Identification of authentic and counterfeit Viagra tablets using near-infrared spectroscopic methods and machine learning algorithms

Sarah Rowlands
Faculty of Science
Liverpool John Moores University
Liverpool, United Kingdom
S.Rowlands@ljmu.ac.uk

Prof Dhiya Al-Jumeily OBE
Faculty of Engineering and Technology
Liverpool John Moores University
Liverpool, United Kingdom
D.Aljumeily@ljmu.ac.uk

Dr Sulaf Assi
Faculty of Science
Liverpool John Moores University
Liverpool, United Kingdom
S.Assi@ljmu.ac.uk

***Abstract*— Counterfeit medicinal and lifestyles products are a global issue that impacts public health. Counterfeit products are often made in unsafe and unsanitary conditions before their release to the public without testing by regulatory bodies. One product that is particularly susceptible to online counterfeiting is Viagra, which is one of the highest selling medicines worldwide. A total of 57 Viagra tablets were used for the study; this included 27 authentic and 30 counterfeit tablets which were measured using near-infrared spectroscopy (NIRS). Spectra obtained using the NIR spectrometer non-destructively were exported into a multi-paradigm numerical computing environment where machine learning algorithms (MLAs) were applied using Matlab 2007a. Four algorithms were used related to correlation in wavelength space (CWS), K-nearest neighbour (KNN), principal component analysis (PCA) and PCA combined with fuzzy C-mean clustering (PCA/FCM). The algorithms were applied unsupervised to the authentic and counterfeit tables with no prior labelling to any of the tablets. The results showed two clear groups/clusters between the authentic and counterfeit tablets. In particular, PCA and PCA/FCM showed further subgroups among the counterfeit tablets that corresponded to their varying manufacturing sources. In summary, the use of NIRS and MLAs proved an effective method for identifying counterfeit Viagra medicines rapidly and non-destructively.**

Keywords—Viagra, Authentic, Counterfeit, Near-infrared spectroscopy, Machine learning algorithms, Principal Component Analysis

I. INTRODUCTION

Counterfeit medicinal and lifestyles products are a global issue that impact public health. Counterfeit products are often made in unsafe and unsanitary conditions before their release to the public without testing by regulatory bodies. The dangers of using unregulated products can include untreated illnesses, poisoning, disease progression and death [1].

One product that is particularly susceptible to online counterfeiting is Viagra, which is one of the highest selling medicines worldwide [2]. In 2021, Los Angeles customs agents seized a shipment from China containing one million counterfeit Viagra pills which were worth an estimated $19million in a shipment [3]. A counterfeit product can have defects in its chemical constituents, physical properties, packaging or even labelling.

We chose to analyse the authentic and counterfeit Viagra tablets using near-infrared Spectroscopy (NIRS). The NIR spectrum reflects both chemical and physical parameters and therefore serves as a 'fingerprint', making this technique the method of choice for detecting counterfeit drugs [4].

This study investigates the use of NIRS alongside machine learning algorithms (MLAs) for the identification of authentic and counterfeit Viagra tablets. The authentic tablets were obtained from the manufacturer, and the counterfeit tablets were obtained through the Korean Food and Drug Administration (KFDA), which were collected worldwide.

II. METHOD

A total of 57 Viagra (100 mg) tablets were used for the study, this included 27 authentic tablets obtained from the manufacturer and 30 counterfeit tablets obtained from the KFDA which were collected worldwide.

Some of the counterfeit tablets were clearly differentiated from the authentic ones (no coating or excess dye); while others were identical copies to the authentic tablets. Figure 1 shows examples of two counterfeit tablets one of which is identical imitation. It is noteworthy to highlight the complexity of the tablets that contained nine excipients in addition to the active pharmaceutical ingredient (API) [6].

Thus, the API in Viagra tablets was sildenafil citrate. Moreover, the excipients included:

(1) In tablet core: microcrystalline cellulose, calcium hydrogen phosphate (anhydrous), croscarmellose sodium and magnesium stearate.

(2) In the tablet film coat: hypromellose, titanium dioxide (E171), lactose monohydrate, triacetin, indigo carmine aluminium lake (E132).

The NIR spectra were measured in diffuse reflectance mode using a FOSS 6500 NIR spectrometer equipped with a rapid content analyser (RCA). The RCA works by allowing a monochromatic incident light to pass through its iris into the sample before then being detected by the lead sulfide (PbS) detector.

Performance tests to check the system suitability, absorbance scale and photometric noise were performed prior to measurement to ensure consistent results are obtained. Wavelength accuracy and bandwidth were assessed during the System Suitability Test of the instrument using an internal reference of polystyrene. 10 scans of the polystyrene standard were obtained by transmission measurement and analysed by the software. For wavelength accuracy, a peak finding algorithm was used to determine the peak maxima. For the

Fig 1. Counterfeit Viagra tablets of different coatings.

bandwidth, it is calculated as the ratio of the absorbance at the polystyrene peak at 2167 nm to the valley at 2154 nm after the offset was corrected.

The spectra were measured with reference to a standard spectralon reference and each spectrum was the mean of 32 scans over the wavelength range 1100-2500 nm at 2 nm intervals [5]. From each tablet, four spectra were collected by placing the tablets in direct contact with the RCA. Two spectra were taken from each side such that the tablets such that the tablet was turned at 90 degrees angle between the two spectra.

Data collected was exported into Matlab 2007a where machine learning algorithms (MLAs) were applied.

Spectral pre-treatment made for the NIR spectra was undergone using standard normal variate and second derivatives (SNV-D2).

The SNV approach corrected the offset of the scattered light and particle size of the sample by autoscaling the absorbance values of the NIR spectra. In this respect, the absorbance of each spectrum y at a wavelength i is fitted into the equation:

$$z_i = \frac{(y_i - \bar{y})}{s}$$

Equation 1

Where, zi is the value at the wavelength i

ȳis the value over the full wavelength range

S is the standard deviation of the values over the full wavelength range

In addition, D2 corrected both the offset and baseline encountered in NIR spectra based on Savitzky-Golay algorithm that underlies fitting a second-order polynomial to each spectrum by least square using 13 data points [7].

MLAs applied included correlation in wavelength space (CWS), K-nearest neighbor based on maximum distance in wavelength space, principal component analysis (PCA) and principal component analysis combined with Fuzzy C-mean clustering (PCA-FCM).

CWS:

CWS was based on the momentum product between spectra. It was based on equation 2 between spectra A and B:

$$r_p = \frac{\sum(A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum(A_i - \bar{A})^2 \sum(B_i - \bar{B})^2}}$$

Equation 2

In this respect, an r value of 1 indicated the perfect correlation (identical spectra); whereas, a value of -1 indicated complete dissimilarity. Hence, it was expected that authentic Viagra tablets' spectra will give r values of 1 against each other. However, considering instrumental noise and sample generated noise it was difficult to achieve an r value of 1 in practice. Therefore, a threshold of 0.95 was considered as a pass [6]. Hence, an r value above 0.95 was expected for authentic tablets against each other. In addition, an r value below 0.95 was expected for counterfeit tablets against authentic tablets.

KNN:

KNN was based on the maximum distance in wavelength space that deployed a training set for each sample to calculate the mean sample spectrum and standard deviation. Then each test spectrum was subtracted from the mean and sample spectrum and divided by the standard deviation at each wavelength. The value obtained for KNN was always positive. There was no definite threshold used, but lower thresholds indicated more similar spectra [7].

PCA:

PCA could be used as a dimension reduction and classification algorithm. PCA is based on reducing dimensions of the data into 2D or 3D based on variances among the data. In this respect, the first dimension (PC) explained the highest variance, the second dimension explained the second highest variance not related to the first, the third explained the third highest variance not related to the first and second and so on… Plotting, the first and second PC (PC scores) showed the variances captured among the spectral data and at the same time showed clusters that corresponded to groups within the spectra. The accuracy of PCA would be assessed by the ability to cluster the authentic from the counterfeit Viagra tablets; where clusters were visualised inside the 95% or 99% equal frequency ellipses.

PCA-FCM:

PCA-FCM also assessed the cluster ability of PCA but in a different approach to the one mentioned above. FCM assigned a membership from each point to each cluster. Data points nearer to each assigned cluster indicated closer membership to this cluster and vice versa. Hence, PCA-FCM allowed to visualise further patterns beyond using PCA with equal frequency ellipses.

III. Results and Discussion

NIRS has emerged over the last two decades as an ideal technique for quality control, process monitoring, product authentication and many other applications. The portability of the technique offered an advantage in carrying the laboratory to the sample where it could be used in diverse settings without the need for extensive laboratory procedures. Subsequently, NIRS was used in this study for authentication of Viagra tablets. Four MLAs were explored for authentication of these tablets of which two had been published in a previous study: being CWS and PCA [5]. Consequently, this study complemented the previous study in [5] by evaluating combining PCA with FCM and using KNN for authentication.

It is noteworthy to highlight the diversity of the datasets used in terms of physicochemical characteristics particularly the counterfeit tablets that evidently varied in manufacturing sources, water content, type of coating and constituents. This has not been the case for authentic tablets that were all manufactured by the same manufacturer and had the same market authorisation holder.

On the other hand, the counterfeit samples had variable characteristics in relation to both the physical and chemical properties. Hence, the coating material and colour varied between counterfeit tablets.

They also varied in chemical consituents related to the following groups:

(1) tablets with excess dose of the API,

(2) tablets with underdose of APIs,

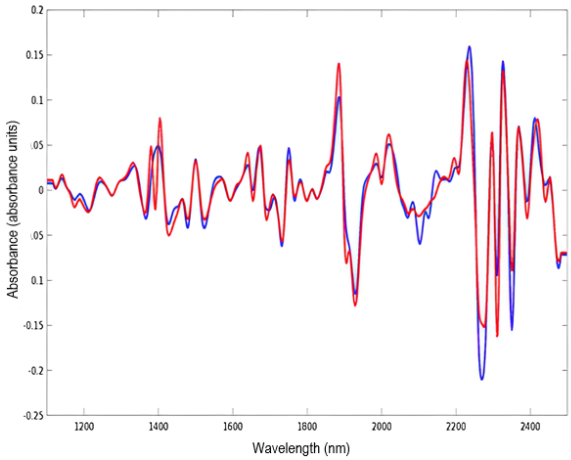

Fig 2. SNV-D2 NIR spectra of authentic (blue) and counterfeit (red) Viagra 100 mg tablets measured by the FOSS 6500 near-infrared spectrometer using a rapid content analyser.

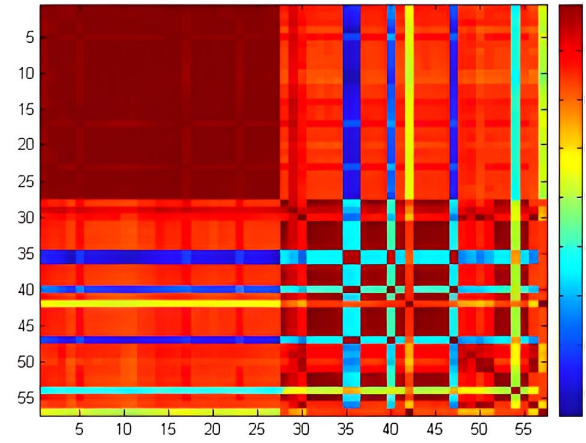

Fig 3. Correlation map of the Standard normal Variate-second Derivative (SNV-D2) spectra of Viagra 100 mg tablets authentic (samples 1 to 27) and counterfeit (samples 28 to 57) measured by the FOSS 6500 NIR spectrometer equipped with a rapid content analyser.

(3) tablets with talc as excipients,

(4) tablets with only 3-4 excipients.

It is noteworthy to mention that reference analysis of the API content has been made using high performance liquid chromatography.

Bearing in mind the physicochemical differences between tablets, NIRS is the ideal technique for their characterisation. The weak signal of NIRS makes the technique ideal for characterising physical properties non-destructively [6]; where the tablets could be measured 'as received' with no sample crushing or using solvents. However, one major drawback is related to the weak NIRS signal when investigating chemical properties. This drawback underlies the overlapping NIR spectra where bands correspond to multiple ingredients within a sample (Figure 2).

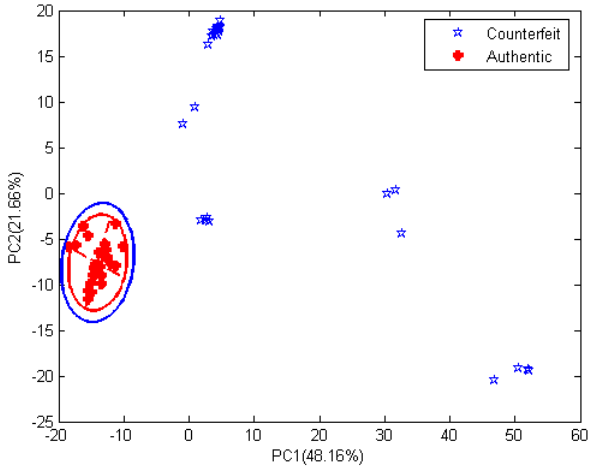

Fig. 4 Principal Component Analysis scores plot of the Standard normal Variate-second Derivative (SNV-D2) spectra of Viagra 100 mg tablets authentic (red crosses) and counterfeit (blue stars) with the 95% (red) and 99% (blue) confidence ellipses drawn around the authentic spectra measured by Near-infrared spectroscopy using a rapid content analyser.

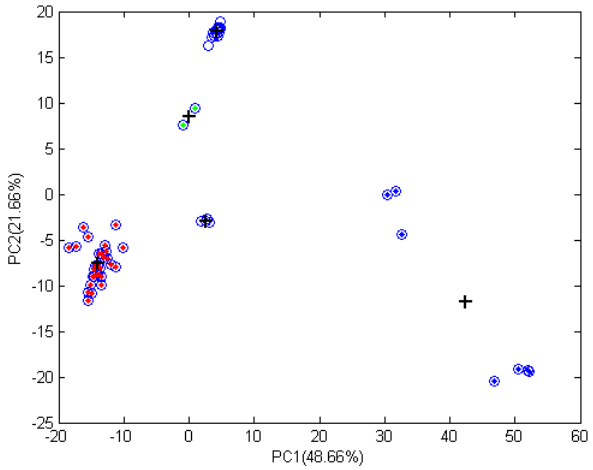

Fig. 5 Principal Component Analysis-FCM clustering classification of the Standard normal Variate-second Derivative (SNV-D2) spectra of Viagra 100 mg tablets authentic (red) and counterfeit (blue) clustered around three centres measured by Near-infrared spectroscopy using a rapid content analyser.

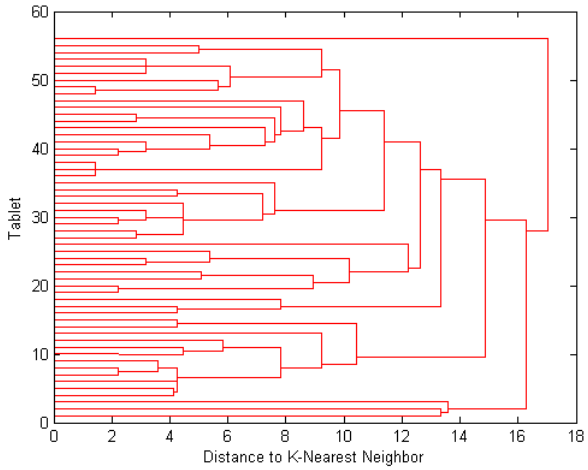

Fig. 6 Cluster tree of authentic (Tablets 1 to 27) and counterfeit (Tablets 28 to 57) Viagra tablets.

Considering the spectral complexity and the overlapping spectra, MLAs are needed in order discriminate between authentic and counterfeit medicines. Thus, four MLAs were applied in this work in an unsupervised approach in order to explore patterns between authentic and counterfeit tablets. Overall, the MLAs used showed to be complementary in medicines' authentication and tracking manufacturing sources. However, each method showed drawbacks alongside the advantages it offered.

In this respect, the first MLA, CWS, showed to be quick, did not require numerous spectra and accurate for identifying authentic tablets that all showed r values > 0.98 (Figure 3). However, few mismatches were seen when CWS was applied to spectra of counterfeit tablets. Hence, few counterfeit tablets that contained the API and common excipients in authnetic tablets gave matches > 0.95 against the authentic tablets. Hence, Figure 3 shows consistent colour between the r values of authentic tablets (dark red colour) corresponding to r values in between 0.98 and 1. On the other hand, counterfeits had range of colours including dark red, light red, orange, yellow, green, light and dark blue. This further to confirmed to the consistency of manufacturing of authentic tablets and poor quality manufacturing of counterfeit tablets. Yet, the false positives encountered for the counterfeit tablets urged the need for using alternative MLAs for analysis.

Subsequently, PCA was applied to the SNV-D2 spectra of the authentic and counterfeit tablets. It is worthnoting that the PCA model was applied to the full spectral range (700 datapoints per spectrum). In this respect, the first two PCs contributed to 69.8% of the variance among the spectra (Figure 4). When the PC scores were plotted, clear grouping were seen between the authentic and counterfeit tablets. The authentic tablets were grouped together that related to their unified manufacturing source despite different purchase sources. There was no inaccuracies in the clustering where all authentic tablet scores fitted within the 95 and 99% equal frequency ellipses. On the other hand, the counterfeit tablets showed five clusters that were clearly distinct from each other. This could be attributed to either manufacturing sources or common excipients within the tablets. However, the sample size was not sufficient to construct equal frequency ellipses around the counterfeit tablets.Therefore, PCA combined with FCM as applied to the authentic and counterfeit tablets (Figure 5). FCM showed five cluster centres corresponding to the different grouping of the authentic and counterfeit tablets. In this respect, the distance of each score to the cluster centre related to its membership where the nearest the score, the highest the membership and vice versa. Five cluster centres were seen (Figure 5): one for the authentic and four for the counterfeit tablets. This could be attributed to the physicochemical similarities between the groups that in turn relate to constituents and manufacturing sources. Yet it was not possible to identify the exact contribute to the individual groups.

Subsequently, KNN was applied to the authentic and counterfeit tablets' spectra. Figure 6 shows the cluster tree of the authentic and counterfeit tablets' spectra calculated based on MDWS. In this sense, distances were variable between the authentic tablets as well as between the counterfeit tablets individually. For instance, the first three tablets showed distance (up to 12) between them and from other tablets. In total, five linked groups were obtained within the cluster tree. Nonethless, these links were ntot related to the authentic and counterfeit sets. This demonstrated the unsuitability of KNN for authenticating counterfeit Viagra tablets; whereas, correlation, PCA and PCA-FCM were more suitable.

IV. Conclusion

This work demonstrated a rapid and non-destructive method for authenticating Viagra tablets rapidly and non-destructively. NIRS offered advantages over other spectroscopic technique in its weak signal that made it ideal for characterising physicochemical properties of the tablets. The technique was non-destructive and allowed to measure the tablets as received with no need for crushing them or changing their morphology. Hence, the method was:

- Simple (did not require sophisticated analysis).
- Green (no need for harmful laboratory solvents).
- Sustainable (samples could be measured several times).
- Cost-effective (no expensive kit required).

Nevertheless, this weak signal attributed to the overlapping spectra of the authentic and counterfeit tablets that necessitates the need for MLAs. MLAs applied to the spectra of authentic and counterfeit tablets showed to be complementary for tablets'authentication. Only KNN showed to be unsuitable for classifying the measured sets of tablets. However, CWS, PCA and PCA-FCM were successful in classifying tablets and tracing manufacturing sources. CWS required less spectral collection as it compared the r value between each two individual spectra based on their momentum product. On the other hand, more spectra were needed for PCA and PCA-FCM in order to ensure accuracy and sufficient variance captured. The latter two MLAs showed rich information corresponding to characteristics within the samples. However, they still required large sample sizes and that could be critical especially in real life scenarios few medicines could be available. Consequently, future work entails applying algorithms that can learn from missing data and incomplete data. Moreover, a complementary spectroscopic technique to NIR could be advantageous. Such techniques include Raman spectroscopy that is specific to APIs and is worth considering in future work.

ACKNOWLEDGMENTS

The authors thank colleagues at UCL School of Pharmacy (UK), Korean FDA and FOSS for facilitating the samples and instrumentation that had been used to carry out the chemical analysis for the study.

REFERENCES

[1] World Health Organisation. (2018). Substandard and falsified medical products. [Online]. World Health Organisation. Last Updated: 31st January 2018. Available at: https://www.who.int/news-room/fact-sheets/detail/substandard-and-falsified-medical-products [Accessed 10 November 2022].

[2] Grey, E. (2016). Tackling counterfeit Viagra. [Online]. Pharmaceutical Technology. Last Updated: 7th August 2016. Available at: https://www.pharmaceutical-technology.com/analysis/featuretackling-counterfeit-viagra-4970635/ [Accessed 11 October 2022].

[3] Spinner, J. (2021). LA customs agents seize $19m in counterfeit Viagra pills. [Online]. Outsourcing Pharma. Last Updated: 13th January 2021. Available at: https://www.outsourcing-pharma.com/Article/2021/01/13/LA-customs-agents-seize-19m-in-counterfeit-Viagra-pills [Accessed 13 October 2022].

[4] Vredenbregt, M.J., Blok-Tip, L., Hoogerbrugge, R., Barends, D.M., de Kaste, D. (2006). Screening suspected counterfeit Viagra® and imitations of Viagra® with near-infrared spectroscopy. Journal of Pharmaceutical and Biomedical Analysis. 40(4), pp.840-849. [Online]. Available at: https://www.sciencedirect.com/science/article/pii/S0731708505005625#bib4 [Accessed 14 October 2022].

[5] Assi, S., Watt, R.A., Moffat, A.C. (2011). Identification of counterfeit medicines from the internet and the world market using near-infrared spectroscopy. Analytical methods. 10(3), pp.2231-2236. [Online]. Available at: https://pubs.rsc.org/en/content/articlelanding/2011/ay/c1ay05227f [Accessed 14 October 2022].

[6] EMC. (2021). Viagra 100 mg film-coated tablets. [Online]. Medicines EMC. Last Updated: January 2021. Available at: https://www.medicines.org.uk/emc/product/7978/smpc#EXCIPIENTS [Accessed 18 October 2022].

[7] Jee, R.D. Near-infrared spectroscopy. In: Moffat, A.C, Osselton, M.D. and Elliot, S. Clark's Analysis of Drugs and Poisons (2011). London: Pharmaceutical Press.