Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

FIELD ASSESSMENT OF ACTIVE INGREDIENT QUANTITY IN PHARMACEUTICAL TABLETS WITH LIMITED CALIBRATION OF NEAR INFRARED SPECTRA: AN APPLICATION TO CIPROFLOXACIN TABLETS

Olatunde Awotunde [a,1], Jin Cai [a,2], Christian Gabriel El Azar [a], Diane Medina [a], Samantha I. Eyolfson [a], Kathleen Hayes [a,3], Christelle Waffo [b], Roland Marini Djang'eing'a [b], Eric M. Ziemons [b], Pierre-Yves Sacré [c], Marya Lieberman [a,,4]*

a Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, IN 46556, USA

b University of Liege (ULiege), CIRM, Laboratory of Pharmaceutical Analytical Chemistry, Department of Pharmacy, Liege, Belgium

c University of Liege (ULiege), CIRM, Research Support Unit in Chemometrics, Department of Pharmacy, Liege, Belgium

Keywords: NIR spectroscopy; chemometrics; machine learning; substandard medicines; predictive model transferability

Abstract

Portable near-infrared (NIR) spectrophotometers have emerged as valuable tools for identifying substandard and falsified pharmaceuticals (SFPs). Integration of these devices with chemometric and machine learning models enhances their ability to provide quantitative chemical insights. However, different NIR spectrophotometer models vary in resolution, sensitivity, and responses to environmental factors such as temperature and humidity, necessitating instrument-specific libraries that hinder the wider adoption of NIR technology. This study addresses these challenges and seeks to establish a robust approach to promote the use of NIR technology in post-market pharmaceutical analysis. We developed support vector machine and partial least squares regression models based on binary mixtures of lab-made ciprofloxacin and microcrystalline cellulose, then applied the models to ciprofloxacin dosage forms that were assayed with high performance liquid chromatography (HPLC). A receiver operating characteristic (ROC) analysis was performed to set spectrophotometer independent NIR metrics to evaluate ciprofloxacin dosage forms as "meets standard," "needs HPLC assay," or "fails standard." Over 200 ciprofloxacin tablets representing 50 different brands were evaluated using spectra acquired from three types of NIR spectrophotometer with 85% of the prediction agreeing with HPLC testing. This study shows that non-brand-specific predictive models can be applied across multiple spectrophotometers for rapid screening of the conformity of pharmaceutical active ingredients to regulatory standard.

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

Introduction

Field screening of pharmaceutical dosage forms using handheld devices is a valuable technique for post-market surveillance of medicines in low-resource settings [1,2]. There are a growing number of commercial products that are advertised for field screening of pharmaceutical dosage forms [3,4]. However, implementation of these devices is hindered by the heterogeneity of pharmaceutical samples encountered in field settings, and by the incompatibility of different manufacturer's instruments and data formats.

Near-infrared spectroscopy (NIRS) is a non-destructive analytical technique with broad applications, including the analysis of pharmaceutical compounds [5,6]. NIRS enables identification and quantification through statistical modeling [7]. Notably, NIRS has emerged as a promising handheld tool for swiftly detecting counterfeit pharmaceuticals through rapid and non-destructive analysis [8]. It has also proven effective in identifying adulterants in genuine formulations and assessing the active pharmaceutical ingredient (API) content in drug formulations [9]. Because NIR spectra are typically broad, chemometric modeling of the data is necessary [10].

Chemometric models for pharmaceutical product identification, with few exceptions [11], are based on libraries of authentic products [12]. The models developed for a given set of authentic products often fail when they are applied to other products in the field, and models developed for one NIR spectrophotometer cannot easily be transferred to other NIR spectrophotometers [13]. Researchers, such as Gryniewicz-Ruzicka *et al*., have demonstrated the discrimination of ciprofloxacin tablets from different manufacturers by combining physical features and chemical signatures, highlighting the variations in formulations [14].

The near-infrared (NIR) spectra obtained from spectrophotometers are susceptible to variations caused by environmental factors, particularly temperature, and humidity [1,15]. Sample temperature variations may influence the intensity and shape of NIR absorption bands if not taken into account during the modeling step [1,2,16]. Humidity levels play a significant role in moisture absorption or desorption in both the sample and the spectrophotometer optics, leading to noticeable alterations in the NIR spectra [17,18]. Moisture can modify the physical and chemical properties of the sample, affecting its reflectance or transmittance properties in the NIR region.

Even under controlled environmental conditions, inherent variations can arise among different spectrophotometers due to discrepancies in design, manufacturing, calibration, and stability, resulting in slight differences in the measured NIR spectra and impacting result accuracy and comparability [19,20]. Because environmental conditions are more controlled in laboratories than in field settings, laboratory-developed statistical models may not be reliable in real-world situations if these influences are not considered during model development.

Each NIR spectrophotometer requires a statistical model to establish an accurate relationship between the acquired spectra and the chemical composition of the samples [21,22]. Over time, spectrophotometers may exhibit changes in performance, resulting in drifts in baseline levels, intensities, or overall spectral features. These drifts can introduce variations in the NIR data, impacting the comparability of measurements taken at different time points or using different

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

spectrophotometers [23, 24]. Spectral noise levels can also vary among NIR spectrophotometers, influencing the signal-to-noise ratio and the reliability of the acquired data. Differences in detector sensitivity, wavelength resolution, and optical alignment can contribute to variations in the noise characteristics of the spectra [8,24]. Furthermore, it is worth noting that the various instruments exhibit dispersion effects influenced by whether they use linear variable filters, MEMS (Micro-Electro-Mechanical Systems), Fabry-Perot interferometer, or Grating + DLP (Digital Light Processing). This diversity arises because the underlying physics of light measurement varies across these technologies. Additionally, variations in data preprocessing methods, such as baseline correction, smoothing, or normalization, can further impact the NIR data. Different spectrophotometers may employ distinct preprocessing techniques, leading to variations in the processed spectra. These instrumental variations underscore the importance of proper calibration, standardization, and validation procedures when utilizing NIR spectrophotometers. In a pharmaceutical environment, the spectrophotometers are qualified

(USP <1058> and <856>) and their performances tracked and ensured before analysis to mitigate this.

In low-resource settings, where pharmaceutical products are sourced from various countries and manufacturers, the extensive variabilities (including environmental conditions) pose a challenge for on-site probing of pharmaceutical qualities by NIR spectrophotometers [25]. Laboratory-made simulated medicines have the potential to address challenges related to such sample variation including excipient interference, tablet coating effects, and environmental influences in field settings [26].

A predictive model was developed for mixtures of pure ciprofloxacin and microcrystalline cellulose using data generated by an Innospectra NIR-M-R2 spectrophotometer. The performance of the models was compared with the gold standard method, HPLC.

2. Experimental

2.1. MATERIALS

Transparent and uncolored size 00 gel capsules were purchased from NOW, Bloomingdale, IL (USA). Ciprofloxacin (CIP) with purity >99%, and microcrystalline cellulose with purity >99% were obtained from Sigma Aldrich (St. Louis, MO, USA).

2.2. LAB FORMULATED MIXTURES

Binary mixtures of CIP and microcrystalline cellulose were formulated as shown in Table S1. To ensure homogeneity, the components were ground and mixed for ~10 min in a clean mortar and pestle, then placed in scintillation vials and vortex-mixed for another 5 min. Lab- made, binary mixture of CIP and crystalline cellulose were subjected to NIR analysis, and the resulting data were explored for an optimized predictive model. Standard normal variate (SNV), second derivative and polynomial Savitzky–Golay transformation, orthogonal projection to latent structures data pretreatments were among preliminary explored for optimized models. Support vector machine and partial least

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

regression (SVR and PLSR) models were developed using 70% of the resulting pretreated spectra data of lab-made CIP formulations. Validation of the models was performed by testing with the remaining, randomly left out 30% data and the respective performance was evaluated to determine optimally performed models.

2.3. DOSAGE FORMS

Samples of CIP dosage forms (n=43) were collected in Kenya from 2020 to 2021, or from Cameroon between 2019 and 2020 (n= 183). These dosage forms were analyzed in 2022 and between 2020 and 2022 respectively. The products were 500 mg oral dosage forms, mostly in the form of tablets packaged in blister packs, some with outer cardboard boxes. Samples (n=43) were stored at 4˚C. IRB approval for sample collection of these 43 products via covert shoppers provided through the University of Notre Dame protocols 18–02–4442 (exp. 2026). These samples were analyzed with a handheld NIR spectrophotometer (NIR-M- R2 USB-powered, InnoSpectra Corporation, Hsinchu, Taiwan). Cameroon samples (n= 183) were analyzed partly in Cameroon and partly in Belgium with another portable handheld NIR spectrophotometer (NIR-S-G1, InnoSpectra Corporation, Hsinchu, Taiwan); these results have been previously published [25,27]. All the dosage forms were analyzed outside the blister by NIR spectrophotometer with barium sulfate that has been pressed into a brick as a reference (Jasco's white standard that comes with the integration sphere) or Spectralon® white diffuse reflectance standard (99%). Each spectrophotometer was controlled by a laptop with the ISC-NIRScan-GUI software (version 3.5.7) in order to collect spectra.

2.4. HIGH PERFORMANCE LIQUID CHROMATOGRAPHY ANALYSIS OF THE BRANDED DRUGS

CIP HPLC analysis was performed on a Waters 2695 separations module equipped with a Waters 2487 Dual-Wavelength Absorbance Detector at an analytical wavelength of 278 nm. A 100 ×4.6 mm XTerra C18 column with 3.5 µm packing size was used for separation. Samples were prepared at ~0.5 mg/mL in mobile phase. 10 µL was injected for analysis and run with an isocratic method of 100% mobile phase with a flow rate of 1.5 mL/min. The mobile phase was prepared by combining 135 mL of acetonitrile with a buffer prepared by diluting 2.9 mL of phosphoric acid in water to 1000 mL and adjusting to pH= 3.0 with triethylamine. The samples from Cameroon were analyzed using a Waters 2695 separation module coupled to a Waters selector valve 7678 and a Waters 996 Photodiode array detector (Waters, Eschborn, Germany) in Belgium and an Agilent 1260 Infinity II HPLC system including a quaternary pump and a DAD 1260 infinity II detector (Agilent Technologies, Santa Clara, CA) in Cameroon. 0.025 M orthophosphoric acid adjusted to pH 3 with triethylamine and methanol (75:25) (v/v) was used as mobile phase and the diluent was 0.025 M orthophosphoric acid adjusted to pH2 with triethylamine and methanol (75:25) (v/v). An X- BRIDGE 2.1*100 mm C18, 3.5 µm was used as stationary phase and a flow of 0.3 mL/min was applied. The method was validated before its implementation [27]. *2.5. NIR data acquisitions*

NIR spectra were acquired on a NIR-M-R2 USB-powered portable spectrophotometer and NIRScan Winform software (InnoSpectra Corporation, Hsinchu, Taiwan). Spectra were acquired by averaging

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

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). A background check was acquired with the same parameters after every five runs of sample analysis with barium sulfate that has been pressed into a brick as a reference (Jasco's white standard that comes with the integration sphere) or in-built reference from the manufacturer. Ten clear capsules were loaded with each respective lab-made formulation, and spectra acquired. For dosage forms, the tablets were scanned outside the blisters but not crushed. At least five spectra of each dosage form were generated by repositioning the tablet between acquisitions. These spectra were then averaged.

For effective sample analysis of pharmaceuticals using the reflectance mode of an NIR instrument, we use a reproducible method for sample presentation [6,8] through a sample holder [28]. New sample holders were 3-D printed to fit the spectrophotometer housing described previously [6]. The new sample holders were designed using Solidworks® software (see file named 'New NIR Cuvette Model (CiproPill) Design 3.SLDPRT' in our GitHub page)[29] to accommodate size 00 capsules or the ciprofloxacin tablets. Each of the samples was presented vertically through our in-house designed sample holder.

This case design is the major difference between the two handheld NIR spectrophotometer since NIR-S-G1was not equipped with the sample holder in this study.

The NIR-S-G1 spectrophotometer is FCC and CE certified, and includes a reflective module integrated with Bluetooth low energy (BLE) function and a battery enclosed within its case. The NIR-M-R2 is identical to NIR-S-G1 but lacks a designed case. The handheld NIR spectrophotometer used for all Cameroon samples was a handheld reflective NIR spectrophotometer (NIR-S-G1, InnoSpectra Corporation, Hsinchu, Taiwan). The NIR-S-G1 was controlled by a laptop with the ISC- NIRScan-GUI software (version 3.5.7, InnoSpectra Corporation, Hsinchu, Taiwan). Each spectrum corresponds to an average of 6 scans in the range of 900–1700 nm (pattern width of 7.03 nm/data point, 2.84 second total acquisition time). The lamp was turned on before starting the analysis (pre-heating phase) until a stable detector's temperature (~ 60°C in Cameroon and ~ 55°C in Belgium) and humidity (~ 10% RH in Cameroon and ~ 0% RH in Belgium) were reached inside the device. This preheating phase is important since the single pixel InGaAs detector of the NIR-S-G1 device has a high sensitivity to temperature at the edges of the spectral range (below 950 and above 1600 nm). A new background was acquired with the same parameters before each sample analysis with a Spectralon® white diffuse reflectance standard (99%) [27].

A benchtop UV-Vis/NIR spectrophotometer, V-670 (Jasco, Japan), with a double-beam, dual grating monochromator utilizing PMT and PbS detectors with 0.5 nm wavelength resolution was used to probe some field collected samples from Kenya. The samples were crushed before being presented to the spectrophotometer with an Integrating Sphere (200–2500 nm) for powders and spectra generated in the absorbance mode. The wavelength range was 900–1700 nm (to align with the portable spectrophotometers) and the spectra were baseline corrected by the instrument. *2.6. Processing and analysis of NIR spectra*

NIR csv data files from the Winform NIRScan software, with header sections containing sample metadata and 228 wavelength/absorption data points were merged and formatted for analysis using Excel. The collated data files in our GitHub repository[29] were uploaded to The Unscrambler X version

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

10.4 (Camo Software, Oslo, Norway) or JupyterLab python 3.0 for data preprocessing and subsequent data analysis.

Standard normal variate (SNV) preprocessing ensures that each spectrum has a standard deviation of one and a mean of zero [30]. Figure S1 shows a comparison of the untreated spectra of twenty ciprofloxacin dosage tablets vs the spectra after SNV pretreatment. Savitzky-Golay filtering (SG) was employed for data smoothing [1]. The key parameters used for SG in python 3.0 were second derivative, second order polynomial and 21 smoothing points. Orthogonal projection to latent structures was used to eliminate unrelated orthogonal variations (noise) from the data. Orthogonal projection to latent structures maximizes spectral features that linearly correlate with the response variable, while the orthogonal component represents the variation that is not correlated with the response [31]. Training spectra data was fit transformed with OPLS using corresponding Y vector for data transformation. To preprocess a new unknown spectrum, we applied orthogonal projection to the latent structures function on the pyopls library using python. The opls.transform allows orthogonal data transformation without a Y vector.

Multivariate data analysis – either support vector machine regression (SVR) with parameters: kernel='sigmoid', gamma = 0.02, C = 2 or partial least square regression (PLS-R) with 13 latent variables – was next applied to develop model for the training data. The mean square error of calibration and prediction (MSEC and MSEP) were used a metric to select the optimized latent variables The models were optimized through manual tuning or grid search of the parameters. In each data model, random splitting of the dataset was performed; 70% of the spectra were used to train the model, 30% to test the model after training. K-fold cross-validation is used for both SVM and PLS-R models. The 'KFold' function is used to split the dataset into 'n_splits' folds, and the models are trained and tested on each fold. This helps assess the model's performance on different subsets of the data, providing a more robust evaluation compared to a single train/test split. 'GridSearchCV' from scikit-learn was employed to perform a grid search over a range of latent variables and selects the model with the best performance based on cross-validated mean squared error. The reported mean square error of prediction (MSEP) and other model parameters are from these validation results.
The key parameters used for SVR and the PLSR models are available in the open-source code in our GitHub repository [29].

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 [29], along with a guided workflow in the Jupyterlab named 'Field assessment of active ingredient quantity in pharmaceutical tablets.ipynb'.

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

3. Results and discussion

3.1. CALIBRATION OF REGRESSION MODELS

The SVR showed better performances than PLSR with mean square error (MSEP) of SNV treated data less than 1%w/w (Table S5). For the orthogonal projection to latent structures (OPLS), PLSR showed better performance than SVR (Table S6). The PLSR model of OPLS and SVR model of SNV treated data gave better separation between the different lab-made formulations of %w/w ciprofloxacin among all the data pretreatments as would be revealed in further studies with field collected samples. Scores of PCA exploratory studies (Fig. 1) showed the sensitivity of OPLS to remove any features that are not correlated with the variation in the targets' concentrations while maximizing the separation between various targets' formulations. These unique characteristics can be explored for discriminating a conformed medicine from adulterated ones.

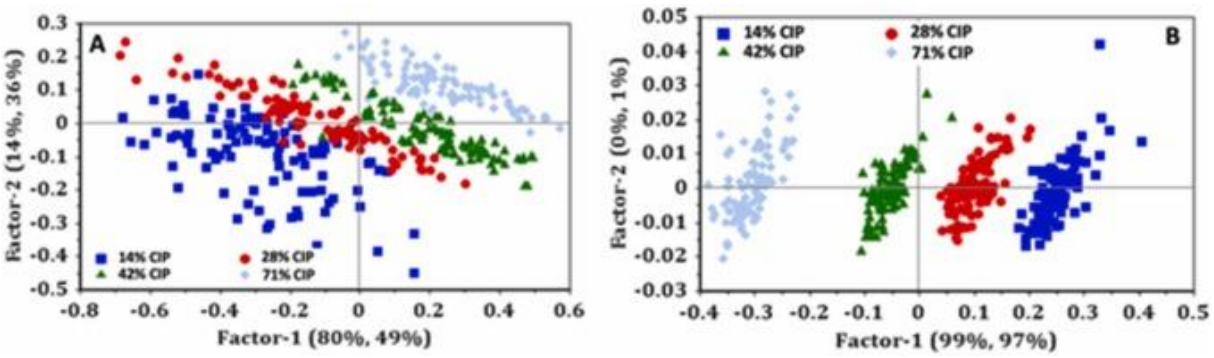


Fig. 1. Scatter plots of principal component analysis scores of A. raw and B. Orthogonal projection to latent structures (OPLS) lab-made ciprofloxacin formulations showing concentration of ciprofloxacin, 200 mg CIP, 400 mg CIP, 600 mg CIP & 1000 mg CIP in 1400 mg binary mixture with cellulose (14%, 28%, 42% and 71% CIP respectively).

The US pharmacopeia monograph provides a well-defined guideline for HPLC assessment of pharmaceuticals including antibiotics like CIP [32]. According to this monograph, quantitative analysis of actual ciprofloxacin in the total pill is required to be between 90% and 110% of the stated dosage to conform with the regulatory standard (the HPLC assay) [32]. However, NIR analysis evaluates the %w/w not the total amount of CIP in the tablets. HPLC analysis was carried out using the USP monograph method on single tablets of the ciprofloxacin tablets. The HPLC analysis of these samples gave the actual API content in each of the tablets. Since weighing the pill mass is a required step in this evaluation, we were able to determine the %w/w of ciprofloxacin in all the collected field samples. Variation in excipients and formulations by manufacturers as well as the allowed pharmacopeia range for conformed pills are common and contributed to our observed wide range of %w/w ciprofloxacin (57 – 76%w/w see Figure S2) in the HPLC analysis. These HPLC results were used to validate the performance of the NIR predictive model(s) that we developed. Further efforts to test the sensitivity of the model(s) were investigated with minor and major dilution of the selected field collected samples with known amount of microcrystalline cellulose simulating substandard products.

3.2. Can lab-made developed models robustly probe collected ciprofloxacin field samples from Kenya?

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

The models developed using PLSR and SVR based models of lab- formulated binary mixture of CIP and microcrystalline cellulose were further applied on 43 branded CIP tablets from Kenya taken out of the respective blisters, (in addition to 30 simulated samples - with minor and major dilution) and the performance of the models was compared to that of HPLC. We explored our previously tested approach that correlates the %w/w of API to total pill mass as a metric to probe and discriminate substandard pharmaceuticals. All CIP tested were tablets with different sizes, shapes, and colors. Our hypothesis was that OPLS (orthogonal projection to latent structures) treated data would minimize those variations by optimizing the key target features, but irrespective of this, we tested other pretreated models (SNV and SNV+SG).

The SNV and SNV+ SG predictive models failed to correctly quantify the %w/w API in the field collected samples, with both the PLSR and SVR models predicting negative %w/w API for all CIP tablets. However, for OPLS (orthogonal projection to latent structures) based models, SVM predicted almost the same value for all tablets while PLS gave a pretty random prediction (compared to HPLC), see Figure S2. The apparent % w/w ciprofloxacin for the undiluted field collected samples was between 53 and 90 for the NIR model and HPLC, this could be explained as due to a measure of variations in statistical methods, tablet formulations and/ or instrumental chemical analysis (Figure S2). All 43 tablets passed HPLC analysis, therefore we selected the above range for the NIR models as metric for conformed quality product than the earlier range of 57 – 76%w/w, which is HPLC based evaluation. We explored further studies to further refine the metric for identification of conformed CIP. In particular, we needed to be sure the NIR model could identify substandard ciprofloxacin.

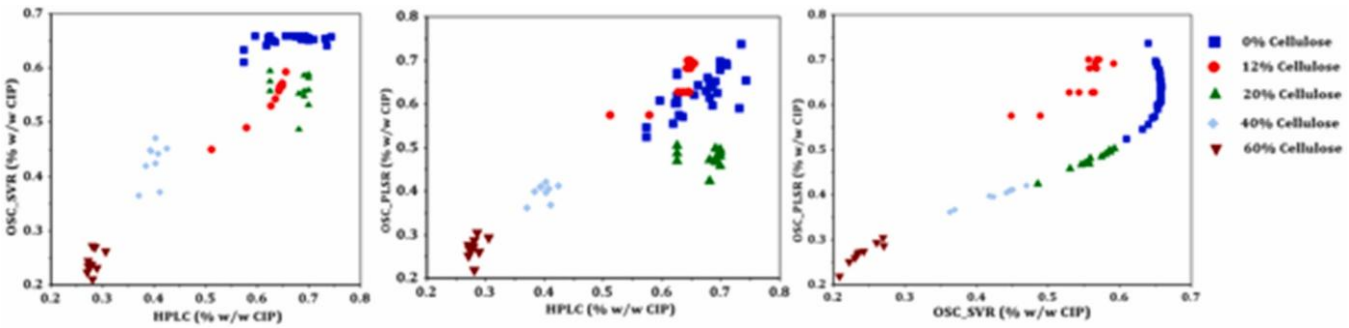


Fig. 2. Scatter plots of the correlation between NIR model predictions A. orthogonal projection to latent structures + support vector machine regression (OPLS+SVR) vs HPLC in %w/w, B. orthogonal projection to latent structures + partial least square regression (OPLS+PLSR) vs HPLC and C orthogonal projection to latent structures + support vector machine regression (OPLS+SVR) vs orthogonal projection to latent structures + partial least square regression (OPLC+PLSR), of crushed ciprofloxacin tablets collected from Kenya diluted with %w of crystalline cellulose simulating broad degree of adulteration.

We investigated hypothetical substandard ciprofloxacin by systematically diluting some branded CIP that passed HPLC assay. 30 of the branded CIP tablets that passed HPLC assay were crushed, placed in clear capsules, and subjected to NIR analysis, and subsequently diluted with 12, 20, 40 and 60% microcrystalline cellulose. Here, we aimed to see how small and large dilution of the authentic product could influence the NIR model predictions. We probed the predictive performance of OPLS+PLSR and OPLS+SVR models with these simulated field samples and observed that a significant proportion of these samples were better discriminated by SVR model (Figs. 2A, and 3) compared to PLSR based model (Fig. 2B). The Support Vector Regression (SVR) model exhibited enhanced discrimination of the borderline between diluted products (contaminated with cellulose) and field-collected samples, with

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

minor exceptions that could be attributed to the non-uniform mixing of the diluents. A unique relationship existed between the two NIR predictive models (Fig. 2C) which we propose to explore in the future works.

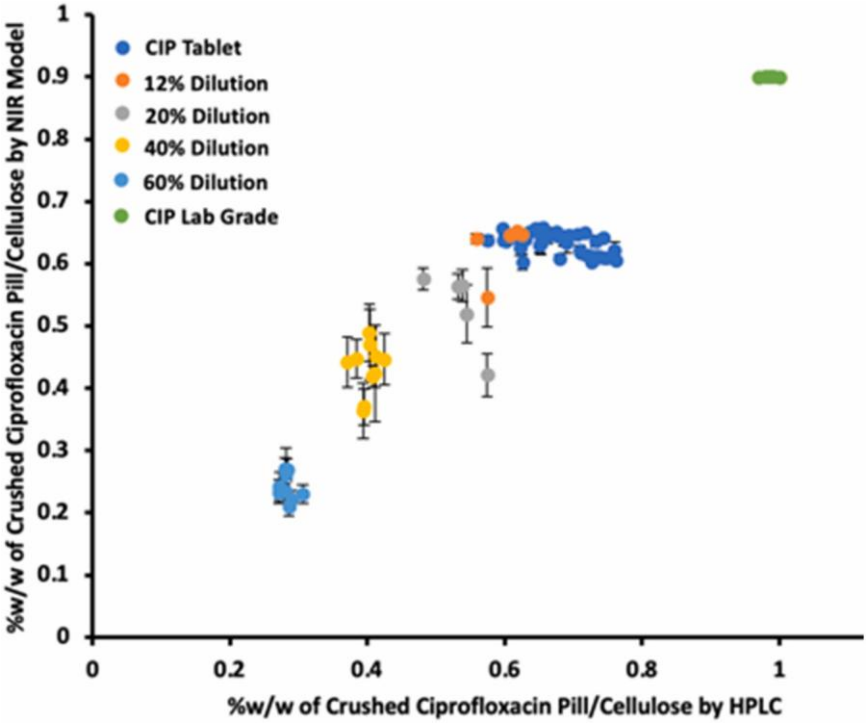


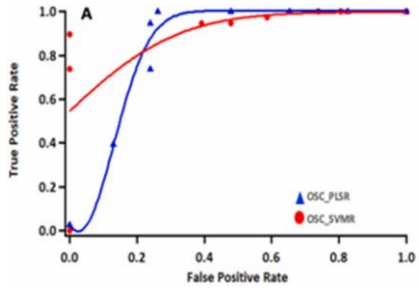
Fig. 3. A. Plot of the correlation between NIR model prediction with OPLS + SVR data analysis, and HPLC analysis of 43 ciprofloxacin tablets collected from Kenya. Date includes 30 samples which are crushed tablets diluted with crystalline cellulose at 12%,20%,40% and 60% dilutions, and 5 samples of pure lab-grade ciprofloxacin.

The reproducibility of the analytical technique was tested by shaking, rotating, and re-presenting the content. Our study revealed a clear distinction between clusters of different dilutions (Fig. 3). Some samples diluted with 12% cellulose successfully met the criteria set by the USP for 500 mg ciprofloxacin, as outlined in the HPLC and NIR model predictions (Fig. 3). This outcome aligns with the permissible tolerance range specified by the USP (90–110% of 500 mg, i.e., 450–550 mg) after the dilution process. The shaking, rotation, or the side of the pill introduced to the NIR light did not have a significant impact on the discriminating ability of the model for these homogeneous samples. This approach summarizes the robustness of the NIR developed model to handle likely challenges that may arise from untrained users.

The results from these 70 samples as evaluated with OPLS+SVR model trained with lab-made binary formulation, were used to establish a multi-class receiver operating characteristic (Fig. 4, Table S2 & S3) [33]. The multiple classes separate the samples into those that are clearly good quality, clearly bad quality, and those that require additional confirmatory testing. All samples that the model predicts to fall between 60% and 79% w/w ciprofloxacin content are classified as pass. If the model predicts less than 49% w/w of ciprofloxacin, the sample is classified as failed. Samples that fall in the predicted range of 50–59%, or above 80%, are recommended for HPLC analysis. This set metric was further refined by evaluating the false positive rate and false negative rate of the predictive model against HPLC assayed results (Fig. 4B). We want our model to have zero tolerance for false positive predictions

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

i.e., be able to identify all samples that would have failed HPLC assay as "failed samples" or "recommend for HPLC". We tested this set metric with spectra of field collected samples to assess the performance of the lab formulation based developed model (OPLS-SVR).



B

	Orthogonal Projection to latent Structures + support vector machine regression model	
Cut-off (%w/w)	False Positive Rate	False Negative Rate
0	1	0
0.30	0.80	0.00
0.40	0.74	0.00
0.45	0.59	0.00
0.50	0.48	0.05
0.55	0.39	0.05
0.60	0.00	0.10
0.65	0.00	0.21
0.70	0.00	0.50

Failed
HPLC Recommended
Passed

Fig. 4. A. ROC plot of NIR-based metric to classify genuine formulation from SFPs B. A metric that doesn't rely on HPLC for assessing the CIP's % w/w content uses predictive values from a NIR model. It is labeled "passed" when no false positive rates are detected, "failed" when false negatives are present, and suggests HPLC analysis for cases in between.

This NIR-based metric was applied to the field collected, 178 ciprofloxacin dosage tablets, produced by different manufacturers across 12 countries. All tablets were assayed by HPLC. As shown in Table 1, the metric reliably showed some consistency with HPLC assay outcomes. All the samples that failed HPLC (6) were either flagged as "Failed" (4) or "HPLC Recommended" (2) by the NIR model. Of the 172 samples that passed HPLC assay, 33 samples (19%) were flagged by the NIR model as suspicious:15 (9%) of them were flagged by NIR model as requiring HPLC while 18 (10%) were classified as failed or requiring further confirmatory testing. The multi-class receiver classifier did not record any false positive for the 178 samples from Cameroon, however, 18 samples classified as "fail" were incorrectly identified (10%) as they all passed HPLC assay. We therefore recommend HPLC assay for all samples that do not pass.

If deployed in the field as a screening test for a post-market surveillance project, the NIR predictive model would catch 100% of the bad quality products, while recommending HPLC analysis of 39 samples instead of 178, a 78% reduction in the cost and time required for HPLC analysis. The lab resources freed up by the use of the screening test could then be applied to analysis of additional post-market samples. This would allow a five-fold expansion of the scope and scale of post-market surveillance, and ultimately detect a greater number of bad quality products in the market.

While the approach employed in this study demonstrated satisfactory performance, we recognize potential challenges associated with its emphasis on %w/w in prediction. Instances may arise where samples falling within the pass range exhibit lower CIP amounts if the total pill mass is below the typical range. Additionally, conforming samples might register %w/w below the pass range due to formulation differences or the use of specific excipients. While these scenarios did not manifest in the present study, our ongoing objective is to assess the model's performance in such circumstances. We aim to refine the metric to reduce the likelihood of categorizing actual non-conforming products as conforming. In cases where uncertainty persists, we recommend conducting further testing or investigating the product in question.

3.3. How well does a predictive model built from one NIR spectrophotometer perform on spectra from another spectrophotometer?

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

The calibration model transfer between different instruments is a challenging task in chemical analysis [20]. However, with the use of OPLS and other data preprocessing, it is possible to apply calibration models on different instruments without the need to re-create the whole calibration sample set [31]. By removing the largest sources of variation that are orthogonal to the analyte variations of interest, OPLS enhances the ability to detect and quantify the analyte signals that are of interest.

This makes it an ideal tool for data preprocessing in the case of multi spectrometer prediction with a calibration model built with a single equipment's model.

To probe the transferability of the spectra recorded on different instruments, we subjected spectra data from two different NIR spectrophotometers to OPLS data pretreatment analysis as discussed in the prior section. This ensures that the data from different instruments is standardized, and that most unwanted variation is eliminated. Spectra of 10 CIP products from Kenya and the samples of each product diluted with 30% or 50% microcrystalline cellulose were analyzed by the two NIR spectrophotometers: a benchtop V-670 (Jasco) and handheld NIR-M-R2, (Innospectra). The benchtop spectrophotometer has higher resolution of about 0.025 nm while the resolution of the NIR-M-R2 is about 3.5 nm. The spectra of the samples from these two distinct NIR spectrophotometers were very similar except that the numbers of data points were 1645 and 228 for the benchtop and handheld spectrophotometers, respectively. To adapt the benchtop spectra dataset to the predictive models developed earlier, the total number of data points was trimmed to 228 by linear interpolation. The spectral features of the benchtop's spectrum were identical before and after the trimming but quite similar to the NIR-M-R2 spectra for the same tablets (Fig. 5). The raw data comprising 228 variables for each sample from the two spectrophotometers were pretreated with OPLS and subsequently analyzed with the previously developed OPLS+SVR and OPLS+PLSR models. A linear correlation was observed between the predicted w/w% values (Fig. 5) from both datasets. All the samples that passed the HPLC assay also passed the NIR-based prediction using the ROC set criteria.

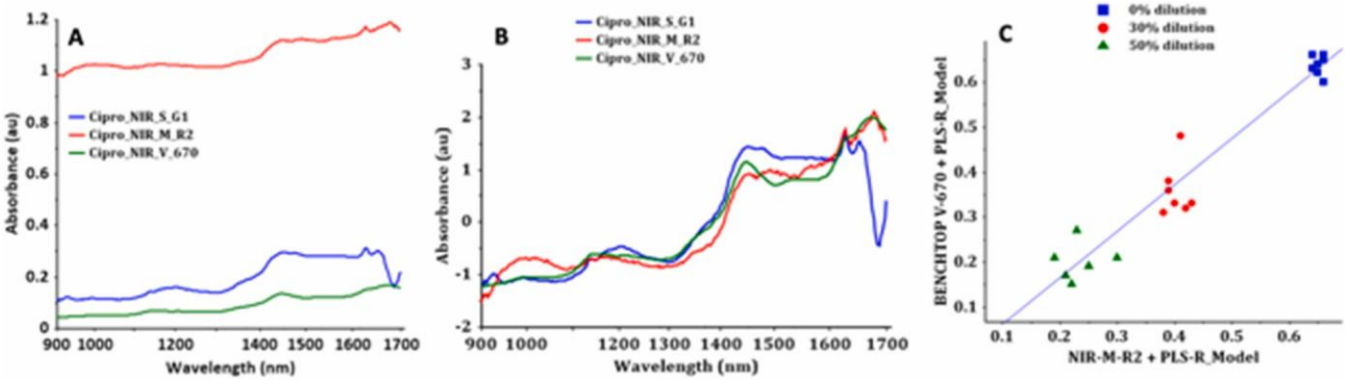


Fig. 5. A. Raw, B. SNV plot of spectra of Ciprofloxacin tablet generated by NIR-M-R2, NIR-S-G1, and Benchtop NIR V-670, JASCO spectrophotometers in Notre Dame, US. C. Correlation between the spectra of Ciprofloxacin tablets generated by NIR-M-R2 and benchtop NIR spectrophotometer using our developed model

The preprocessed data from our NIR-M-R2 spectrophotometer were used to build the predictive models discussed throughout this work. We tested the robustness of the NIR-M-R2 based model with data that had been generated from another NIR spectrophotometer (NIR-S-G1) in Cameroon (partly discussed earlier). Rapid change in environmental conditions is common in tropical countries, thus we were interested in the robustness of our lab-made binary mixture models to probe pharmaceuticals despite such conditions [24,34].

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

We evaluated the %w/w ciprofloxacin in 178 ciprofloxacin tablets measured with the NIR-S-G1 spectrophotometer through our models, trained with spectra lab-formulated binary mixtures obtained on the NIR-M-R2 spectrometer. Here, we aimed to check the robustness of the predictive models in probing pharmaceuticals of different origins (Cameroon, India, United Kingdom, Germany, China, Turkey, USA, Austria, Spain, Nigeria, France, and Belgium), analyzed under different environmental conditions by another model of the spectrometer. The pass threshold is set at greater than or equal to 60%w/w, HPLC analysis was recommended for borderline of 59–50%w/w and the "failed" threshold is set at less than or equal to 40%w/w (see Fig. 3). Using the set metric, the predictive model showed a satisfactory performance as discussed earlier. With this prediction performance, we aim to optimize the model and test it on spectra from a broader range of NIR spectrophotometers from multiple manufacturers.

Table 1
NIR predictive model with OPLS+SVR data analysis applied to 178 dosage forms from Cameroon, including good quality ciprofloxacin (n=172) and 6 "bad quality" products, substandard ciprofloxacin (n=6).

	True Condition	
NIR Model	**Pass HPLC (n¼172)**	**Fail HPLC (n¼6)**
Pass (n=139)	139	0
HPLC Recommended (n = 17)	15	2
Fail (n=22)	18	4

Overall, the use of orthogonal projection to latent structures simplifies the process of applying calibration models between different instruments. This approach reduces the time and cost associated with creating and re-analyzing calibration sample sets, and can be optimized using a broader range of samples for field settings. It also improves the accuracy and reliability of the transferred model, making it a valuable tool in chemical analysis.

4. Conclusion

This study highlights the significant potential of portable Near-IR as a valuable tool for post-market surveillance of pharmaceutical dosage forms in low-resource settings. The research presents predictive models for probing the %w/w content of ciprofloxacin in different field collected samples, demonstrating the potential to estimate a key pharmaceutical quality attribute without relying on HPLC. The use of HPLC-

independent metric can significantly reduce the time and cost required for HPLC analysis, enabling more extensive post-market surveillance and improved detection of substandard pharmaceutical products.

One key strength of this research is the exploration of model applicability between different NIR spectrophotometers. The use of orthogonal projection to latent structures (OPLS) for data preprocessing enhances the transferability of calibration models, addressing instrument variations. The developed models have demonstrated robustness in diverse environmental conditions and with samples from various manufacturers and countries. This indicates their potential to serve as an

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

effective tool for post-market pharmaceutical quality assessment, facilitating a broader scope of surveillance and enhancing the detection of substandard products.

However, it's important to acknowledge certain limitations of the developed method. One notable limitation is that the models were primarily based on laboratory-made binary mixtures. While these models have shown promise, their real-world applicability to a wide range of pharmaceutical products and formulations requires further validation and optimization. Additionally, the study does not account for the full spectrum of potential variations and complexities encountered in low- resource settings, such as variations in excipients, tablet coating effects, and other environmental factors. Further research and refinement of the models are necessary to ensure their reliability and broad acceptance in post-market surveillance.

Research contributions

Olatunde Awotunde designed the 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.

Jin Cai, Christian Gabriel El Azar, and Diane Medina generated the spectra data for Kenya samples, lab-made formulations, and HPLC assay data at Lieberman lab.

Samantha I. Eyolfson and **Dr. Kathleen Hayes** and supported by Olatunde, contributed to the HPLC analysis of most field-collected samples.

Christelle Waffo provided all NIR spectra for the field-collected samples from Cameroon and the corresponding HPLC results.

Roland Marini Djang'eing'a provided all NIR spectra for the field- collected samples from Cameroon and the corresponding HPLC results.

Eric Ziemons provided all NIR spectra for the field-collected samples from Cameroon and the corresponding HPLC results and review the manuscript.

Pierre-Yves Sacre provided all NIR spectra for the field-collected samples from Cameroon and the corresponding HPLC results and review the manuscript.

Marya Lieberman conceived the experimental design, edited the manuscript, and was responsible for overall supervision and funding of this work.

Author Agreement Statement

We the underlisted authors declare that this manuscript is original, has not been published before and is not currently being considered for publication elsewhere. We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript

has been approved by all of us. We understand that the Corresponding Author is the sole contact for the Editorial process. He is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs Signed by all authors.

Funding

This work was supported by NSF (CMMI-1842369, OISE-1559497). Pulte Initiative, University of Notre Dame for Pamoja summer award, Lucy Institute of Data and Society, University of Notre Dame for Lucy fellowship and American Chemical Society for the Fall 2022 Pfizer Travel Award.

CRediT authorship contribution statement

Marya Lieberman: Writing – review & editing, Resources, Project administration, Funding acquisition, Conceptualization. **Pierre-Yves Sacre:** ´ Writing – review & editing, Validation, Supervision, Methodology, Formal analysis. **Jin Cai:** Methodology. **Olatunde Awotunde:** Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **Christian Gabriel El Azar:** Methodology. **Samantha I Eyolfson:** Methodology. **Diane Medina:** Methodology. **Christelle Waffo:** Writing – original draft, Methodology. **Kathleen Hayes:** Validation, Supervision, Methodology. **Eric M Ziemons:** Supervision. **Roland Marini Djang'eing'a:** Methodology.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work has also benefited from the facilities available through the Notre Dame Center for Environmental Science and Technology (CEST). **Appendix A. Supporting information**

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jpba.2024.116189.

References

[1] C. Pasquini, Near infrared spectroscopy: a mature analytical technique with new perspectives – a review, Anal. Chim. Acta 1026 (2018) 8–36, https://doi.org/ 10.1016/j.aca.2018.04.004.

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

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

[3] R. Cockburn, P.N. Newton, E.K. Agyarko, D. Akunyili, N.J. White, The global threat of counterfeit drugs: why industry and governments must communicate the dangers, PLoS Med 2 (4) (2005) 0302–0308, https://doi.org/10.1371/journal. pmed.0020100.

[4] C. Caillet, et al., Evaluation of portable devices for medicine quality screening: lessons learnt, recommendations for implementation, and future priorities, PLoS Med. 18 (9) (2021) 1–12, https://doi.org/10.1371/journal.pmed.1003747.

[5] J. Linderholm, P. Geladi, N. Gorretta, R. Bendoula, A. Gobrecht, Near infrared and hyperspectral studies of archaeological stratigraphy and statistical considerations, Geoarchaeology 34 (3) (2019) 311–321, https://doi.org/10.1002/gea.21731.

[6] O. Awotunde, et al., Discrimination of substandard and falsified formulations from genuine pharmaceuticals using NIR Spectra and machine learning, Anal. Chem. 94 (37) (Sep. 2022) 12586–12594, https://doi.org/10.1021/acs.analchem.2c00998.

[7] Y. Tie, et al., Spectroscopic techniques combined with chemometrics for fast on-site characterization of suspected illegal antimicrobials, Talanta 217 (April) (2020) 121026, https://doi.org/10.1016/j.talanta.2020.121026.

[8] Q. Chai, J. Zeng, D. Lin, X. Li, J. Huang, W. Wang, Improved 1D convolutional neural network adapted to near-infrared spectroscopy for rapid discrimination of Anoectochilus roxburghii and its counterfeits, J. Pharm. Biomed. Anal. 199 (2021) 114035, https://doi.org/10.1016/j.jpba.2021.114035.

[9] E. Ziemons, J. Mantanus, P. Lebrun, E. Rozet, B. Evrard, P. Hubert, Acetaminophen determination in low-dose pharmaceutical syrup by NIR spectroscopy, J. Pharm. Biomed. Anal. 53 (3) (2010) 510–516, https://doi.org/10.1016/j. jpba.2010.06.003.

[10] H. Chen, Z. Lin, C. Tan, Fast quantitative detection of sesame oil adulteration by near-infrared spectroscopy and chemometric models, Vib. Spectrosc. 99 (September) (2018) 178–183, https://doi.org/10.1016/j.vibspec.2018.10.003.

[11] O. Awotunde, et al., Mitigating the impact of gelatin capsule variability on detection of substandard and falsified pharmaceuticals with near-IR spectroscopy, Anal. Methods (2024), https://doi.org/10.1039/D4AY00001C.

[12] I. Storme-Paris, et al., Challenging Near InfraRed Spectroscopy discriminating ability for counterfeit pharmaceuticals detection, Anal. Chim. Acta 658 (2) (2010) 163–174, https://doi.org/10.1016/j.aca.2009.11.005.

[13] K.B. Bec, J. Grabska, C.W. Huck, Principles and applications of miniaturized near- infrared (NIR) spectrometers, Jan. 21, Chem. - A Eur. J. 27 (5) (2021) 1514–1532, https://doi.org/10.1002/chem.202002838.

[14] N. Fuenffinger, S. Arzhantsev, C. Gryniewicz-Ruzicka, Classification of ciprofloxacin tablets using near-infrared spectroscopy and chemometric modeling, Appl. Spectrosc. 71 (8) (2017) 1927–1937, https://doi.org/10.1177/ 0003702817699624.

[15] EUROPEAN PHARMACOPOEIA 5.0 2.2.40. Near-infrared spectrophotometry. Accessed: Oct. 03, 2023. [Online]. Available: https://www.ema.europa.eu/en/ documents/scientific-guideline/draft-guideline-use-near-infrared-spectroscopy- pharmaceutical-industry-data-requirements-new_en-0.pdf

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

[16] M. Razuc, A. Grafia, L. Gallo, M.V. Ramírez-Rigo, R.J. Romanach, Near-infrared ˜ spectroscopic applications in pharmaceutical particle technology, Drug Dev. Ind. Pharm. 45 (10) (2019) 1565–1589, https://doi.org/10.1080/ 03639045.2019.1641510.

[17] A. Candolfi, R. De Maesschalck, D.L. Massart, P.A. Hailey, Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA 19 (1999) 923–935.

[18] A.K. Deisingh, Pharmaceutical counterfeiting (Royal Society of Chemistry), Analyst 130 (3) (2005) 271–279, https://doi.org/10.1039/b407759h.

[19] J. Engel, et al., Breaking with trends in pre-processing? TrAC - Trends Anal. Chem. 50 (2013) 96–106, https://doi.org/10.1016/j.trac.2013.04.015.

[20] C. Liang, H.F. Yuan, Z. Zhao, C.F. Song, J.J. Wang, A new multivariate calibration model transfer method of near-infrared spectral analysis, Chemom. Intell. Lab.
Syst. 153 (2016) 51–57, https://doi.org/10.1016/j.chemolab.2016.01.017.

[21] J. Grabska, K.B. Bec, S. Mayr, C.W. Huck, Theoretical simulation of near-infrared spectrum of piperine: insight into band origins and the features of regression models, Appl. Spectrosc. 75 (8) (2021) 1022–1032, https://doi.org/10.1177/ 00037028211027951.

[22] Z. Xu, et al., A correlation-analysis-based wavelength selection method for calibration transfer, Spectrochim. Acta A Mol. Biomol. Spectrosc. vol. 230 (2020) 118053, https://doi.org/10.1016/j.saa.2020.118053.

[23] A. Heinz, M. Savolainen, T. Rades, C.J. Strachan, Quantifying ternary mixtures of different solid-state forms of indomethacin by Raman and near-infrared spectroscopy, Eur. J. Pharm. Sci. 32 (3) (2007) 182–192, https://doi.org/10.1016/ j.ejps.2007.07.003.

[24] Z. Xiaobo, Z. Jiewen, M.J.W. Povey, M. Holmes, M. Hanpin, Variables selection methods in near-infrared spectroscopy, Anal. Chim. Acta vol. 667 (1–2) (2010) 14–32, https://doi.org/10.1016/j.aca.2010.03.048.

[25] C.A.W. Tchounga, et al., Prevalence of poor quality ciprofloxacin and metronidazole tablets in three cities in Cameroon, Am. J. Trop. Med. Hyg. 108 (2) (2023) 403–411, https://doi.org/10.4269/ajtmh.22-0221.

[26] G. Lawson, J. Ogwu, and S. Tanna, Quantitative screening of the pharmaceutical ingredient for the rapid identification of substandard and falsified medicines using reflectance infrared spectroscopy, 2018, doi: 10.1371/journal.pone.0202059.

[27] C.A.W. Tchounga, et al., Usefulness of medicine screening tools in the frame of pharmaceutical post-marketing surveillance (August, Aug), PLoS One 18 (8) (2023), https://doi.org/10.1371/journal.pone.0289865.

[28] W.L. Yoon, R.D. Jee, A.C. Moffat, Optimisation of sample presentation for the near- infrared spectra of pharmaceutical excipients, Analyst 123 (5) (1998) 1029–1034, https://doi.org/10.1039/a800358k.

[29] O. Awotunde *et al.*, Field assessment of active ingredient quantity in pharmaceutical tablets with limited calibration of near infrared spectra - An application to ciprofloxacin tablets. Accessed: Dec. 01, 2023. [Online]. Available: https://github.com/aawotund/Field-assessment-of-active-ingredient-quantity-in- pharmaceutical-tablets-with-limited-calibration.

[30] C. Pasquini, Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications, J. Braz. Chem. Soc. 14 (2) (2003) 198–219, https://doi. org/10.1590/S0103-50532003000200006.

[31] M. Blanco, J. Coello, I. Montoliu, M.A. Romero, Orthogonal signal correction in near infrared calibration, Anal. Chim. Acta 434 (1) (2001) 125–132, https://doi. org/10.1016/S0003-2670(01)00820-0.

Published in : Journal of Pharmaceutical and Biomedical Analysis (2024), p. 116189
DOI: 10.1016/j.jpba.2024.116189
Status : Postprint (Author's version)

[32] United States Pharmacopeial Convention, USP Monographs_ Ciprofloxacin Tablets, Monograph Development-Antibiotics USP29-, vol. 30(6). p. 1977, 2006. Accessed: Nov. 26, 2023. [Online]. Available: http://www.pharmacopeia.cn/v29240/ usp29nf24s0_m4170.html.

[33] J.N. Mandrekar, Receiver Operating Characteristic Curve in Diagnostic Test Assessment, 2010.

[34] A. Gombas, I. Antal, and P. Szabor´evesz, Quantitative determination of ´ crystallinity of alpha-lactose monohydrate by Near Infrared Spectroscopy (NIRS), vol. 256, pp. 25–32, 2003, doi: 10.1016/S0378-5173(03)00059-0.