analytical chemistry

pubs.acs.org/ac Article

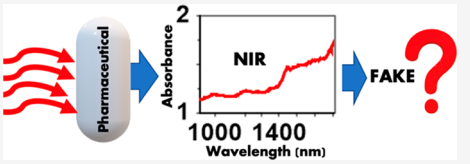
Discrimination of Substandard and Falsified Formulations from Genuine Pharmaceuticals Using NIR Spectra and Machine Learning

Olatunde Awotunde, Nicholas Roseboom, Jin Cai, Kathleen Hayes, Revati Rajane, Ruoyan Chen, Abdullah Yusuf, and Marya Lieberman*

Cite This: https://doi.org/10.1021/acs.analchem.2c00998 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Near-infrared (NIR) spectroscopy is a promising technique for field identification of substandard and falsified drugs because it is portable, rapid, nondestructive, and can differentiate many formulated pharmaceutical products. Portable NIR spectrometers rely heavily on chemometric analyses based on libraries of NIR spectra from authentic pharmaceutical samples. However, it is difficult to build comprehensive product libraries in many low- and middle-income countries due to the large numbers of manufacturers who supply these markets, frequent unreported changes in materials sourcing and product formulation by the manufacturers, and general lack of cooperation in providing authentic samples. In this work, we show that a simple library of lab-formulated binary mixtures of an active pharmaceutical ingredient (API) with two diluents gave good performance on field screening tasks, such as discriminating substandard and falsified formulations of the API. Six data analysis models, including principal component analysis and support-vector machine classification and regression methods and convolutional neural networks, were trained on binary mixtures of acetaminophen with either lactose or ascorbic acid. While the models all performed strongly in cross-validation (on formulations similar to their training set), they individually showed poor robustness for formulations outside the training set. However, a predictive algorithm based on the six models, trained only on binary samples, accurately predicts whether the correct amount of acetaminophen is present in ternary mixtures, genuine acetaminophen formulations, adulterated acetaminophen formulations, and falsified formulations containing substitute APIs. This data analytics approach may extend the utility of NIR spectrometers for analysis of pharmaceuticals in low-resource settings.



1. INTRODUCTION

Detection of substandard and falsified pharmaceuticals (SFPs) in field settings is a still unmet challenge for analytical chemistry, particularly in locations where regulatory resources are inadequate or where individuals or organizations have developed strong networks to bypass regulatory authorities. The impact of SFPs on patient health and medical systems costs is enormous.[1−3] These products may be directly harmful to patients or may lack efficacy in treating illness, leading to poor clinical outcomes; they also contribute to the development of antimicrobial resistance and reduce trust in the entire medical system.[4] In low- and middle-income countries (LMICs), the World Health Organization (WHO) estimates that one in ten products sold is an SFP, constituting a significant fraction of health care expenditures.[5−8] While pharmaceutical companies, distributors, and regulators are making efforts to ensure a proper and secure supply chain for the safe delivery of drugs to end users, the fight against SFPs remains a great threat to public health.[9−11]

One strategy for the detection of SFPs involves empowering stakeholders (regulators, pharmacists, or even patients) through point-of-use technology that can presumptively identify low-quality products. The most widely available point-of-use devices are portable spectrophotometers, which conduct Raman, infrared, or near-infrared (NIR) analysis of the vibrational modes of the substances found in pharmaceutical products. The suitability of these technologies for detection of SFPs in field settings has been reviewed recently by Kovacs et al. and Roth et al.[9,12] In a multistage study, Caillet et al. evaluated 41 technologies for detection of SFP antimalarial drugs; 12 were selected for laboratory testing, and the most promising 6 were tested in a simulated LMIC pharmacy setting. Of the 12 devices that were selected for laboratory evaluation, three were NIR spectrometers. Difficulties in building suitable reference libraries for the NIR spectrometers were commonly cited disadvantages of this technology both in lab settings[13,14] and for field use: "... reference library creation and updates will incur significant costs ... some 'reference' samples contained API content outside pharmacopeial limits, despite being procured from what were thought to be reliable sources."[15] In the field, users

Received: March 2, 2022
Accepted: August 25, 2022

often chose the incorrect reference libraries from options provided on the NIR spectrometer software, which could cause false classification of good-quality products as fakes, "There appeared to be a lack of awareness that different brands of the same API may contain different excipients, resulting in need for different reference libraries."[16]

The complexity often associated with pharmaceutical products has been a major challenge toward building a robust model for point of use technology in global assessment of SFPs. NIR is attractive for applications in pharmaceutical and medical fields because it is nondestructive, requires little to no sample preparation, and is very fast to operate.[17] Pasquini and Roggo et al. recognized the unique ability of NIR to give useful information about complex samples including petroleum oil, wood, polymers, and pharmaceuticals.[18−20] This information ranges from classification, authentication, or identification to quantification of target analytes in complex matrices.[21]

Unlike IR spectra, which can be associated with fundamental vibrational modes of small groups of atoms, NIR spectra consist of overlapping bands arising from overtones and combinations of bond vibrations. The spectra are also affected by intermolecular interactions. The strength of the vibrational excitations depends on the polarizability of the dipole moment and its degree of anharmonicity, so bonds between hydrogen and heavier elements such as S, O, N, and C contribute to high-intensity peaks in the NIR. Extraction of chemical information from the raw NIR spectra requires a chemometric treatment of the data, which can in some cases predict the chemical composition of a sample as well as the concentration of its constituents based on the wavelength and intensities of NIR peaks.[22−25]

Conventional model-building approaches such as partial least-squares and principal component regressions have been used extensively to gather useful information from NIR data.[25−27] While diverse properties of the targets have been studied using these approaches, the emergence of machine learning (ML) has brought a new dimension to the advancement of NIR technology.[28−32]

The chemometric analysis of pharmaceutical formulations classically relies on models that are trained to recognize specific brands of a pharmaceutical product using proprietary libraries and algorithms to identify targets. The brand-specific approach maximizes accuracy at the cost of robustness. These models are useful for product authentication[21,33−36] or manufacturing process control,[27,37,38] but they are not optimal for detection of SFPs because they tend to give false alarms for formulations or brands outside their training set.[39,40] The reliance on libraries of authentic products is a major barrier to the SFP use case.[19,28,33,34] Problems can arise when manufacturers use different excipients, forms of the API (e.g., different crystal polymorphs and particle sizes), and formulation technologies (e.g., coated vs noncoated tablets) in different brands of the same product, or even in different batches of the same brand. These formulation differences can create differences in the NIR spectra that result in identification of a good quality product as an SFP. In theory, manufacturers of pharmaceuticals and portable spectrophotometers could cooperate to build and update comprehensive libraries. In practice, there are thousands of manufacturers and hundreds of portable spectrophotometer brands, and the necessary cooperation has not coalesced. In some cases, "authentic products" provided for a library have been found to be substandard or falsified.[41]

In this work, we aimed to maximize the robustness of NIR identification of formulations of acetaminophen corresponding to SFPs, without needing to build brand-specific libraries. Acetaminophen, or paracetamol, is an inexpensive analgesic and antipyretic on the WHO list of essential medicines, which has been a target of falsification.[42,43] A group of chemometrics approaches and convolutional neural networks (CNNs) were trained on simple lab-formulated binary mixtures, and we then attempted to evaluate the strengths and weaknesses of each of these methods for analysis of samples from outside the training sets. Next, we combined the methods into an algorithm to answer the types of presumptive questions that a drug regulator might ask during postmarket surveillance activities: is this product likely to be falsified? Is it likely to be substandard? The final algorithm was then tested with samples from outside the training set, including commercial acetaminophen products adulterated with an inert filler; lab-made samples that simulated good, substandard, and falsified acetaminophen formulations; and 20 other pharmaceuticals.

2. EXPERIMENTAL SECTION

2.1. Materials and Methods. Pure acetaminophen (AC) was purchased from Sigma-Aldrich. Alpha lactose monohydrate (LA) with purity greater than 99% and isoniazid (IS) with purity greater than 99% were obtained from Sigma-Aldrich, while USP-grade L-ascorbic acid (AA) was purchased from VWR Life Science. Compositions of binary and ternary mixtures are shown in Table S1. Acetaminophen is a common API for pain relief, lactose is a common excipient within the pharmaceutical industry, while ascorbic acid has antioxidant properties and is an ingredient in some cold-relief medicines.

The acetaminophen dosage forms were Tylenol and TopCare brands purchased at a Martin's supermarket in South Bend, Indiana, in 2021. They are TopCare Extra Strength Sweet Coat (TCESSC), TopCare Regular Strength (TCRES), Tylenol Extra Strength (TEXST), Tylenol Rapid Release (TRARE), Tylenol Regular Strength (TREST). Double "00" gelatin caps (NOW, IL) were bought online from Amazon. High-performance liquid chromatography (HPLC) assays (Figure S1) showed all were of good quality. Since dosage forms contain excipients in addition to API, it is common to see API content that goes above 100% of the number of milligrams stated on the package. These branded products were "adulterated" as shown in Table S2.

Twenty APIs and excipients were assessed in these studies: antipyrine, cefuroxime axetil, chloramphenicol, chloroquine, dapsone, digoxin, D-penicillamine, hydrochlorothiazide, isatin, L-citrulline, levofloxacin, lovastatin, metformin hydrochloride, pravastatin, simvastatin, spironolactone, sulfamethoxazole, tetracycline hydrochloride, uric acid, and Verapamil hydrochloride.

2.2. Instrumentation. All spectra were acquired on an LMMI58000060 USB-powered portable NIR spectrophotometer manufactured by InnoSpectra Corporation, Hsinchu, Taiwan. Data were collected from 900 to 1700 nm with a digital resolution of about 4 nm/data point, using 20 scans; acquisition of a spectrum took 20 s. Each spectrum contained 228 data points generated from ISC NIRScan Winform GUI software.

2.3. Data Analysis. The NIR spectrophotometer produces data files with a header section containing sample and spectrum metadata and a list of wavelength and absorption data (an example is the data file labeled exrawdataspec in our

https://doi.org/10.1021/acs.analchem.2c00998
Anal. Chem. XXXX, XXX, XXX−XXX

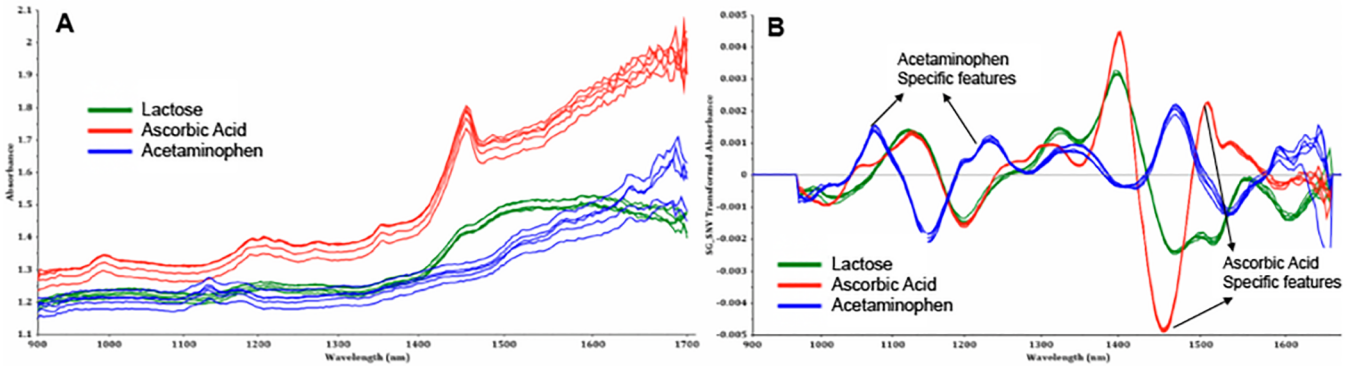


Figure 1. Original spectra (A) and the transformed spectra (B) of pure AC, LA, and AA. This signal processing treatment substitutes each data point with a smoothed estimate resulting from a polynomial regression transformation.

GitHub repository.[44] The raw data files were saved in a folder and transformed to a worksheet containing the key parameters of each spectrum. A macro of this transformation (NIRMacro-Transformation) is in a folder tagged R-PLAYING.[44] The NIR intensity versus wavelength data are saved as RawNIRData. Other key parameters necessary for the different quantitative and qualitative analysis methods are in a csv file named Sample Info. The data files, SampleInfo and RawNIRData, were merged and imported to a multivariate analysis software package, The Unscrambler X version 10.4 (Camo Software, Oslo, Norway). Our raw data, code, and step-by-step directions for data pretreatment, chemometric analysis, and machine learning analysis are archived in our GitHub repository.[44]

3. RESULTS AND DISCUSSION

3.1. Overall Strategy. Nine hundred NIR spectra were generated to train a range of different chemometric models. The spectra included pure acetaminophen, lactose, ascorbic acid, and binary mixtures of acetaminophen with either lactose or ascorbic acid. Each set of spectra was divided into a training set (70%), a validation set for optimizing the training (15%), and a test set (15%) that was reserved for evaluating the performance of the trained and optimized models. After this initial assessment, each model was applied to 400 spectra of ternary mixtures of the same three compounds to evaluate how well it performed with samples outside the boundaries of its training set.

An algorithm combining six of the models trained on the binary mixture data set was developed. Our hypothesis was that the algorithm would be more accurate in quantifying acetaminophen and more specific in rejecting samples that were significantly different from acetaminophen than any of the component models alone. In practical terms, the first capability is related to identification of pharmaceutical products that are substandard due to dilution with an expected excipient, while the second capability is closer to the task of identifying falsified products, degraded products, or ones that are "cut" with an unexpected filler.

To gauge the robustness of the algorithm and the component models on more realistic pharmaceutical samples, we evaluated NIR spectra from five brands of acetaminophen capsules and tablets as well as samples of these dosage forms that had been diluted with lactose. We further challenged the algorithm with a set of blinded lab-made samples that included formulations corresponding to good quality, substandard, and falsified acetaminophen. Lastly, we generated 20 samples from laboratory-grade APIs to test whether the algorithm could reliably reject these substitute pharmaceuticals. For all these experiments, the algorithm and the underlying setups of the models that contributed to the algorithm (e.g., the PCA loadings, SIMCA hyperplane coordinates, and CNN loadings) were held constant based on the original binary mixture training.

Our experimental design incorporated both quantitative and qualitative approaches by exploring molar ratio and percentage of the API and excipients formulated in the lab as a metric for quantitative and qualitative assessments of APIs in actual formulations. We have not optimized the algorithm for accurate concentration determination yet, because the main question for detection of SFPs is whether a substance meets or fails the API content standards.

3.2. Raw Spectra and Data Pretreatment. Raw spectra were generated from the NIR spectrometer enclosed in a three-dimensional (3D) printed case (Figure S3) with a holder for a gel capsule filled with samples.[45−48] A preliminary study (Figure S5) suggested that 100 spectra of a given sample would be sufficient to classify or quantify unknown samples in a supervised learning algorithm. We introduced variability in the samples and sample measurements by generating the 100 spectra of each sample using 10 gelatin capsules filled with a portion of the sample, each repositioned and measured 10 times. The spectra generated for acetaminophen (AC), lactose (LA), and ascorbic acid (AA) were significantly different by visual inspection, and mixtures of these compounds gave additive spectra (Figures 1A & S6).

The raw spectra were normalized and smoothed (Figure 1B). To remove offsets and multiplicative effects, we applied standard normal variate (SNV) preprocessing,[19,49−51] ensuring that every spectrum had a standard deviation of one and a mean of zero (Figure S6). Noise was reduced through Savitzky-Golay smoothing (SG) yielding second-derivative spectra (Figures 1B & S7B) that were subjected to data analysis.[50−52] SG smoothing may distort the shape and the intensity of the spectral bands thereby limiting the optimal model performance.[53−55] However, preliminary studies and exploratory data analysis were carried out to ensure an optimized outcome. Figure S8 shows the scores of the SNV-only (A) treatment, raw data (B), and SNV plus SG (C) treatment of the sample data. SNV plus SG gave superior separation of the samples.

https://doi.org/10.1021/acs.analchem.2c00998
Anal. Chem. XXXX, XXX, XXX−XXX

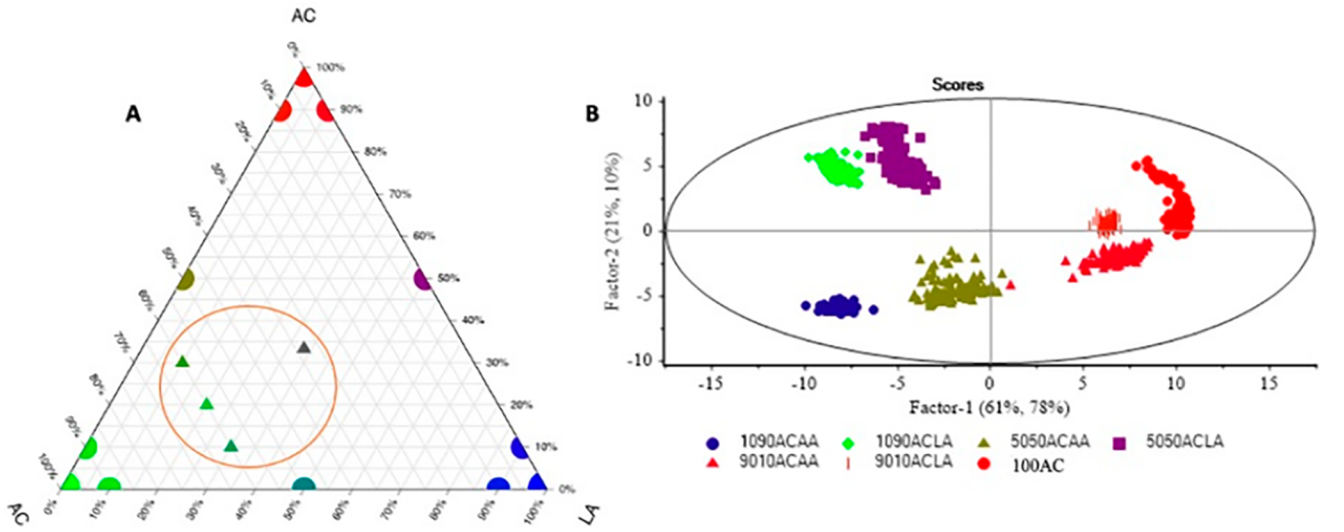


Figure 2. NIR spectra of mixtures of acetaminophen (AC) with lactose (LA) and/or ascorbic acid (AA) were used to train and evaluate the data analysis models. (a) Compositions of the binary and ternary (circled) mixtures in mole%. (b) PC scores for principal components 1 and 2, showing the clustering of NIR spectra corresponding to different binary formulations.

3.3. Classification and Regression Analysis for Qualitative and Quantitative Assessment and Predictions. *3.3.1. Classification.* To answer questions that are relevant for end users or regulators of pharmaceutical products such as acetaminophen, we approached the classification with two questions in mind. In case 1, is the formulation consistent with authentic product (authentic), or not consistent with authentic product (suspicious)? And in case 2, is the sample's AC content within the allowed threshold provided by the regulators (meets the 90−110% API content standard, OK), present but below the required threshold (substandard, SUB), or not determined (No API or a substitute API, FAIL)? After training and optimizing the PCA/SIMCA (Figure 2B) and SVM models on 85% of the 900 pure and binary mixture spectra, the classification accuracies of each model on the Case 1 and Case 2 tasks were evaluated using the remaining 15% of the spectra.

Not surprisingly, both models performed well when tested against samples similar to their training set. The more computationally intensive SVM model gave a validation accuracy of 100% in case 1 and 99.9% in case 2. SIMCA based on PC scores gave a validation accuracy of 100% in case 1 and 89% in case 2. However, all SIMCA's misclassifications in case 2 were due to the SIMCA algorithm classifying FAIL samples as SUB. Since both categories would fail to meet the regulatory standard for API content, the accuracy for discriminating between samples that fail or meet the standard was 100%.

Next, without any additional training, the models were tested with 400 spectra of lab-formulated ternary mixtures of AC mixed with both LA and AA. The prediction accuracy for both classes was 100% for SIMCA, while SVM was 100% in case 1 and 96% in case 2.

The traditional approach to NIR data analysis is to generate spectra from authentic samples of the products that are being analyzed in the field setting and use these to train various data analytics models. These models give excellent results for samples that fall into their training set but may fail for samples outside the training set. We followed this traditional approach to build PCA and SVM models for three Tylenol and two TopCare brands of acetaminophen tablets and capsules. Each of the five brands is designated as 100% in "brand content." Mixtures were then prepared in varying proportions with lactose (LA, w/w) to simulate adulteration of the branded products. In addition to the branded product, samples were prepared with 90%, 50%, or 10% brand content. Five SIMCA and five SVM models were trained, one for each brand; test results gave classification accuracies of 96−100%. However, when each model trained on one brand of acetaminophen was tested on the reserved samples from the other brands of acetaminophen, the classification accuracies dropped to 60−76% for the SIMCA models and 98−100% for the SVM models.

Each model was then tested on binary mixtures of acetaminophen with lactose or ascorbic acid. While SVM performed well in detecting falsified samples (case 1) with classification accuracy of 100%, its accuracy for detection of substandard samples (case 2) was between 79% and 87%, depending on which brand the model had been trained on. SIMCA's classification accuracy was 100% for all the binary samples in both cases. Next, we tested each of the branded models on ternary mixtures of acetaminophen with lactose and ascorbic acid. SVM yielded classification accuracy of between 70 and 75% for both case 1 and case 2 classifications, while SIMCA failed to classify any of the samples correctly. Here we observed that, when the test samples possess similar characteristics to the training samples, SIMCA offers a better performance, while analysis involving different matrices yielded better performance with the SVM algorithm. This finding aligned with Ŕacz et al.'s work on SIMCA exploration in classification as an algorithm that anchors on similarities among samples within the same class/group rather than on the differences between the groups, making it more effective as a classification method for identical samples.[56]

Since the models that were trained on lab-produced binary mixtures gave better performance for classification of ternary mixtures than the models that were trained on specific brands, we used lab-produced binary mixtures for the next sets of experiments, which focused on semiquantitative assessment.

https://doi.org/10.1021/acs.analchem.2c00998
Anal. Chem. XXXX, XXX, XXX−XXX

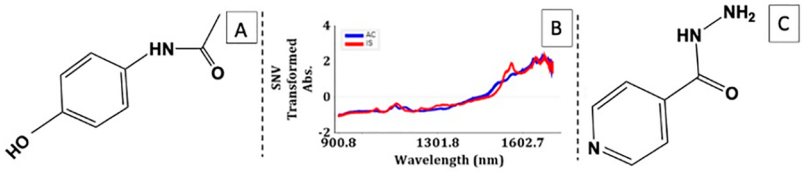


Figure 3. SNV transformed spectra of acetaminophen, AC (A), spectra plots (B), and isoniazid, IS (C).

3.3.2. Regression. We developed regression models based on PC and SVM in order to detect substandard products. Because our experimental data do not span the concentration space evenly but focus on the regulatory threshold level near 90% API content, the resulting regression models should not be seen as tools for quantitative assay, particularly for API levels that are different from the regulatory threshold levels.

Partial least-squares regression (PLS-R) creates a linear model in reduced dimensions through which the observables can be used to predict concentration values for unknown samples. SVM-R, a nonparametric technique, uses support vectors and kernel functions in its predictions.[19] The PLS-R and SVM-R models were again trained using 85% (70% training set and 15% validation set) of the 900 spectra generated from the binary mixtures; some of the binary mixtures were diluted with lactose and others with ascorbic acid, but the models were trained only to predict the concentration of the acetaminophen component. The regression accuracies for these models were evaluated using the 15% of the spectra that had not been used in the training. The prediction accuracy for these regressions is defined as the predicted API content divided by the actual API content. PLS-R yielded a prediction accuracy of 98%, while SVM-R yielded a prediction accuracy of 99%.

The models trained on the binary mixtures were then employed to quantify acetaminophen in the lab-formulated ternary mixtures. PLS-R and SVM-R performed exceptionally well in the quantitation of the API, as the predicted values are all within the expected quantity with prediction accuracy of greater than 96%. These findings are important because the most encountered types of bad-quality pharmaceuticals are substandard products, and existing field screening devices are generally bad at detecting these.

As expected, regression models trained on specific brands performed better in analysis of those brands than models trained on the binary mixtures.

PLS-R and SVM-R models were developed by training the models on each brand's product either pure or adulterated with varying amounts of lactose, LA, and then, their performances were evaluated. The prediction accuracies of all five models ranged from 96% to 99%. SVM-R performed slightly better than PLS-R (Table S3) with better root-mean-square error (RMSEV) and correlation coefficients.

Although the SVM-R and PLS-R models that had been trained on specific brands gave superior performance in analysis of those brands, when they were used to evaluate acetaminophen content in ternary mixture samples, their performance dropped significantly, to 85 and 82% prediction accuracy, respectively.

3.3.3. Convolutional Neural Networks. Neural networks have been setting new benchmarks for regression and classification tasks in many fields for the past decade.[57−60] Zhang et al. used multimodal convolutional neural networks (CNNs) to classify NIR data of tobacco samples from four different countries.[61] We explored a similar approach for acetaminophen samples. One major difference of our work in comparison with that of Zhang et al. is that we trained our CNN models for regression tasks instead of classification. With a regression objective, the model can focus on predicting the molar or mass content of AC instead of forcibly classifying a data sample into one of the class labels. CNN has been explored both in data and image analysis. In this work, our goal was to use CNN to broaden the conventional "linear" approaches, which was why we tried multiple neural network (NN) designs, including two-dimensional (2D) CNN. We tried several CNN designs; a more detailed discussion of the network architectures, model parameters, and optimization can be found in the Supporting Information. The NIR data can be viewed either as a one-dimensional (1D) vector that can be analyzed with 1D-CNN or as a grayscale image that can be analyzed with 2D-CNN (Figure S4). We varied the network architecture[31,59,60] of both the 1D and 2D models (Table S4) to optimize their performance.

Each of the models was trained using 85% (70% training set and 15% validation set) of the NIR spectra from the lab-formulated binary mixtures of AC with AA or LA. The models were then tested on the remaining 15% of the NIR spectra. All the CNN models performed above 96% prediction accuracy when tested on the lab-formulated binary samples, with root-mean-square error of prediction (RMSEP) values less than 5%. This confirms that neural networks are generally robust for detecting pharmaceutical products that are substandard due to dilution with an expected component. However, substandard pharmaceutical products often contain unexpected components, and neural networks are notoriously fragile when confronted with data from outside their training sets,[62] so we next tested them with the ternary mixture data. Each model showed good robustness despite the common weakness of supervised learning algorithms in external validations.[63] The RMSEP of each model was surprisingly less than 10% (Table S4). However, when the CNN models were used to analyze the blinded samples (see Supporting Information) corresponding to good quality, substandard, or falsified acetaminophen formulations, their performances were poor for the falsified formulations, leading to average RMSEP errors for the acetaminophen concentrations greater than 30% (Table S4). The CNN models failed completely for the blinded samples containing isoniazid (IS) (whose NIR spectrum resembles that of acetaminophen, see Figure 3) whereas they performed well for starch (a polysaccharide whose functional groups are similar to those found in lactose).

3.4. Integration of PCA, SVM, and CNN Models. In order to maximize the robustness of detection of SFPs, we developed an algorithm that combines the strengths of six of the models described in sections 3.3.1, 3.3.2, and 3.3.3. The models included SIMCA and SVM classification, PLS and SVM regressions, and LeNet and 2D CNN models (the CNNs that gave the best RMSEP for the "blinded" samples, Table

https://doi.org/10.1021/acs.analchem.2c00998
Anal. Chem. XXXX, XXX, XXX−XXX

S4). All models in the algorithm use the lab-produced binary mixture data set as their training set.

The goal of the algorithm is to classify an NIR spectrum of an unknown pharmaceutical sample as "OK" (meets standard), "SUB" (substandard), or "FAIL" (implies an unexpected component is present or the acetaminophen concentration is less than 10%). The algorithm prediction pathway is shown in Figure 4. Five "cells" each contribute one vote toward the final prediction. The first step in the algorithm is to compare the classes assigned by the two classification models SVM-C and SIMCA. If these classes (OK/SUB/FAIL) agree, the content of the "classification" cell is set to that value. If the SVM-C and SIMCA models do not agree, the classification cell content is set to FAIL. The next step is to model the acetaminophen content using the four regression models (PLS-R, SVM-R, LeNet, and 2D-CNN). Each of these models has an individual voting cell. If the value of a model's predicted concentration of acetaminophen is between 0.80 and 1.20, the model votes OK, and if the predicted concentration is less than 0.8 or greater than 1.20, the model votes SUB. If the model predicts a concentration that is less than 0.30, the model's vote switches to FAIL. Finally, the algorithm counts the votes. The algorithmic prediction is the result found in at least three of the voting cells in a majority rule approach. If there is no majority vote, the algorithmic prediction is set to FAIL.

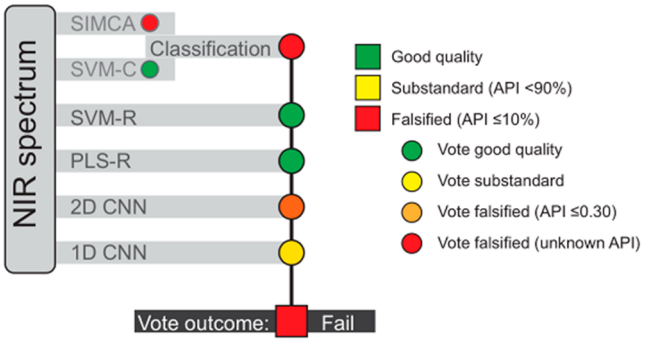


Figure 4. Flowchart for the algorithm. If the two classification models disagree, their vote is set to FAIL; if they agree, their vote is set to that value. For the regression models, predictions between 0.80 and 1.20 are set to OK, predictions at most 0.30 are classified as FAIL, and predictions between 0.30 and 0.80 or greater than 1.20 are classified as SUB.

The algorithm was tested on NIR spectra from three classes of validation samples, including (a) binary mixtures that covered the training set compositions and extended to totally unfamiliar compositions (these samples were independent from the original binary training/testing set), (b) five authentic acetaminophen formulations, some of which were intentionally adulterated with lactose, and (c) 20 other common pharmaceuticals.

3.4.1. Algorithm Performance on "Lab-Made" Samples. To see if the algorithm could distinguish samples from both inside and outside its training set, it was used to classify 140 spectra corresponding to binary and ternary mixtures of pure acetaminophen and lactose or ascorbic acid as well as pure isoniazid (IS), an isoniazid/acetaminophen mixture, and cornstarch. Isoniazid was selected because several regions of the AC and IS spectra are superficially similar (Figure 3), and starch is a polysaccharide whose NIR spectrum mimics that of lactose to some degree. The algorithm correctly identified the three samples that met the monograph API content standard (Table S8), correctly identified the two binary mixtures whose API content was substandard, identified both binary mixtures

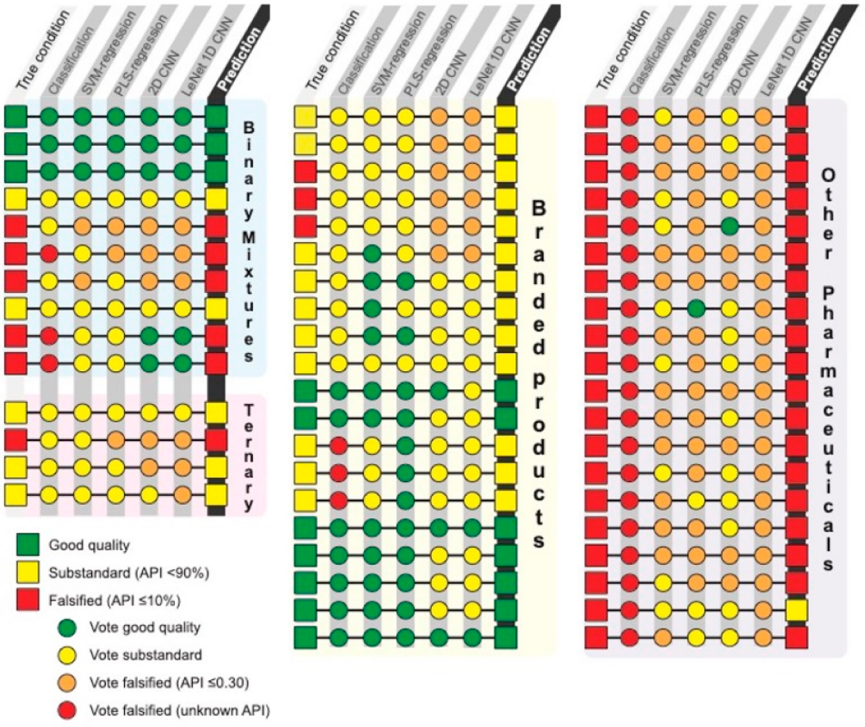


Figure 5. Summary of the model and algorithm predictions for binary and ternary mixtures of acetaminophen with lactose and ascorbic acid, branded acetaminophen products with lactose diluent, and other pharmaceuticals.

https://doi.org/10.1021/acs.analchem.2c00998
Anal. Chem. XXXX, XXX, XXX−XXX

with 10% API content as falsified, and could not classify the samples containing either isoniazid or starch, correctly indicating that an unexpected NIR-active substance was present in each for a classification of falsified. Three ternary mixtures were correctly classified as substandard, and the fourth (containing just 10% acetaminophen) was classified as falsified (Figure 5, Table S7). No individual model performed better than the algorithm in classification of the binary and ternary samples.

3.4.2. Algorithm Performance on Acetaminophen Dosage Forms. The algorithm was then challenged with NIR spectra generated from five brands of acetaminophen tablets, some cut with lactose to simulate substandard or falsified products. These samples were either good quality (OK, API content > 90%), substandard (SUB, 90% > API content > 10%), or falsified (FAIL, API content < 10%). The algorithm identified all the good quality and substandard formulations correctly; three falsified formulations containing slightly less than 10% API were identified as substandard. SIMCA classified 617 out of the 1249 spectra correctly, including identifying all the seriously adulterated samples (containing less than 90% of brand content) as substandard (SUB). However, SIMCA was unable to classify three of the branded drugs (TREST, TRARE, and TEXST) with 10% lactose content (Table S5). This created a vote of FAIL from the classification cell, which was overridden by the SUB votes from the regression cells in the next step of the algorithm. The more computationally intensive SVM-C model classified all the samples correctly. SVM-R performed poorly by predicting most spectra of samples that had been cut to 50% AC content as good quality. LeNet and the 2DCNN model mistakenly identified several of the branded products as substandard, not surprising given the unfamiliar matrix of these products, but were outvoted by the classification cell and the other two regression models.

3.4.3. Algorithm performance on other pharmaceuticals. Finally, we investigated the ability of the algorithm to discriminate acetaminophen from other APIs. Twenty common pharmaceuticals containing a broad range of functional groups were selected, and five NIR spectra were acquired from each substance. The algorithm then classified each spectrum as either FAIL, SUB, or OK; all the samples were appropriately classified as falsified (FAIL) except for uric acid, which was classified as substandard (Table S6).

4. CONCLUSION

Portable NIR spectrometers could be a game-changing technology for uncovering SFP products in low-resource settings, but the need to build libraries of spectra from authentic samples of the products that are to be analyzed has hampered their implementation in low- and middle-income countries. Efforts have been made to identify counterfeit acetaminophen products in the past, but most of the approaches used a brand-specific library in training the algorithm.[64,65] Here, we trained six models using NIR spectra that were generated from lab-prepared binary mixtures of acetaminophen and two cutting agents. All the models performed very well when evaluated individually through the normal validation process. However, when we challenged each model to evaluate the quality of NIR spectra in 54 samples that were different from the training samples, each model had weak spots. For example, the classification voting cell erred in 17% of the cases because these methods tended to identify samples with unfamiliar excipients as substandard or falsified. The two regression models erred in 27% of the cases; they had trouble identifying substandard products unless the API content was below 50% and mistook several of the substitute APIs for substandard acetaminophen. The two neural nets erred in 32% of the cases; both misidentified isoniazid or acetaminophen/isoniazid mixtures as good quality acetaminophen and misidentified three good quality acetaminophen products as substandard.

A simple voting algorithm was used to combine the strengths of the six models. The algorithm predicted the correct classification in 93% of the cases. The improved robustness of the NIR data analysis with a simple combination of six models is encouraging. Much future work will be required to optimize the choice of the classification/regression methods in the ensemble, minimize the computational overhead, and weight the contributions of individual models to the overall decision. This is outside the scope of this work, but it is important to investigate. To support that work, we are building up a data set that includes more pharmaceuticals, a wider range of dosage forms, and possibly, other types of spectroscopic data such as IR. We do not yet know whether the selection of the diluents used in the binary mixtures needs to be optimized for the algorithm to recognize other active pharmaceutical ingredients, and it is possible that different data analysis models or different weighting of the individual models in the algorithm could further improve the overall performance. Further work is in progress to broaden these investigations to other active pharmaceutical ingredients that are used in medicines on the WHO essential medicines list and to evaluate the performance of the NIR analysis on a wide range of pharmaceutical products from many manufacturers across the globe.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c00998.

Additional experimental details, materials, and methods, including link to GitHub archive for raw data and code, description of compositions of all samples, sample spectra before and after processing, tables showing numerical prediction accuracies and RMSEV or RMSEP values of individual models, and predictions of the individual models and combined algorithm for NIR spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Marya Lieberman − *Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States;* orcid.org/0000-0003-3968-8044; Email: mlieberm@nd.edu

Authors

Olatunde Awotunde − *Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States*

Nicholas Roseboom − *Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States*

https://doi.org/10.1021/acs.analchem.2c00998
Anal. Chem. XXXX, XXX, XXX−XXX

Jin Cai − *Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States*
Kathleen Hayes − *Department of Chemistry and Biochemistry, University of Notre Dame, Notre Dame, Indiana 46556, United States;* orcid.org/0000-0003-1217-0050
Revati Rajane − *Precise Software Solutions Inc, Rockville, Maryland 20850, United States*
Ruoyan Chen − *Precise Software Solutions Inc, Rockville, Maryland 20850, United States*
Abdullah Yusuf − *Precise Software Solutions Inc, Rockville, Maryland 20850, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c00998

Notes

The authors declare the following competing financial interest(s): Precise Software Solutions, which performed the CNN modeling of the NIR data, is a startup company developing spectroscopic and data analytic tools for assessing the quality of pharmaceuticals.

■ ACKNOWLEDGMENTS

The authors thank ternaryplot.com for access to their software. This work was supported by NSF (CMMI-1842369, OISE-1559497). An undergraduate research stipend for N.R. was provided by Precise Software Solutions, Inc.

■ REFERENCES

(1) WHO. Seventieth World Health Assembly Update, 29 May 2017;Vol. *41*.http://www.who.int/mediacentre/news/releases/2017/dementia-immunization-refuguees/en/.
(2) Callister, L. C. *MCN Am. J. Matern. Nurs.* **2019**, *44* (6), 361−361.
(3) World Health Organization. (2019). The WHO Member State Mechanism on Substandard and Falsified Medical Products: How WHO Member States Work Together to Safeguard Access to Safe and Effective Medicines, Vaccines and Other Medical Products; World Health Organization, 2019. https://apps.who.int/iris/handle/10665/332635?locale-attribute=pt&.
(4) The Pathology of Negligence (Report of the Judicial Inquiry Tribunal to Determine the Causes of Deaths of Patients at the Punjab Institute of Cardiology, Lahore 2011−2012); 2012, https://padproject.nd.edu/assets/206307/pathology_of_negligence_pic_drug_inquiry_report_2012.pdf.
(5) Roth, L.; Bempong, D.; Babigumira, J. B.; Banoo, S.; Cooke, E.; Jeffreys, D.; Kasonde, L.; Leufkens, H. G. M.; Lim, J. C. W.; Lumpkin, M.; Mahlangu, G.; Peeling, R. W.; Rees, H.; Ndomondo-Sigonda, M.; Stergachis, A.; Ward, M.; Nwokike, J. *Global Health* **2018**, *14* (1), 1−12.
(6) Chen, H.-H.; Higgins, C.; Laing, S. K.; Bliese, S. L.; Lieberman, M.; Ozawa, S. *Med. Access @ Point Care* **2021**, *5*, No. 23992026209803.
(7) Cockburn, R.; Newton, P. N.; Agyarko, E. K.; Akunyili, D.; White, N. J. *PLoS Med.* **2005**, *2* (4), 0302−0308.
(8) Mackey, T. K.; Liang, B. A.; York, P.; Kubic, T. *Am. J. Trop. Med. Hyg.* **2015**, *92* (6), 59−67.
(9) Roth, L.; Biggs, K. B.; Bempong, D. K. Substandard and Falsified Medicine Screening Technologies. *AAPS Open* **2019**, *5* (1). DOI: 10.1186/s41120-019-0031-y.
(10) Land, T. *Nature* **1992**, *355* (6357), 192.
(11) Martino, R.; Malet-Martino, M.; Gilard, V.; Balayssac, S. *Anal Bioanal Chem.* **2010**, *398*, 77−92.
(12) Kovacs, S.; Hawes, S. E.; Maley, S. N.; Mosites, E.; Wong, L.; Stergachis, A. *PLoS One* **2014**, *9* (3), e90601.
(13) Zambrzycki, S. C.; Caillet, C.; Vickers, S.; Bouza, M.; Donndelinger, D. V.; Geben, L. C.; Bernier, M. C.; Newton, P. N.; Fernández, F. M. *PLoS Negl. Trop. Dis.* **2021**, *15* (9), 1−19.
(14) Caillet, C.; Vickers, S.; Zambrzycki, S.; Luangasanatip, N.; Vidhamaly, V.; Boutsamay, K.; Boupha, P.; Lubell, Y.; Fernández, F. M.; Newton, P. N. *PLoS Negl. Trop. Dis.* **2021**, *15* (9), 1−6.
(15) Caillet, C.; Vickers, S.; Vidhamaly, V.; Boutsamay, K.; Boupha, P.; Zambrzycki, S.; Luangasanatip, N.; Lubell, Y.; Fernandez, F. M.; Newton, P. N. *PLoS Med.* **2021**, *18* (9), 1−12.
(16) Caillet, C.; Vickers, S.; Zambrzycki, S.; Fernández, F. M.; Vidhamaly, V.; Boutsamay, K.; Boupha, P.; Peerawaranun, P.; Mukaka, M.; Newton, P. N. *PLoS Negl. Trop. Dis.* **2021**, *15* (9), 1−22.
(17) Srivastava, H. K.; Wolfgang, S.; Rodriguez, J. D. *Anal. Chim. Acta* **2016**, *914*, 91−99.
(18) Pasquini, C. *J. Braz. Chem. Soc.* **2003**, *14* (2), 198−219.
(19) Roggo, Y.; Chalus, P.; Maurer, L.; Lema-Martinez, C.; Edmond, A.; Jent, N. *J. Pharm. Biomed. Anal.* **2007**, *44* (3), 683−700.
(20) Pasquini, C. *Anal. Chim. Acta* **2018**, *1026*, 8−36.
(21) Sarraguça, M. C.; Cruz, A. V.; Amaral, H. R.; Costa, P. C.; Lopes, J. A. *Anal. Bioanal. Chem.* **2011**, *399* (6), 2137−2147.
(22) Meglen, R. R. *Mar. Chem.* **1992**, *39*, 217−237.
(23) Dégardin, K.; Guillemain, A.; Guerreiro, N. V.; Roggo, Y. *J. Pharm. Biomed. Anal.* **2016**, *128*, 89−97.
(24) Liang, C.; Yuan, H. F.; Zhao, Z.; Song, C. F.; Wang, J. J. *Chemom. Intell. Lab. Syst.* **2016**, *153*, 51−57.
(25) Deidda, R.; Sacre, P. Y.; Clavaud, M.; Coïc, L.; Avohou, H.; Hubert, P.; Ziemons, E. *TrAC - Trends Anal. Chem.* **2019**, *114*, 251−259.
(26) Grabska, J.; Beć, K. B.; Mayr, S.; Huck, C. W. *Appl. Spectrosc.* **2021**, *75* (8), 1022−1032.
(27) Mainka, D. L.; Link, A. *Anal. Methods* **2019**, *11* (40), 5185−5194.
(28) Blanco, M.; Peguero, A. *TrAC - Trends Anal. Chem.* **2010**, *29* (10), 1127−1136.
(29) Chen, Y.; Bin, J.; Zou, C.; Ding, M. *J. Anal. Methods Chem.* **2021**, *2021*, 1.
(30) Ma, J.; Sheridan, R. P.; Liaw, A.; Dahl, G. E.; Svetnik, V. *J. Chem. Inf. Model.* **2015**, *55* (2), 263−274.
(31) Guo, C.; Pleiss, G.; Sun, Y.; Weinberger, K. Q. On Calibration of Modern Neural Networks. 34th International Conference on Machine Learning, ICML 2017. 2017; pp 2130−2143.
(32) Lecun, Y.; Bengio, Y.; Hinton, G. *Nature* **2015**, *521* (7553), 436−444.
(33) Said, M. M.; Gibbons, S.; Moffat, A. C.; Zloh, M. *Int. J. Pharm.* **2011**, *415*, 102−109.
(34) Torniainen, J.; Afara, I. O.; Prakash, M.; Sarin, J. K.; Stenroth, L.; Töyräs, J. *Anal. Chim. Acta* **2020**, *1108*, 1−9.
(35) Dowell, F. E.; Maghirang, E. B.; Fernandez, F. M.; Newton, P. N.; Green, M. D. *J. Pharm. Biomed. Anal.* **2008**, *48*, 1011−1014.
(36) Wiedemair, V.; Huck, C. W. *Talanta* **2018**, *189* (March), 233−240.
(37) Wu, H.; Tawakkul, M.; White, M.; Khan, M. A. *Int. J. Pharm.* **2009**, *372* (1−2), 39−48.
(38) Skibsted, E. T. S.; Boelens, H. F. M.; Westerhuis, J. A.; Witte, D. T.; Smilde, A. K. *J. Pharm. Biomed. Anal.* **2006**, *41* (1), 26−35.
(39) Cantor, S. L.; Hoag, S. W.; Ellison, C. D.; Khan, M. A.; Lyon, R. C. *AAPS PharmSciTech* **2011**, *12* (1), 262−278.
(40) Rodionova, O. Y.; Titova, A. V.; Balyklova, K. S.; Pomerantsev, A. L. *Talanta* **2019**, *205* (5), 120150.
(41) Xiang, D.; Konigsberger, M.; Wabuyele, B.; Hornung, K.; Cheney, J. *Analyst* **2009**, *134* (7), 1405−1415.
(42) Ziémons, E.; Mantanus, J.; Lebrun, P.; Rozet, E.; Evrard, B.; Hubert, P. *J. Pharm. Biomed. Anal.* **2010**, *53* (3), 510−516.
(43) Lawson, G.; Ogwu, J.; Tanna, S. *PLoS One* **2018**, *13*, e0202059 DOI: 10.1371/journal.pone.0202059.
(44) Awotunde, O.; Roseboom, N.; Cai, J.; Rajane, R.; Ruoyan, C.; Lieberman, M. Discrimination-of-substandard-and-falsified-formulations-from-genuine-drugs-using-NIR-and-ML https://github.com/

aawotund/Discrimination-of-substandard-and-falsified-formulations-from-genuine-drugs-using-NIR-and-ML.
(45) Sparén, A.; Malm, M.; Josefson, M.; Folestad, S.; Johansson, J. *Appl. Spectrosc.* **2002**, *56* (5), 586−592.
(46) Yoon, W. L.; Moffat, A. C. *Analyst* **1998**, *123* (5), 1029−1034.
(47) Dalvi, H.; Langlet, A.; Colbert, M. J.; Cournoyer, A.; Guay, J. M.; Abatzoglou, N.; Gosselin, R. *Talanta* **2019**, *195*, 87−96.
(48) Mahakul, R.; Nath Thatoi, D.; Choudhury, S.; Patnaik, P. *Mater. Today Proc.* **2021**, *41*, 340−346.
(49) Blanco, M.; Coello, J.; Iturriaga, H.; Maspoch, S.; Pagès, J. *Chemom. Intell. Lab. Syst.* **2000**, *50* (1), 75−82.
(50) Rajalahti, T.; Kvalheim, O. M. *Int. J. Pharm.* **2011**, *417* (1−2), 280−290.
(51) Martens, H. *J. Chemom.* **2015**, *29* (11), 563−581.
(52) Heil, K.; Schmidhalter, U. *Sensors* **2021**, *21* (4), 1−24.
(53) Engel, J.; Gerretzen, J.; Szymańska, E.; Jansen, J. J.; Downey, G.; Blanchet, L.; Buydens, L. M. C. *Chem.* **2013**, *50*, 96−106.
(54) Beć, K. B.; Huck, C. W. *Front. Chem.* **2019**, *7* (2), 1−22.
(55) De Araújo, S. C.; Kawano, Y. *J. Appl. Polym. Sci.* **2002**, *85*, 199−208.
(56) Rácz, A.; Gere, A.; Bajusz, D.; Héberger, K. *RSC Adv.* **2018**, *8* (1), 10−21.
(57) Balabin, R. M.; Smirnov, S. V. *Anal. Chim. Acta* **2011**, *692* (1−2), 63−72.
(58) Neubauer, C. *IEEE Transactions on Neural Networks* **1998**, 685−696.
(59) Petrovic, J.; Chansanroj, K.; Meier, B.; Ibric, S.; Betz, G. *Eur. J. Pharm. Sci.* **2011**, *44*, 227−234.
(60) Ni, Y.; Song, R.; Kokot, S. *Spectrochim. ACTA PART A Mol. Biomol. Spectrosc.* **2012**, *96*, 252−258.
(61) Zhang, L.; Ding, X.; Hou, R. *J. Anal. Methods Chem.* **2020**, *2020*, 1.
(62) Le Page, A. L.; Ballot, E.; Truntzer, C.; Derangère, V.; Ilie, A.; Rageot, D.; Bibeau, F.; Ghiringhelli, F. *Sci. Rep.* **2021**, *11* (1), 1−8.
(63) Patil, A.; Rane, M. *Smart Innov. Syst. Technol.* **2021**, *195*, 21−30.
(64) Scafi, S. H. F.; Pasquini, C. *Analyst* **2001**, *126* (12), 2218−2224.
(65) Eustaquio, A.; Blanco, M.; Jee, R. D.; Moffat, A. C. *Anal. Chim. Acta* **1999**, *383*, 283−29.

https://doi.org/10.1021/acs.analchem.2c00998
Anal. Chem. XXXX, XXX, XXX−XXX