Detection Method of Falsified Medicines by Using a Low-Cost Raman Scattering Spectrometer Combined with Soft Independent Modeling of Class Analogy and Partial Least Squares Discriminant Analysis

Tomoko Sanada, Naoko Yoshida, Kazuko Kimura, and Hirohito Tsuboi

Clinical Pharmacy and Healthcare Sciences, Faculty of Pharmacy, Institute of Medical, Pharmaceutical and Health Sciences, Kanazawa University; Kakuma-machi, Kanazawa, Ishikawa 920–1192, Japan; AI Hospital/Macro Signal Dynamics Research and Development Center, Institute of Medical, Pharmaceutical and Health Sciences, Kanazawa University; Kakuma-machi, Kanazawa, Ishikawa 920–1192, Japan; and Medi-Quality Security Institute, Graduate School of Medical Sciences, Kanazawa University; Kakuma-machi, Kanazawa, Ishikawa 920–1192, Japan.

Received December 27, 2020; accepted February 26, 2021

There are many reports of falsified medicines that may cause harm to patients. A rapid and simple method of identifying falsified medicines that could be used in the field is required. Although Raman scattering spectroscopy has become popular as a non-destructive analysis, few validation experiments on falsified medicines that are actually distributed on the market have been conducted. In this study, we validated a discriminant analysis using an ultra-compact, portable, and low-cost Raman scattering spectrometer combined with multivariate analysis. The medicines were three types of erectile dysfunction therapeutic tablet and one type of antifungal tablet: tadalafil (Cialis), vardenafil hydrochloride (Levitra), sildenafil citrate (Viagra), and fluconazole (Diflucan), which is sometimes advertised as female Viagra. For each medicine, the authentic standard product and products obtained by personal import via the internet (genuine or falsified) were used. Discriminant analyses were performed on the Raman spectra combined with soft independent modeling of class analogy (SIMCA) and partial least squares discriminant analysis (PLS-DA). It was possible to identify all falsified samples by SIMCA using the standard product model for all four products. Using the PLS-DA using the PLS models of the four standard products, falsified Levitra and Diflucan samples were classified correctly, although some falsified Cialis and all Viagra samples also belonged to the standard class. In this study, SIMCA might be more suitable than PLS-DA for identifying falsified medicines. A spectroscopic module that combines the low-cost Raman scattering spectroscopy with SIMCA might contribute to the rapid identification of falsified medicines in the field.

Key words discriminant analysis; falsified medicine; Raman scattering spectroscopy; low-cost analyzer; soft independent modeling of class analogy; partial least squares discriminant analysis

INTRODUCTION

Falsified medicines that intentionally give false or misleading information about the product pose a risk to human health. According to a report from the WHO in 2017, it is estimated that 10% of medical products are substandard or falsified in low- and middle-income countries (LMICs). In Japan, the main inflow route of falsified medicines is personal import via the internet. Although the government has taken measures, such as closing illegal internet sites and calling people’s attention to the risk of personal import, falsified medicines have been confirmed in Japan. To prevent the spread of falsified medicines, it is necessary to develop detection methods for falsified medicines, as the Asia Pacific Economic Cooperation and WHO point out. Since the main detection methods reported thus far have been destructive analyses, such as HPLC and TLC, a rapid and simple analytical method is needed, especially to test a large number of samples. Various portable spectroscopic analyzers have been developed to confirm the identity of batches of starting materials following the Pharmaceutical Inspection Convention and Pharmaceutical Inspection Co-operation Scheme.

As a rapid non-destructive analysis, spectroscopic analyses such as near-IR (NIR) spectroscopy, Fourier transform IR spectroscopy, NMR spectroscopy, and Raman spectroscopy have been developed. NIR spectroscopy is the major non-destructive method in the pharmaceutical field that does not require sample pretreatment. Raman spectroscopy is also becoming popular and its utility is being evaluated. Although there are various portable analyzers that can be used in the field, Raman scattering analyses are relatively higher cost, and this can inhibit their introduction in some areas, such as medical institutions and LMICs. Therefore, we examined the application of an ultra-compact Raman scattering analyzer developed in Japan (C13560), which is a low-cost portable analyzer that can be used in the field.

Multivariate analysis can be combined with Raman scattering analysis and NIR spectroscopy because it processes and discriminates huge amounts of spectral data. The methods known include principal component analysis (PCA), which is used for qualitative analysis; classical least squares (CLS), principal component regression (PCR), and partial least squares (PLS) are used for quantitative analysis. PCA is an unsupervised pattern recognition method and is used for data compression and dimension removal. Soft independent modeling of class analogy (SIMCA) and PLS-discriminant analysis (DA) are supervised pattern recognition. Spectroscopic analysis combined with SIMCA and PLS-DA has often been applied to the analysis of such items as foods, oils, inks, banknotes, and cloth, and its effectiveness has been confirmed in various fields. SIMCA is an unsupervised pattern recognition method and is used for data compression and dimension removal. Soft independent modeling of class analogy (SIMCA) and PLS-discriminant analysis (DA) are supervised pattern recognition. Spectroscopic analysis combined with SIMCA and PLS-DA has often been applied to the analysis of such items as foods, oils, inks, banknotes, and cloth, and its effectiveness has been confirmed in various fields.

© 2021 The Pharmaceutical Society of Japan
has been evaluated. Although the usefulness of spectroscopic analysis combined with multivariate analysis is also being investigated in pharmaceuticals, the method should be validated with falsified products that are actually distributed on the market. Falsified medicines distributed on the market may show no or wrong information about their contents, contain no active pharmaceutical ingredient, and some falsified medicines are very similar to the genuine products because companies manufacturing falsified products have improved their techniques.

In our previous study, which used an ultra-compact Raman scattering spectrometer to discriminate between genuine and falsified tablets by PCA, the methods of discriminant analysis were not considered. To solve this issue, we developed a detection method that combines Raman scattering spectroscopy with SIMCA and PLS-DA. We used authentic standard products as a calibration set, and medicines actually distributed on the market (genuine and falsified) as a test set. The medicines were four products for which falsified products have been reported before. Tadalafil (Cialis), vardenafil hydrochloride (Levitra), and sildenafil citrate (Viagra) tablets, which are medicines for erectile dysfunction (ED), and fluconazole (Diflucan) tablet is an antifungal medicine, which is sometimes advertised as female Viagra online.

MATERIALS AND METHODS

Materials The medicines were Cialis 20-mg tablets, Levitra 20-mg tablets, Viagra 100-mg tablets, and Diflucan 100-mg tablets. Authentic standards of Cialis 20-mg and Levitra 20-mg tablets were obtained legally through the Japanese market. Authentic standards of Viagra 100-mg and Diflucan 100-mg tablets were donated from Pfizer Inc. (New York, NY, U.S.A.) because the dosage form is not approved in Japan. We tested the samples (genuine and falsified) obtained by personal import via the internet. The authenticity of the samples was judged by the manufacturers, and the quality was confirmed using the HPLC method used in previous studies. We also obtained one Diflucan falsified sample from the manufacturer.

Cialis samples (n = 33) consisted of one standard product, nine genuine, and 23 falsified samples. Levitra samples (n = 23) consisted of one standard product, nine genuine, and 13 falsified samples. Viagra samples (n = 23) consisted of one standard product, four genuine, and 18 falsified samples. Diflucan samples (n = 12) consisted of one standard product, nine genuine, and two falsified samples. Since there was only one falsified Diflucan sample obtained by personal import, one other Diflucan falsified sample was donated by the manufacturer.

Raman Scattering Analysis The Raman spectrum of the tablet surface was measured using an ultra-compact Raman scattering analyzer (C13560; Hamamatsu Photonics KK, Shizuoka, Japan). The wavelength was 785 nm, the power was 15 mW, and the exposure time was one second. We used the focus guide provided by Hamamatsu Photonics to adjust the laser focus on every tablet measurement.

Five standard tablets were measured five times on both the front and back sides of the tablet resulting in 50 spectral data being obtained, and the average value was calculated. Each of the genuine and falsified sample tablets was measured five times at different locations on one side, and the average value was calculated. There were only two Diflucan falsified tablets, so each tablet was measured five times on both the front and back side of the tablet. The average was calculated, and the four spectral data of the falsified Diflucan sample were obtained.

Multivariate Analysis

Processing of Raman Spectra

Multivariate analysis was performed using The Unscrambler X 10.5 (CAMO Software; Oslo, Norway). The spectral data were processed with Savitzky-Golay method using a second-order polynomial and a window size of 15.

SIMCA

SIMCA is a commonly supervised classification method based on the PCA model. In PCA, the principal component (PC) is defined as a new variable from the whole data set. The direction with the largest overall variation is the first principal component (PC-1), and the direction with the second-largest variation, which crosses PC-1, is the second principal component (PC-2). In SIMCA, the test samples are compared with the defined PCA model and classified according to their analogy to the class.

The PCA model was constructed for the four products. The 50 spectral data of standard products were used for the PCA model as the calibration set. The optimal number of principal components was determined by cross-validation. The data of samples obtained by personal import or from the manufacturer were treated as the test set.

PLS-DA

PLS-DA is a classification method based on PLS modeling using several different classes. The regression model was created by specifying a group using a dummy variable. The class is represented as the dummy constants “0” or “1,” and the dummy variable for the class is assigned “1” if the sample belonged to the class and “0” if not. The cutoff value for the Y-predicted (Y-pred) plot is 0.5. In other words, the sample is assigned as class “1” if Y-pred >0.5, and the sample is classified as class “0” if Y-pred <0.5.

First, we performed PLS-DA with PLS models using genuine and falsified classes; however, the spectra of falsified samples were too diverse to be treated as one class and an appropriate model could not be created. Therefore, the PLS regression model was created by assigning the four standard products classes as the objective variable, and by assigning Raman spectra as the explanatory variable. The dimensions of the PLS models were determined by cross-validation. The coefficient of determination (R²) and the root mean square error (RMSE) of the predictions were used to evaluate the model. PLS-DA was performed on all test samples for class classification.

RESULTS

Raman Scattering Analysis The Raman spectra of standard products were obtained by using the average of the 50 spectral data. The Raman spectra of the four standard products were different, but the major peak of the spectra was similar among Cialis, Levitra, and Viagra (Fig. 1a). Figure 1b shows the pre-processed Raman spectra of the standard products. The Raman spectra of all the samples were obtained by using the average of five spectral data for each sample. Ten Cialis falsified samples and 11 Levitra falsified samples had no peak, and the spectra appears as noise around the intensity of 50000–60000 because of the fluorescence. Because the spectra of standard products did not show fluorescence, it...
was thought that the fluorescence was due to the composition of the sample, not the laser wavelength or detector. Figure 2b shows the pre-processed Raman spectra of all samples.

**SIMCA** The sample distance to model (Si) represents the distance from the PCA model built using the standard products to the sample. Leverage (Hi) represents how different the sample is from all other samples. The Si vs. Hi plot shows the classification of the sample visually. When both Si and Hi are within the boundary limits, the sample is judged to belong to the model. In this study, sample discrimination was evaluated with a 95% confidence level.

Genuine and falsified Cialis samples were classified by SIMCA using the Cialis standard product model, and the optimal number of principal components was three. The Si vs. Hi
Fig. 3. Si vs. Hi Plot for Cialis
The filled circle represents genuine samples \((n = 9)\), and the open circle represents falsified samples \((n = 23)\). The plots of some falsified samples were overlapped. The red line shows the 95% confidence interval for the standard PCA model. Samples that fall within the limit of both Si and Hi are considered to belong to the model. (Color figure can be accessed in the online version.)

Fig. 4. Si vs. Hi Plot for Levitra
The filled square represents genuine samples \((n = 9)\), and the open square represents falsified samples \((n = 13)\). The plots of some genuine and falsified samples were overlapped. The red line shows the 95% confidence interval for the standard PCA model. Samples that fall within the limit of both Si and Hi are considered to belong to the model. (Color figure can be accessed in the online version.)

Fig. 5. Si vs. Hi Plot for Viagra
The filled triangle represents genuine samples \((n = 4)\), and the open triangle represents falsified samples \((n = 18)\). The red line shows the 95% confidence interval for the standard PCA model. Samples that fall within the limit of both Si and Hi are considered to belong to the model. (Color figure can be accessed in the online version.)

Fig. 6. Si vs. Hi Plot for Diflucan
The filled diamond represents genuine samples \((n = 9)\) and, the open diamond represents falsified samples \((n = 4\), four spectral data were obtained by two Diflucan falsified samples\). The red line shows the 95% confidence interval for the standard PCA model. Samples that fall within the limit of both Si and Hi are considered to belong to the model. (Color figure can be accessed in the online version.)
plot obtained by SIMCA is shown in Fig. 3. The genuine samples were within the limits of both Si and Hi and belonged to the model. All falsified samples were out of the limits of both Si and Hi, and the variation among the falsified samples was large. Genuine and falsified Levitra samples were classified by SIMCA using the Levitra standard product model, and the optimal number of principal components was four. Si vs. Hi plot obtained by SIMCA is shown in Fig. 4. The genuine samples were within the limits of both Si and Hi and belonged to the model. All falsified samples were out of the limits of both Si and Hi. Si was particularly large in one of the falsified samples, which meant this sample was significantly different from the model. Genuine and falsified Viagra samples were classified by SIMCA using the Viagra standard product model, and the optimal number of principal components was four. Si vs. Hi plot obtained by SIMCA is shown in Fig. 5. The genuine samples were within the limits of both Si and Hi and belonged to the model. All falsified samples were within the limits of Hi and outside the limits of Si, and one of which was almost on the boundary line. The other four falsified samples were within the limits of Si and outside the limits of Hi, and one of which was almost on the boundary. Ten falsified samples were out of the limits of both Si and Hi, and the variation among falsified samples was large. The loading plots show that the standard PCA models for Cialis, Levitra, and Viagra were affected by similar wavenumbers (Fig. 7). For all four products, Cialis, Levitra, Viagra, and Diflucan, SIMCA using the standard PCA models showed that all genuine samples belonged to the standard model, and all falsified samples did not belong to the standard model. SIMCA using the standard PCA model could discriminate falsified samples with a rate of 100% at the 95% confidence level. SIMCA was also performed on all samples using the four standard models to confirm whether they belonged to the correct standard model. The Coomans plot shows the distance between the sample and the two PCA class models visually. When the plot of the sample is close to 0, it means that the sample is close to the model. 56) For Levitra, Viagra, and Diflucan, all genuine samples belonged to the correct standard model, and all falsified samples did not belong to any model. For Cialis, all genuine samples belonged to the standard Cialis model, while the two falsified samples belonged to the standard Viagra model. Although only two falsified samples belonged to the standard Viagra model as a result of SIMCA classification using Si and Hi, a Coomans plot showed five Cialis falsified samples were close to the standard Viagra model (p = 0.05) (Fig. 8). Therefore, it was shown that when all samples were classified using the four standard models, some samples belonged to the standard model of another product (Table 1).

**PLS-DA** PLS-DA was performed using the PLS regression models of the four standard products, and the optimal number of factors was five. From the loading plot of the first factor of the PLS, the model was affected by the peak around 510, 600, 630, and 670 cm$^{-1}$ (Fig. 9). The PLS regression model of Cialis showed an $R^2 = 0.944$ and an RMSE = 0.102. All genuine Cialis samples were correctly assigned to the class with a Y-pred of over 0.5, but 12 falsified Cialis samples also belonged to the class. In addition, 11 falsified Levitra samples and two falsified Viagra samples also belonged to the class. All genuine Levitra, Viagra, and Diflucan samples did not belong to the class (Fig. 10). The PLS regression model of Levitra showed an $R^2 = 0.962$ and an RMSE = 0.084. All genuine Levitra samples were correctly assigned to the class with a Y-pred of over 0.5, and all other samples did not belong to the class (Fig. 11). The PLS regression model of Viagra showed an $R^2 = 0.959$ and an RMSE = 0.088. All Viagra samples had a Y-pred of over 0.5, so it was not possible to discriminate falsified Viagra samples. One falsified Cialis, one falsified Levitra and two falsified Diflucan samples were classified by SIMCA using the Viagra standard product model, and the optimal number of principal components was four. Si vs. Hi plot obtained by SIMCA is shown in Fig. 5. The genuine samples were within the limits of both Si and Hi and belonged to the model. The four falsified samples were within the limits of Hi and outside the limits of Si, and one of which was almost on the boundary line. The other four falsified samples were within the limits of Si and outside the limits of Hi, and one of which was almost on the boundary. Ten falsified samples were out of the limits of both Si and Hi. Genuine and falsified Diflucan samples were classified by SIMCA using the Diflucan standard product model, and the optimal number of principal components was four. Si vs. Hi plot obtained by SIMCA is shown in Fig. 6. The genuine samples were within the limits of both Si and Hi and belonged to the model. All falsified samples were out of the limits of both Si and Hi, and the variation among falsified samples was large. The loading plots show that the standard PCA models for Cialis, Levitra, and Viagra were affected by similar wavenumbers (Fig. 7). For all four products, Cialis, Levitra, Viagra, and Diflucan, SIMCA using the standard PCA models showed that all genuine samples belonged to the standard model, and all falsified samples did not belong to the standard model. SIMCA using the standard PCA model could discriminate falsified samples with a rate of 100% at the 95% confidence level. SIMCA was also performed on all samples using the four standard models to confirm whether they belonged to the correct standard model. The Coomans plot shows the distance between the sample and the two PCA class models visually. When the plot of the sample is close to 0, it means that the sample is close to the model. 56) For Levitra, Viagra, and Diflucan, all genuine samples belonged to the correct standard model, and all falsified samples did not belong to any model. For Cialis, all genuine samples belonged to the standard Cialis model, while the two falsified samples belonged to the standard Viagra model. Although only two falsified samples belonged to the standard Viagra model as a result of SIMCA classification using Si and Hi, a Coomans plot showed five Cialis falsified samples were close to the standard Viagra model (p = 0.05) (Fig. 8). Therefore, it was shown that when all samples were classified using the four standard models, some samples belonged to the standard model of another product (Table 1).

**PLS-DA** PLS-DA was performed using the PLS regression models of the four standard products, and the optimal number of factors was five. From the loading plot of the first factor of the PLS, the model was affected by the peak around 510, 600, 630, and 670 cm$^{-1}$ (Fig. 9). The PLS regression model of Cialis showed an $R^2 = 0.944$ and an RMSE = 0.102. All genuine Cialis samples were correctly assigned to the class with a Y-pred of over 0.5, but 12 falsified Cialis samples also belonged to the class. In addition, 11 falsified Levitra samples and two falsified Viagra samples also belonged to the class. All genuine Levitra, Viagra, and Diflucan samples did not belong to the class (Fig. 10). The PLS regression model of Levitra showed an $R^2 = 0.962$ and an RMSE = 0.084. All genuine Levitra samples were correctly assigned to the class with a Y-pred of over 0.5, and all other samples did not belong to the class (Fig. 11). The PLS regression model of Viagra showed an $R^2 = 0.959$ and an RMSE = 0.088. All Viagra samples had a Y-pred of over 0.5, so it was not possible to discriminate falsified Viagra samples. One falsified Cialis, one falsified Levitra and two falsified Diflucan samples were classified by SIMCA using the Viagra standard product model, and the optimal number of principal components was four. Si vs. Hi plot obtained by SIMCA is shown in Fig. 5. The genuine samples were within the limits of both Si and Hi and belonged to the model. The four falsified samples were within the limits of Hi and outside the limits of Si, and one of which was almost on the boundary line. The other four falsified samples were within the limits of Si and outside the limits of Hi, and one of which was almost on the boundary. Ten falsified samples were out of the limits of both Si and Hi. Genuine and falsified Diflucan samples were classified by SIMCA using the Diflucan standard product model, and the optimal number of principal components was four. Si vs. Hi plot obtained by SIMCA is shown in Fig. 6. The genuine samples were within the limits of both Si and Hi and belonged to the model. All falsified samples were out of the limits of both Si and Hi, and the variation among falsified samples was large. The loading plots show that the standard PCA models for Cialis, Levitra, and Viagra were affected by similar wavenumbers (Fig. 7). For all four products, Cialis, Levitra, Viagra, and Diflucan, SIMCA using the standard PCA models showed that all genuine samples belonged to the standard model, and all falsified samples did not belong to the standard model. SIMCA using the standard PCA model could discriminate falsified samples with a rate of 100% at the 95% confidence level. SIMCA was also performed on all samples using the four standard models to confirm whether they belonged to the correct standard model. The Coomans plot shows the distance between the sample and the two PCA class models visually. When the plot of the sample is close to 0, it means that the sample is close to the model. 56) For Levitra, Viagra, and Diflucan, all genuine samples belonged to the correct standard model, and all falsified samples did not belong to any model. For Cialis, all genuine samples belonged to the standard Cialis model, while the two falsified samples belonged to the standard Viagra model. Although only two falsified samples belonged to the standard Viagra model as a result of SIMCA classification using Si and Hi, a Coomans plot showed five Cialis falsified samples were close to the standard Viagra model (p = 0.05) (Fig. 8). Therefore, it was shown that when all samples were classified using the four standard models, some samples belonged to the standard model of another product (Table 1).
flucan samples also belonged to the class. All genuine Cialis, Levitra, and Diflucan did not belong to the class (Fig. 12). The PLS regression model of Diflucan showed an $R^2 = 0.978$ and an RMSE = 0.064. All genuine Diflucan samples were correctly assigned to the class with a Y-pred of over 0.5, and all other samples did not belong to the class (Fig. 13).

In PLS-DA, genuine and falsified samples could be correctly distinguished in Levitra and Diflucan samples. However, some falsified Cialis samples and all the falsified Viagra samples could not be discriminated from the standards. Although some falsified samples belonged to the classes of other products, none of the genuine samples belonged to other product classes. One falsified Levitra sample was excluded because the deviation was greater than 0.5.

Table 1. The Result of SIMCA Using the Four Standard Product Models ($n = 89$)

| Samples        | Cialis genuine ($n = 9$) | Cialis falsified ($n = 23$) | Levitra genuine ($n = 9$) | Levitra falsified ($n = 13$) | Viagra genuine ($n = 4$) | Viagra falsified ($n = 18$) | Diflucan genuine ($n = 9$) | Diflucan falsified ($n = 4^*$) |
|----------------|--------------------------|-----------------------------|--------------------------|----------------------------|--------------------------|----------------------------|-----------------------------|-----------------------------|
| Standard Cialis model | 9                        | 0                           | 0                        | 0                           | 0                        | 0                           | 0                           | 0                           |
| Standard Levitra model  | 0                        | 0                           | 9                        | 0                           | 0                        | 0                           | 0                           | 0                           |
| Standard Viagra model    | 0                        | 2                           | 0                        | 0                           | 4                        | 0                           | 0                           | 0                           |
| Standard Diflucan model   | 0                        | 0                           | 0                        | 0                           | 0                        | 0                           | 9                           | 0                           |

*Four spectral data were obtained from two falsified Diflucan samples.

Fig. 9. PLS Score-Loading Plots

The loading plots on factor-1 are shown. The PLS model was built using standard Cialis ($n = 50$), standard Levitra ($n = 50$), standard Viagra ($n = 50$), and standard Diflucan ($n = 50$).

**DISCUSSION**

Discriminant analysis of falsified medicines was performed using a low-cost Raman spectrometer combined with SIMCA and PLS-DA. A differentiating feature of this work is to establish the application of methods for detecting falsified medicines that are available on the market. Additionally, the applicability of the detection method was evaluated using multiple ED-related medicines actually distributed.

In SIMCA, it was possible to classify the falsified samples completely for all four products. Since all the falsified samples of Cialis, Levitra, and Diflucan were outside the limits of both Si and Hi, it was considered that genuine and falsified samples could be distinguished by SIMCA (Figs. 3, 4, 6). For the Viagra tablets, both Si and Hi were outside the limits in ten falsified samples, and some falsified samples were difficult to judge based on Si or Hi alone (Fig. 5). This might be because the Raman spectra of falsified Viagra samples were similar to that of the standard tablet (Figs. 1, 2). Some falsified Cialis samples were judged to belong to the standard Viagra model (Table 1). A likely factor is that Cialis, Levitra, and Viagra are all film-coated tablets, and the Raman spectra are influenced by additive agents in the coating layer. The standard PCA models of Cialis, Levitra, and Viagra were all affected by peaks of similar wavenumber, suggesting that the three standard PCA models were similar (Fig. 7). The loading plot of Diflucan was different from Cialis, Levitra, and Viagra, and it might be because Diflucan is an uncoated tablet. Only two falsified Cialis samples were classified as belonging to the standard Viagra models in the SIMCA, but the Coomans...
plot suggested that five falsified Cialis samples were close to the standard Viagra model (Fig. 8). The Coomans plot allows visual determination of which of the two PCA models the sample belongs to. It was considered that these falsified Cialis samples were closer to the standard Viagra PCA model than Cialis because of the slight difference in additives. These falsified Cialis samples might not be mistakenly judged as Viagra tablets because the appearance of the tablets resembles a standard Cialis tablet. Since all genuine samples were correctly classified, genuine samples might not be mistakenly judged as falsified or other products.

In PLS-DA, the genuine samples belonged to the correct class in all four products. In Levitra and Diflucan tablets, only the genuine samples of each product belonged to the class, so PLS-DA might be useful method for detecting falsified products (Figs. 11, 13). However, some falsified Cialis samples and all falsified Viagra samples could not be completely distinguished, so it was difficult to classify falsified samples for Cialis and Viagra correctly in PLS-DA. In addition, some falsified samples belonged to another product class (Figs. 10, 12). The PLS model was affected by around 510 and 630 cm\(^{-1}\) from the loading plot (Fig. 9), and Cialis, Levitra, and Viagra all showed major peaks around 510 and 630 cm\(^{-1}\). Because major peaks were similar among Cialis, Levitra and Viagra, some falsified samples could belong to the standard model of another product because of slight differences in additives.
In PLS-DA, falsified samples were sometimes judged to be in the class, but there might be no possibility of misclassifying genuine samples because genuine samples were correctly discriminated.

Although there have been many reports on the classification of medicines by SIMCA and PLS-DA, there are few studies focusing on genuine and falsified products that are actually distributed on the market, and further studies were required to demonstrate practical use in the field. Although PLS-DA is considered to be a useful discriminant analysis, some falsified products showed completely different spectra from others, and a PLS regression model using a single falsified class could not be created. PLS-DA may have success when comparing products of similar composition made by different manufacturers, or using simulated falsified products formulated in the laboratory. However, this study reveals that the method of creating PLS models using falsified products of uncertain origin is not suitable for discrimination. The model could be improved by classifying falsified products into multiple classes by spectral type, however, falsified products are inherently undefined making their classification difficult. In addition, authenticity results are required before discriminant analysis.

Although research using low-cost NIR spectrometry has been reported, there are few studies using low-cost Raman
spectroscopy, and so its use needed to be validated. NIR makes it easy to identify components because of its rich spectral library, but the spectral peaks are broad and are affected by water content and granularity. The spectral peaks of the Raman spectra are sharp, but Raman scattering spectroscopy might not be suitable for samples with weak Raman scattering or strong fluorescence. This study demonstrated the usefulness of a low-cost Raman scattering analyzer coupled with SIMCA. Although the comparison of Raman spectra and PCA score plots provide visual discrimination, SIMCA enables discrimination of falsified products based on a unified standard. Falsified products can be detected using the standard product model in situations where the standard product is available, and does not require authenticity results. Although further demonstration experiments are necessary, it is thought that the establishment of discriminant analysis algorithms will enable the implementation of Raman scattering analysis combined with SIMCA for the detection of falsified products by enhancing the library of standard products in the future. By developing an application to execute SIMCA in real-time, it is possible to identify falsified products rapidly anytime and anywhere. Furthermore, the compactness and low-cost of the analyzer may lead to not only prevent the spread of falsified products in LMICs but also detect falsified products at medical institutions and customs all over the world.

In this study, although each tablet was measured five times, only one tablet was used per sample because of the limited number of tablets. The medicines analyzed in this study were three types of ED therapeutic tablet, and one antifungal tablet advertised as female Viagra. The validation experiments with other types of medicines would lead to social implementation.

CONCLUSION

Raman scattering spectroscopy using a low-cost analyzer combined with SIMCA was useful for detection of falsified medicines actually distributed on the market. SIMCA was more suitable than PLS-DA because some falsified medicines could not be discriminated by PLS-DA.

Acknowledgments This work was supported by SHIBUYA SCIENCE CULTURE AND SPORTS FOUNDATION in FY 2019. Sample collection was supported by Health and Labor Sciences Research Grants from the Ministry of Health, Labour and Welfare, Japan [Grant Nos. H23-chikyukiboshi-006 and H26-chikyukibo A-shitei-003]. We thank Pfizer Inc. (Tokyo, Japan) for donating the standard Viagra tablet, standard Diflucan tablet, and falsified Diflucan tablet. We thank James Murray, PhD for editing a draft of this manuscript.

Conflict of Interest The authors declare no conflict of interest.

REFERENCES

1) WHO. “Substandard and falsified medical products, Fact sheet, 2018.” : https://www.who.int/news-room/fact-sheets/detail/substandard-and-falsified-medical-products, accessed 9 April, 2020.
2) WHO. “Assembly document A70/23, Appendix 3 to Annex, 2017.”: http://apps.who.int/gb/ebwha/pdf_files/WHA70/A70_23-en.pdf, accessed 20 January, 2020.
3) WHO. “Counterfeit medicines, Fact sheet, Revised November 2006.”: https://www.who.int/medicines/definition/en/third_global_congress/third_global_congress_ref_z3.pdf, accessed 6 December, 2020.
4) WHO. “WHO global surveillance and monitoring system for substandard and falsified medical products, reports and executive summary, November 2017.” : https://www.who.int/medicines/regulation/ssffc/publications/GSMReport_EN.pdf?ua=", accessed 10 August, 2020.
5) Sato D. Counterfeit medicines—Japan and the world. Yakugaku Zasshi, 134, 213–222 (2014).
6) Ministry of Health, Labour and Welfare. “Pharmaceutical products to be careful in personal import.”: https://www.mhlw.go.jp/topics/bukyoku/iyaku/kojinyunya/050609-1.html, accessed 28 September, 2020.
7) Ministry of Health, Labour and Welfare. “Unregistered and unlicensed drug purchase survey in 2011.”: https://www.mhlw.go.jp/st/ houdou/2985200002dk5t.html, accessed 28 September, 2020.
8) WHO. “Technology to tackle SF Medical Products.”: https://www.who.int/medicines/regulation/ssffc/technologies/en/, accessed 28 September, 2020.
9) Asia Pacific Economic Cooperation. APEC supply chain integrity toolkit, roadmap for supply chain security.
10) Visser BJ, de Vries SG, Bache EB, Meerveld-Gerrits J, Kroon D, Boersma J, Agranjdi ST, van Vugt M, Grobuesch MP. The diagnostic accuracy of the hand-held Raman spectrometer for the identification of anti-malarial drugs. Malar. J. 15, 160 (2016).
11) Sacré PY, Deconinck E, De Beer T, Courtois S, Vancauwenbergh R, Chiap P, Crommen J, De Beer JO. Comparison and combination of spectroscopic techniques for the detection of counterfeit medicines. J. Pharm. Biomed. Anal., 53, 445–453 (2010).
12) Hajjou M, Qin Y, Bradhy S, Bempong D, Lukulay P. Assessment of the performance of a handheld Raman device for potential use as a screening tool in evaluating medicines quality. J. Pharm. Biomed. Anal., 74, 47–55 (2013).
13) Nagy B, Farkas A, Borbás E, Vass P, Nagy ZK, Marosi G. Raman spectroscopy for process analytical technologies of pharmaceutical secondary manufacturing. AAPS PharmSciTech, 20, l (2019).
14) Ciza PH, Sacré PY, Waffo C, Coic L, Avhou H, Mbinze JK, Ngono R, Marinri RD, Hubert P, Ziemons E. Comparing the qualitative performances of handheld NIR and Raman spectrophotometers for the detection of falsified pharmaceutical products. Talanta, 202, 469–478 (2019).
15) de Veij M, Vandenhoopect, Poemens L. Detection of counterfeit erectile dysfunction drugs with Raman spectroscopy. Spectroscopy Europe, 20, 7–10 (2008).
16) Sakio T, Nagase H, Hakaoka T, Yoshida N, Hirakawa J, Macha S, Hiroshima T, Ikeda Y, Tsuboi H, Kimura K. Survey to identify substandard and falsified tablets in several Asian countries with pharmacopoeial quality control tests and principal component analysis of handheld Raman spectroscopy. Am. J. Trop. Med. Hyg., 98, 1643–1652 (2018).
17) Kakio T, Yoshida N, Macha S, Moriguchi K, Hiroshima T, Ikeda Y, Tsuboi H, Jiruska K. Classification and visualization of physical and chemical properties of falsified medicines with handheld Raman spectroscopy and X-ray computed tomography. Am. J. Trop. Med. Hyg., 97, 684–689 (2017).
18) Mbinze JK, Sacré PY, Yemoa A, Mavar Tayey Mbay J, Habyalimala V, Kalenda N, Hubert P, Marinri RD, Ziemons E. Development, validation and comparison of NIR and Raman methods for the identification and assay of poor-quality oral quinine drops. J Pharm Biomed Anal., 111, 21–27 (2015).
19) Kudelski A. Analytical applications of Raman spectroscopy. Talanta, 76, 1–8 (2008).
20) Roth S, Khetrapal S, Ball D. Portable screening devices for medicine quality: putting power into the hands of regulators in low-
