data (Fig. 4). Similar performance trends were observed in the SVM-R models. Additional details and the Python version of this work can be found in the Jupyter notebook and our GitHub repository.[42]

We replicated the aforementioned procedures using doxycycline, which is commonly formulated in capsules. Regression models were trained using NIR spectra derived from laboratory mixtures of doxycycline and lactose, as outlined in Table S2.† Promisingly, the regression performance of these models (Fig. S8†) were comparable to those observed in the isoniazid models.

We applied the doxycycline model, trained on lab-made mixtures to a group of commercial doxycycline capsules obtained from Kenya and Liberia. Our goal was to see whether the model trained on 14%, 28%, 42% and 71% mixtures of doxycycline in a lactose matrix could predict doxycycline content in commercial products. This is a challenging test because the commercial products come in smaller capsules with different colors (Fig. S4†) and mostly used microcrystalline cellulose (rather than lactose) as an excipient.

3.5 There are large variations in the solid concentration of doxycycline packaged in commercial capsules that meet the assay standard for API content

When we tried to validate the method using doxycycline dosage forms, we encountered a new problem. Pharmacopoeial standards governing API content are based on the cumulative amount of API present in the dosage unit, with different manufacturers employing varying quantities of diverse excipients when formulating these units. Excipients, inactive ingredients within the formulation, contribute substantially to the total mass of the pill.[56,57] The wide range of potential identities and concentrations of excipients is an inherent limitation to the utility of NIR spectrophotometers for analysis of real dosage forms, particularly in low- and middle-income country (LMIC) settings.

For field screening of pharmaceutical dosage forms, the goal is to confirm the identity and amount of the API present in each capsule. However, due to the limited penetration depth, NIR does not measure the absolute amount of material in a capsule—it measures the concentration of the API, usually in units of %w/w. The assessment of API amount depends on the total weight of powder inside the capsule and the weight percentage of the API in that powder (% w/w). For on-site evaluations, the weight of powder can be measured using an affordable milligram balance.[58]

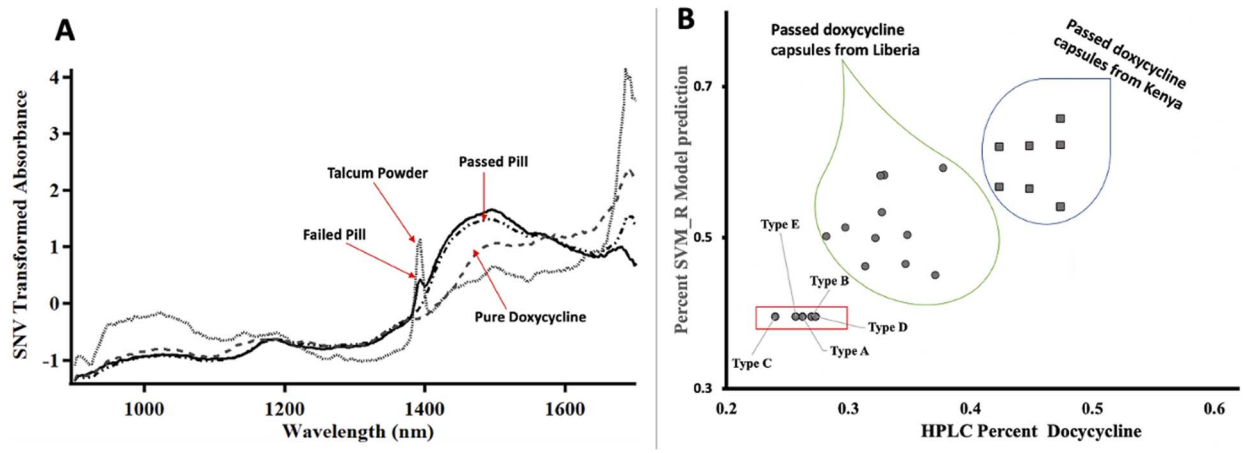

Fig. 5 (A) NIR spectra of two branded samples of doxycycline (one that failed HPLC assay and one that passed HPLC assay), pure doxycycline, and talcum powder. (B) Scatter plot showing actual and predicted % w/w doxycycline in twenty-three commercial samples of doxycycline. The samples in the red box all failed HPLC assay and were found to be adulterated with talc.

In order to evaluate the likely range of API concentrations that the NIR method might encounter, we used the monograph limits for API amount (90–120 mg for a nominal 100 mg dose of doxycycline) and the measured amount of powder in doxycycline capsules from 23 brands to calculate the range of doxycycline content that would still pass assay. The capsules were made by six manufacturers in three different countries (India, Germany, and China) and all were purchased from retail pharmaceutical stores in Africa. The upper and lower limit of these 100 mg pill masses were 322 mg and 154 mg. For these brands, the API concentration of the powder could in theory range from 28% to 62% w/w without violating the monograph standard for API content.[59] The actual doxycycline content in the capsules (as calculated from HPLC assay values and pill masses) ranged from 31 to 51% (Table S3†), except for 5 samples that did not pass the HPLC assay and had doxycycline content between 24 and 27%.

When the NIR models were applied to these 23 field-collected samples, the predicted concentration of doxycycline in % w/w (Fig. 5) ranged from 20.3% to 41.6%. Using the standard that good doxycycline pills should have %w/w API between 28% and 62%, the five bad quality samples failed while the 18 good quality samples passed the NIR screening evaluation. This shows that the HPLC-derived %w/w of API in field collected samples can provide a useful benchmark for non-brand specific NIR models. However, the NIR-estimated values are systematically underpredicted, probably because the microcrystalline cellulose and other excipients used in the field-collected doxycycline samples were too different from the lactose used to train the NIR regression models. Pharmaceutical monographs regulate the absolute quantity of API present in the dosage units, not the %w/w concentration of the API. When we used the NIR prediction of % w/w and the amount of powder in each pill to predict the actual mass of the doxycycline in the 23 dosage forms, the API mass was systematically underestimated. Only 48% of the 23 samples passed the total API content evaluation by NIR, while according to the "gold standard" HPLC assay, 78% of the 23 samples passed. This problem is likely to be more serious for dosage forms, like these doxycycline capsules, that include relatively large amounts of excipients.

4. Conclusion

The O-PLS pre-treatment demonstrated significant efficacy in mitigating capsule interferences within our lab-formulated samples. Its seamless integration into the NIR workflow and its potential for analyzing field samples underscore its strengths. However, a notable limitation was its inability to account for excipient interferences, resulting in the underestimation of API content in field samples. Monographs do not specify the exact excipients or the quantity of excipients to be used in a formulated product, except for requiring use of pharmaceutical grade excipients. Consequently, manufacturers use different excipients and different quantities of excipients in products that are nominally the same drug at the same dosage level. This variability will be a problem for any NIR models that are not trained on authentic samples. While our models exhibited promising accuracy in distinguishing between genuine and substandard pharmaceuticals, the vast variability in excipient concentrations and formulations among commercial products presented challenges.[60] These complexities, especially in the context of the capsule's total weight and the resultant % w/w API, emphasize the need for further research and refinement. Our future efforts will focus on determining how to select excipients for the lab-based training models that will give more robust models for analysis of field collected samples. Hyperspectral imaging, though effective, is cost-prohibitive for LMICs.[61–64] While using a transmission mode might mitigate coating interference, many portable devices do not support this feature, and it requires crushing the pills, which is difficult in a field screening scenario. This problem could also impact NIR models that are trained on authentic samples if the manufacturers alter the excipient composition or content, and the model is not updated.

Author contributions

Olatunde Awotunde designed the NIR experiments, performed HPLC analysis of some of the samples, was responsible for all the data analysis and preparation of all plots and graphics, and drafted the manuscript with the exception of SEM-EDX, XRF and HPLC experimental methodologies. Kathleen Hayes and Sarah Honegger performed HPLC analysis of some of the samples. Jiaqi Lu, Jin Cai and Nicholas Roseboom generated NIR spectra. Jin Cai and Nicholas Rosenboom designed 3D-printed instrument cases and sample holders for the NIR spectrophotometer. Alyssa Wicks collected the SEM-EDX data. Ornella Joseph collected and analyzed X-ray fluorimetry data and drafted the XRF experimental methodology section. Marya Lieberman conceived the experimental design, edited the manuscript, and was responsible for overall supervision and funding of this work.

Conflicts of interest

The authors do not have any conflict of interest in this work.

Acknowledgements

The authors would like to thank Rachel Roller providing access to materials used for this work. This work has also benefited from the facilities available through the Notre Dame Integrated Imaging Facility (NDIIF) and Center for Environmental Science and Technology (CEST). This work was supported by NSF (CMMI-1842369, OISE-1559497). OA received funding from the University of Notre Dame (Pamoja summer award, fellowship through Lucy Institute for Data and Society, and received partial support through the American Chemical Society Fall 2022 Pfizer Travel Award.

References

1 J. Luypaert, D. L. Massart and Y. Vander Heyden, Near-infrared spectroscopy applications in pharmaceutical

analysis, *Talanta*, 2007, **72**(3), 865–883, DOI: 10.1016/j.talanta.2006.12.023.

2 M. M. Said, S. Gibbons, A. C. Moffat and M. Zloh, Near-infrared spectroscopy (NIRS) and chemometric analysis of malaysian and UK paracetamol tablets: a spectral database study, *Int. J. Pharm.*, 2011, **415**, 102–109, DOI: 10.1016/j.ijpharm.2011.05.057.

3 R. Martino, M. Malet-Martino, V. Gilard and S. Balayssac, Counterfeit drugs : analytical techniques for their identification, *Anal. Bioanal. Chem.*, 2010, **398**, 77–92, DOI: 10.1007/s00216-010-3748-y.

4 E. Ziémons, J. Mantanus, P. Lebrun, E. Rozet, B. Evrard and P. Hubert, Acetaminophen determination in low-dose pharmaceutical syrup by NIR spectroscopy, *J. Pharm. Biomed. Anal.*, 2010, **53**(3), 510–516, DOI: 10.1016/j.jpba.2010.06.003.

5 M. Sendanyoye, Validation of HPLC-UV method for determination of amoxicillin trihydrate in capsule, *Ann. Adv. Chem.*, 2018, 055–072, DOI: 10.29328/journal.aac.1001014.

6 B. M. Couillaud, P. Espeau, N. Mignet and Y. Corvis, State of the art of pharmaceutical solid forms: from crystal property issues to nanocrystals formulation, *ChemMedChem*, 2019, **14**(1), 8–23, DOI: 10.1002/cmdc.201800612.

7 A. J. O'Neil, R. D. Jee and A. C. Moffat, Measurement of the percentage volume particle size distribution of powdered microcrystalline cellulose using reflectance near-infrared spectroscopy, *Analyst*, 2003, **128**(11), 1326–1330, DOI: 10.1039/b307263k.

8 M. Razuc, A. Grafia, L. Gallo, M. V. Ramírez-Rigo and R. J. Romañach, Near-infrared spectroscopic applications in pharmaceutical particle technology, *Drug Dev. Ind. Pharm.*, 2019, **45**(10), 1565–1589, DOI: 10.1080/03639045.2019.1641510.

9 A. B. Garmarudi, M. Khanmohammadi, N. Khoddami and K. Shabani, Near infrared spectrometric analysis of titanium dioxide nano particles for size classification, in *IEEE International Conference on Nanotechnology*, 2010, pp. 451–453, DOI: 10.1109/NANO.2010.5697778.

10 M. Blanco, J. Coello, H. Iturriaga, S. Maspoch and J. Pagès, NIR calibration in non-linear systems: different PLS approaches and artificial neural networks, *Chemom. Intell. Lab. Syst.*, 2000, **50**(1), 75–82, DOI: 10.1016/S0169-7439(99)00048-9.

11 A. U. Vanarase, M. Alcalà, J. I. Jerez Rozo, F. J. Muzzio and R. J. Romañach, Real-time monitoring of drug concentration in a continuous powder mixing process using NIR spectroscopy, *Chem. Eng. Sci.*, 2010, **65**(21), 5728–5733, DOI: 10.1016/j.ces.2010.01.036.

12 O. Y. Rodionova, A. V. Titova, K. S. Balyklova and A. L. Pomerantsev, Detection of counterfeit and substandard tablets using non-invasive NIR and chemometrics – A conceptual framework for a big screening system, *Talanta*, 2019, **205**, 120150, DOI: 10.1016/j.talanta.2019.120150.

13 T. Rajalahti and O. M. Kvalheim, Multivariate data analysis in pharmaceutics: a tutorial review, *Int. J. Pharm.*, 2011, **417**(1–2), 280–290, DOI: 10.1016/j.ijpharm.2011.02.019.

14 I. Storme-Paris, H. Rebiere, M. Matoga, C. Civade, P. A. Bonnet, M. H. Tissier and P. Chaminade, Challenging near infrared spectroscopy discriminating ability for counterfeit pharmaceuticals detection, *Anal. Chim. Acta*, 2010, **658**(2), 163–174, DOI: 10.1016/j.aca.2009.11.005.

15 M. C. Sarraguça, A. V. Cruz, H. R. Amaral, P. C. Costa and J. A. Lopes, Comparison of different chemometric and analytical methods for the prediction of particle size distribution in pharmaceutical powders, *Anal. Bioanal. Chem.*, 2011, **399**(6), 2137–2147, DOI: 10.1007/s00216-010-4230-6.

16 M. Blanco and M. Alcalá, Content uniformity and tablet hardness testing of intact pharmaceutical tablets by near infrared spectroscopy: a contribution to process analytical technologies, *Anal. Chim. Acta*, 2006, **557**(1–2), 353–359, DOI: 10.1016/j.aca.2005.09.070.

17 N. Fuenffinger, S. Arzhantsev and C. Gryniewicz-Ruzicka, Classification of ciprofloxacin tablets using near-infrared spectroscopy and chemometric modeling, *Appl. Spectrosc.*, 2017, **71**(8), 1927–1937, DOI: 10.1177/0003702817699624.

18 Y. Tie, C. Duchateau, S. Van de Steene, C. Mees, K. De Braekeleer, T. De Beer, E. Adams and E. Deconinck, Spectroscopic techniques combined with chemometrics for fast on-site characterization of suspected illegal antimicrobials, *Talanta*, 2020, **217**, 121026, DOI: 10.1016/j.talanta.2020.121026.

19 K. B. Beć, J. Grabska and C. W. Huck, Principles and applications of miniaturized near-infrared (NIR) spectrometers, *Chem.–Eur. J.*, 2021, 1514–1532, DOI: 10.1002/chem.202002838.

20 O. Awotunde, N. Roseboom, J. Cai, K. Hayes, R. Rajane, R. Chen, A. Yusuf and M. Lieberman, Discrimination of substandard and falsified formulations from genuine pharmaceuticals using NIR spectra and machine learning, *Anal. Chem.*, 2022, **94**(37), 12586–12594, DOI: 10.1021/acs.analchem.2c00998.

21 A. Eustaquio, M. Blanco, R. D. Jee and A. C. Moffat, Determination of paracetamol in intact tablets by use of near infrared transmittance spectroscopy, *Anal. Chim. Acta*, 1999, **383**, 283–290.

22 A. K. Deisingh, Pharmaceutical counterfeiting, *Analyst*, 2005, 271–279, DOI: 10.1039/b407759h.

23 H. Zhang, D. Hua, C. Huang, S. K. Samal, R. Xiong, F. Sauvage, K. Braeckmans, K. Remaut and S. C. De Smedt, Materials and technologies to combat counterfeiting of pharmaceuticals: current and future problem tackling, *Adv. Mater.*, 2020, **32**(11), 1905486, DOI: 10.1002/adma.201905486.

24 U.S. Government Publishing Office, *Counterfeit Drugs: Fighting Illegal Supply Chains*, 2015, https://www.govinfo.gov/content/pkg/CHRG-113hhrg88828/html/CHRG-113hhrg88828.htm.

25 R. Cockburn, P. N. Newton, E. K. Agyarko, D. Akunyili and N. J. White, The global threat of counterfeit drugs: why

industry and governments must communicate the dangers, *PLoS Med.*, 2005, **2**(4), 0302–0308, DOI: 10.1371/journal.pmed.0020100.

26 D. L. Mainka and A. Link, Near-infrared spectroscopic identification and quantification of active pharmaceutical ingredients in closed capsules: a feasibility study for pediatric doses, *Anal. Methods*, 2019, **11**(40), 5185–5194, DOI: 10.1039/c9ay01241a.

27 C. Caillet, S. Vickers, S. Zambrzycki, F. M. Fernández, V. Vidhamaly, K. Boutsamay, P. Boupha, P. Peerawaranun, M. Mukaka and P. N. Newton, A comparative field evaluation of six medicine quality screening devices in laos, *PLoS Negl. Trop. Dis.*, 2021, **15**(9), 1–22, DOI: 10.1371/journal.pntd.0009674.

28 C. Caillet, S. Vickers, V. Vidhamaly, K. Boutsamay, P. Boupha, S. Zambrzycki, N. Luangasanatip, Y. Lubell, F. M. Fernandez and P. N. Newton, Evaluation of portable devices for medicine quality screening: lessons learnt, recommendations for implementation, and future priorities, *PLoS Med.*, 2021, **18**(9), 1–12, DOI: 10.1371/journal.pmed.1003747.

29 C. Caillet, S. Vickers, S. Zambrzycki, N. Luangasanatip, V. Vidhamaly, K. Boutsamay, P. Boupha, Y. Lubell, F. M. Fernández and P. N. Newton, Multiphase evaluation of portable medicines quality screening devices, *PLoS Negl. Trop. Dis.*, 2021, **15**(9), 1–6, DOI: 10.1371/journal.pntd.0009287.

30 S. C. Zambrzycki, C. Caillet, S. Vickers, M. Bouza, D. V. Donndelinger, L. C. Geben, M. C. Bernier, P. N. Newton and F. M. Fernández, Laboratory evaluation of twelve portable devices for medicine quality screening, *PLoS Negl. Trop. Dis.*, 2021, **15**(9), 1–19, DOI: 10.1371/journal.pntd.0009360.

31 S. Vickers, M. Bernier, S. Zambrzycki, F. M. Fernandez, P. N. Newton and C. Caillet, Field detection devices for screening the quality of medicines: a systematic review, *BMJ Global Health*, 2018, **3**, e000725, DOI: 10.1136/bmjgh-2018-000725.

32 L. L. Augsburger and S. W. Hoag, *Pharmaceutical Dosage Forms: Capsules*, 2006, vol. 1999.

33 X. Wu, S. Yin, Q. Dong, B. Liu, Y. Wang, T. Sekino, S. W. Lee and T. U. V. Sato, Visible and near-infrared lights induced NO_x destruction activity of (Yb,Er)-NaYF4/C-TiO_2 composite, *Sci. Rep.*, 2013, **3**, 1–8, DOI: 10.1038/srep02918.

34 S. Ozawa, D. R. Evans, S. Bessias, D. G. Haynie, T. T. Yemeke, S. K. Laing and J. E. Herrington, Prevalence and estimated economic burden of substandard and falsified medicines in low- and middle-income countries: a systematic review and meta-analysis, *JAMA Netw. Open*, 2018, **1**(4), 1–22, DOI: 10.1001/jamanetworkopen.2018.1662.

35 M. Blanco, J. Coello, I. Montoliu and M. A. Romero, Orthogonal signal correction in near infrared calibration, *Anal. Chim. Acta*, 2001, **434**(1), 125–132, DOI: 10.1016/S0003-2670(01)00820-0.

36 R. M. Balabin, E. I. Lomakina and R. Z. Safieva, Neural network (ANN) approach to biodiesel analysis: analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy, *Fuel*, 2011, **90**(5), 2007–2015, DOI: 10.1016/J.FUEL.2010.11.038.

37 H. Wu, M. Tawakkul, M. White and M. A. Khan, Quality-by-design (QbD): an integrated multivariate approach for the component quantification in powder blends, *Int. J. Pharm.*, 2009, **372**(1–2), 39–48, DOI: 10.1016/j.ijpharm.2009.01.002.

38 R. M. Balabin and S. V. Smirnov, Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data, *Anal. Chim. Acta*, 2011, **692**(1–2), 63–72, DOI: 10.1016/J.ACA.2011.03.006.

39 H. Sohi, Y. Sultana and R. K. Khar, Taste masking technologies in oral pharmaceuticals: recent developments and approaches, *Drug Dev. Ind. Pharm.*, 2004, **30**(5), 429–448, DOI: 10.1081/DDC-120037477.

40 M. Lieberman, N. Myers, M. Berta, S. Bliese, K. Hayes, M. Wilfinger and O. Awotunde, *HPLC Methodology Manual for Distributed Pharmaceutical Analysis Laboratory (DPAL)*, South Bend, 2023.

41 S. L. Bliese, M. Berta and M. Lieberman, Involving students in the distributed pharmaceutical analysis laboratory: a citizen-science project to evaluate global medicine quality, *J. Chem. Educ.*, 2020, **97**(11), 3976–3983, DOI: 10.1021/acs.jchemed.0c00904.

42 O. Awotunde, J. Lu, J. Cai, O. Joseph, A. Wicks and M. Lieberman, *Mitigating Gelatin Capsule Impact on Detection of Substandard and Falsified Pharmaceuticals with Near-IR Spectroscopy*, https://github.com/aawotund/Handling-the-interference-effects-of-gelatin-capsule-composition-and-excipient-concentration-on-dete, accessed 2023 Auguest 25.

43 C. Pasquini, Near infrared spectroscopy: fundamentals, practical aspects and analytical applications, *J. Braz. Chem. Soc.*, 2003, **14**(2), 198–219, DOI: 10.1590/S0103-50532003000200006.

44 Y. Roggo, P. Chalus, L. Maurer, C. Lema-Martinez, A. Edmond and N. Jent, A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies, *J. Pharm. Biomed. Anal.*, 2007, **44**(3 SPEC. ISS.), 683–700, DOI: 10.1016/j.jpba.2007.03.023.

45 K. Heil and U. Schmidhalter, An evaluation of different nir-spectral pre-treatments to derive the soil parameters c and n of a humus-clay-rich soil, *Sensors*, 2021, **21**(4), 1–24, DOI: 10.3390/s21041423.

46 C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives – A review, *Anal. Chim. Acta*, 2018, **1026**, 8–36, DOI: 10.1016/j.aca.2018.04.004.

47 M. Hoffmann and F. Noé, *Generating Valid Euclidean Distance Matrices*, 2019.

48 M. B. Whitfield and M. S. Chinn, *Near Infrared Spectroscopic Data Handling and Chemometric Analysis with the R Statistical Programming Language: A Practical Tutorial*, 2017, DOI: 10.1177/0967033517740768.

49 R. S. Ortiz, K. C. Mariotti, N. V. Schwab, G. P. Sabin, W. F. C. Rocha, E. V. R. de Castro, R. P. Limberger, P. Mayorga, M. I. M. S. Bueno and W. Romão, Fingerprinting of sildenafil citrate and tadalafil tablets in pharmaceutical formulations *via* X-ray fluorescence (XRF)

spectrometry, *J. Pharm. Biomed. Anal.*, 2012, **58**(1), 7–11, DOI: 10.1016/j.jpba.2011.09.005.

50 P. C. Trackman, *J. Cell. Biochem.*, 2005, **96**(5), 927–937, DOI: 10.1002/jcb.20605.

51 S. Assi, B. Arafat, K. Lawson-Wood and I. Robertson, Authentication of antibiotics using portable near-infrared spectroscopy and multivariate data analysis, *Appl. Spectrosc.*, 2021, **75**(4), 434–444, DOI: 10.1177/0003702820958081.

52 J. F. Kauffman, M. Dellibovi and C. R. Cunningham, Raman spectroscopy of coated pharmaceutical tablets and physical models for multivariate calibration to tablet coating thickness, *J. Pharm. Biomed. Anal.*, 2007, **43**(1), 39–48, DOI: 10.1016/j.jpba.2006.06.017.

53 S. Romero-Torres, J. D. Pérez-Ramos, K. R. Morris and E. R. Grant, Raman spectroscopy for tablet coating thickness quantification and coating characterization in the presence of strong fluorescent interference, *J. Pharm. Biomed. Anal.*, 2006, **41**(3), 811–819, DOI: 10.1016/j.jpba.2006.01.033.

54 A. Sparén, M. Hartman, M. Fransson, J. Johansson and O. Svensson, Matrix effects in quantitative assessment of pharmaceutical tablets using transmission raman and near-infrared (NIR) spectroscopy, *Appl. Spectrosc.*, 2015, **69**(5), 580–589, DOI: 10.1366/14-07645.

55 R. De Maesschalck, D. Jouan-Rimbaud and D. L. Massart, The mahalanobis distance, *Chemom. Intell. Lab. Syst.*, 2000, **50**(1), 1–18, DOI: 10.1016/S0169-7439(99)00047-7.

56 A. Haywood, *Pharmaceutical Excipients-Where Do We Begin?*, 2011, vol. 34, https://www.australianprescriber.com.

57 C. G. Abrantes, D. Duarte and C. P. Reis, An overview of pharmaceutical excipients: safe or not safe?, *J. Pharm. Sci.*, 2016, 2019–2026, DOI: 10.1016/j.xphs.2016.03.019.

58 N. M. Myers, M. W. Maina, P. M. Were, R. Karwa, S. D. Pastakia, J. C. Sharp, J. L. Luther, A. Cooper, S. L. Bliese, N. Oberhof, D. Aldulaimi and M. Lieberman, Lab on paper: assay of beta-lactam pharmaceuticals by redox titration, *Anal. Methods*, 2019, **11**(37), 4741–4750, DOI: 10.1039/c9ay01547g.

59 United States Pharmacopeial Convention, USP monographs_antibiotics tablets, *Monogr. Dev. Antibiot. USP29*, 2006, **30**(6), 1977.

60 D. Reker, S. M. Blum, C. Steiger, K. E. Anger, J. M. Sommer, J. Fanikos and G. Traverso, Inactive ingredients in oral medications, *Sci. Transl. Med.*, 2019, **11**(483), 1–14, DOI: 10.1126/scitranslmed.aau6753.

61 C. Gomez, P. Lagacherie and G. Coulouma, Regional predictions of eight common soil properties and their spatial structures from hyperspectral vis-NIR data, *Geoderma*, 2012, **189–190**, 176–185, DOI: 10.1016/J.GEODERMA.2012.05.023.

62 T. Selige, J. Böhner and U. Schmidhalter, High resolution topsoil mapping using hyperspectral image and field data in multivariate regression modeling procedures, *Geoderma*, 2006, **136**(1–2), 235–244, DOI: 10.1016/J.GEODERMA.2006.03.050.

63 A. De Man, J. S. Uyttersprot, P. F. Chavez, F. Vandenbroucke, F. Bovart and T. De Beer, The application of near-infrared spatially resolved spectroscopy in scope of achieving continuous real-time quality monitoring and control of tablets with challenging dimensions, *Int. J. Pharm.*, 2023, **641**, 123064, DOI: 10.1016/j.ijpharm.2023.123064.

64 C. P. Meza, M. A. Santos and R. J. Romañach, Quantitation of drug content in a low dosage formulation by transmission near infrared spectroscopy, *AAPS PharmSciTech*, 2006, E206–E214, http://www.aapspharmscitech.org.

