Science & Justice 62 (2022) 327–335

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Research Paper

Near-infrared spectroscopy combined with chemometrics to classify cosmetic foundations from a crime scene

Svetlana Skobeeva, Alana Banyard, Brian Rooney, Ravtej Thatti, Baljit Thatti, John Fletcher *

School of Life Sciences, Pharmacy and Chemistry, Faculty of Science, Engineering and Computing, Kingston University, Penrhyn Road, Kingston upon Thames KT1 2EE, UK

ARTICLE INFO

Keywords:
Cosmetic
Foundation
Trace evidence
Chemometrics
Near-infrared
PCA-LDA

ABSTRACT

Cosmetic smears are a form of trace evidence that can link the crime scene, suspects, and victims. Foundation and lipstick are the most common sources of cosmetics that can easily smear, with most current research focused on the evidential analysis of lipsticks. This research aims to create a database of cosmetic foundations on different materials and to access the robustness of using Near-infrared with chemometrics as a non-destructive technique to identify unknown samples collected from a crime scene. Small amounts of six shades of three brands of foundations were smeared on clothing materials, which were then analysed with a combination of Near-infrared with chemometric analysis. Principle component analysis (PCA) was used to reduce data dimensionality and explore potential patterns in sample separation and Linear Discriminant Analysis (LDA) was utilised to assign unknown samples to one of the established classes. The selected techniques proved to be promising for database construction and as a preliminary method of analysis, with 93% of the spectra being correctly classified. Notably, darker foundation shades were less likely to be correctly classified (90% classified correctly) compared to lighter ones (96.7% classified correctly). This could not be improved with Standard Normal Variate (SNV) data pre-treatment or selecting specific NIR regions. This finding is of particular importance; according to the Crime Survey for England and Wales (year ending March 2020) police recorded sexual offences demonstrated that those in Mixed and Black or Black British ethnic groups were significantly more likely to be a victim of sexual assault compared to White, Asian or Other ethnic groups. It is, therefore, crucial to add a wide range of foundation shades, particularly of darker tones, to the future database.

1. Introduction

Trace evidence is small amounts of recoverable material, residue, impressions or marks that are used in forensic case work [1]. The analysis of these forms of evidence varies depending on the type of trace and any additional testing required of the specimen [2]. It is not uncommon for a multiplexed analytical approach to be used for the most discriminative and accurate results in identification of the trace evidence [3]. Non-destructive techniques are preferred to preserve the evidence in its original form and amount, especially in cases when the investigation or analytical strategy needs to be repeated or revised at a future date [3]. A comparison between a questioned sample and reference sample, for instance through chemical analysis, will reveal the possibility of a common origin [4] and allows links to be made between the crime scene, suspect or victim or all the above [5].

Cosmetic products are substances intended for external use on the body, to perfume, beautify or nurture [6], any of which can produce a smear when physical contact occurs with another surface. This smear can be collected as trace evidence from a crime scene. The most abundant expression of cosmetic smears is produced from lipsticks and foundation due to their highly transferable nature [7]. Foundation is applied at relatively large amounts onto the face and therefore has a greater chance of coming into contact with another surface.

The composition of cosmetic foundation can be split into bases, fillers, and pigments [6,8]. Base aids smooth application of the product onto the skin [6], filler prevents shine, absorbs oil, and helps even distribution of colour pigments across the face [8]. A mix of pigments is used to achieve a certain shade, even out the skin tone and reduce the appearance of blemishes [8]. These pigments also add an undertone to the foundation, that are described as cold, neutral or warm [9–11].

* Corresponding author.
E-mail addresses: k1814727@kingston.ac.uk (S. Skobeeva), k1612535@kingston.ac.uk (A. Banyard), b.rooney@kingston.ac.uk (B. Rooney), r.thatti@kingston.ac.uk (R. Thatti), b.thatti@kingston.ac.uk (B. Thatti), j.fletcher@kingston.ac.uk (J. Fletcher).

https://doi.org/10.1016/j.scijus.2022.03.002
Received 29 September 2021; Received in revised form 23 February 2022; Accepted 6 March 2022
Available online 8 March 2022

The three brands, selected for the development of a spectroscopic database of cosmetics to identify crime scene samples, use water as a base, dimethicone as a filler, and titanium dioxide and iron oxide as pigments. As well as the foundations individually, the NIR may show interferences from the different clothing materials used as a matrix. then, spectral subtraction may be necessary to get clearer results [12,13].

Cases that will significantly benefit from progression in cosmetic evidence analysis are contact crimes, such as sexual assault. 58,845 sexual assault cases were recorded by the police in England and Wales in 2020, all of which could involve a transfer of cosmetic evidence [14]. Traces of cosmetics may also be discovered on commonplace items, such as drinking cups, tissues, face masks etc. Currently, however, there are limited robust methods of analysis for cosmetic evidence, and no guidelines or standards set [15], despite the importance of cosmetic evidence being established in 1912 [16].

Near-infrared spectroscopy (NIR) is a non-destructive analytical technique that provides insight into matrices with complex composition [17,18], such as pharmaceuticals [18,19] and makeup [15,20]. NIR includes the range between the visible spectral range and the mid-infrared region [21]. Prominence of absorption bands in the NIR region (wavelength range of 780–2526 nm or wave number range 12820–3959 cm^{-1}) is a result of overtones and combinations of fundamental vibrations of –CH, –NH, –OH (and –SH) functional groups [17,18].

Mid-IR absorptions bands directly correspond to specific chemical bonds [22], whereas NIR absorption bands are typically overlapping, broader and much weaker than mid-infrared ones [17,18]. This aspect of NIR complicates spectra interpretation since the untreated spectra cannot be interpreted without additional statistical analysis, but it can also be advantageous both in transmittance and reflectance mode. The low absorption coefficient allows the light to penetrate deeper into the sample, thus permitting direct analysis of samples with strong absorbing or high scattering properties without additional sample pre-treatment [17,18].

One of the major benefits of utilising the NIR spectra rather than the IR spectra is that NIR can differentiate between samples of the same constituents since the overtones and combinations will differ. Dubois and co-workers [18] found this aspect of NIR particularly useful in identification of counterfeit antimalarial tablets. A combination of NIR imaging and Principal Component Analysis (PCA) identified subtle differences within the pharmaceuticals [18]. Pharmaceutical research focusing on the identification of counterfeit drugs shows that NIR spectroscopy is capable of detecting even slight changes in composition, including differences in particle size. Research in identifying counterfeit Viagra showed a success rate of 98% for correctly identifying the presence or absence of a key ingredient [19].

Due to complexity of the raw NIR spectra, statistical data analysis, or chemometrics, is often utilised to interpret obtained results. Chemometrics is the extraction of information from chemical systems using data statistics [21]. It is necessary to use chemometrics in this instance as the similar chemical composition of the foundations makes spectral analysis alone challenging [20,22–23]. PCA is an unsupervised multivariate analysis technique, which reduces variance and assigns patterns and groupings to the data [24], thus allowing to explore whether there are compositional or physical differences between the studied samples. It does this by creating new principal components whose linear combination will approximate the original data [25]. The first principal component (PC1) will always account for the greatest variability in the data, with subsequent PCs accounting for ever decreasing amounts [26].

Linear Discriminant Analysis (LDA) [19,27] is a supervised pattern recognition method that uses "training objects" with known class membership to establish a rule for grouping new unknown objects (samples) into a correct class. Linear Discriminant Function enables differentiation between different classes of samples by maximising distinguishing features between classes and minimising scatter within each class. A disadvantage [28] of LDA usually noted by other authors is that the number of samples must exceed the number of variables. This can be overcome by performing PCA on the data first [29,30], thus reducing data dimensionality whilst retaining as much variance as possible. Afterwards, LDA is applied to the PC scores matrix to construct prediction and classification LDA models. Distance from the centre of each sample to the centre of each modelled class is calculated and the samples are assigned to the class with the lowest distance value.

Despite potential challenges that may be encountered in accurate discrimination of foundations due to similar constituents, there are numerous successful examples of NIR being used under similar circumstances in various fields such as pharmaceuticals [18,19], identification of documents [25] and cosmetics [31].

In this work, the NIR combined with PCA and LDA were explored for rapid and accurate identification of cosmetic samples. Liquid foundations in a range of shades from three different brands were smeared on four different materials and prepared samples represented cosmetic evidence that could be found at a crime scene. The combination of NIR, PCA, and LDA was examined in its capability to create a spectroscopic database of a variety of cosmetic samples. Once a robust classification model is developed, with specificity, sensitivity, and accuracy above 90%, it will enable unknown samples, i.e., from a crime scene/victim/suspect, to be identified via a database match.

2. Materials and methods

2.1. Materials and sample preparation

Three high street brands of liquid foundation cosmetics were selected, NYX, e.l.f., and L'Oréal Paris. Six shades from each brand were included, ranging from lightest to darkest shades available within each brand (Table 1).

The four materials chosen for the database reflect common materials found in typical clothing: a blue top (95% cotton and 5% elastane), blue jeans (100% cotton denim), a white top (100% polyester) and a blue jumper (100% acrylic). Each material was cut into squares, approximately 5 cm × 5 cm, one square of each material for each shade of foundation. A small amount (500 mg – 1 g) of each foundation was smeared once onto each type of material using a spatula. The individual samples were then packaged in small tamper-proof evidence bags using tweezers and fully labelled with the date, analyst's signature and the foundation and material specification. They were then sealed until the analysis was carried out. PPE and basic decontamination procedure were maintained throughout sample preparation. In total, 76 samples of 16 foundation shades, each smeared on four different clothing materials, were prepared for analysis, including the four blank material controls.

2.2. Near infra-red spectroscopy (NIR)

The Buchi (Postfach, Switzerland) NIR – Flex N500 model of the spectrometer was used, in conjunction with the NIRWare and NIRCal software. Each sample was taken out of the sealed evidence bag and placed on a clean bench, where spectra were collected using a NIRFlex fibre optic probe, and each spectrum was constructed from the average of 16 scans. The probe was placed over the smear so that the tip was perpendicular to the sample. The smear was divided into six equal sections and 5 spectra were collected from each, 30 spectra from one smear in total. Calibration (system suitability and background collection) was performed as required by the system. The spectra were then exported as Excel spreadsheets for further analysis. This process was repeated for each smear of each foundation brand, shade and clothing material combination.

Table 1
List of foundations used in the analysis. Within each brand the shades are in the order of from lightest (first brand row) to darkest (last brand row). Lighter foundation shades are samples 1–3, 7–9, 13–15 and darker foundation shades are samples 4–6, 10–12, 16–18.

Sample Number	Code	Brand	Range	Shade Name	Batch Number
1	NYXV	NYX Professional Makeup	Born to Glow	Vanilla	
2	NYXWV	NYX Professional Makeup	Born to Glow	Warm Vanilla	MF T40W
3	NYXG	NYX Professional Makeup	Born to Glow	Golden	MF S50W
4	NYXG	NYX Professional Makeup	Born to Glow	Mocha	MF S50W
5	NYXC	NYX Professional Makeup	Born to Glow	Cocoa	MF S50W
6	NYXDW	NYX Professional Makeup	Born to Glow	Deep Walnut	MF S50W
7	ENA	e.l.f	Flawless finish	Natural	
8	ELI	e.l.f	Flawless finish	Light ivory	
9	ENU	e.l.f	Flawless finish	Nude	
10	ECO	e.l.f	Flawless finish	Coco	
11	ENUT	e.l.f	Flawless finish	Nutmeg	
12	EBO	e.l.f	Flawless finish	Ebony	
13	LRS	L'Oréal Paris	True Match	Rose Sand	16 T 8 01
14	LCB	L'Oréal Paris	True Match	Creamy Beige	16 T 9 01
15	LGS	L'Oréal Paris	True Match	Golden Sun	16 T 7 00
16	LDG	L'Oréal Paris	True Match	Deep Golden	
17	LES	L'Oréal Paris	True Match	Espresso	16 N 2 02
18	LEB	L'Oréal Paris	True Match	Ebony	16S 3 03

2.3. Chemometrics

2.3.1. Data pre-treatment

All data analysis was carried out using the Unscrambler X software version 10.3. The data was imported into the software and transposed. Application of pre-treatments was explored, such as Standard Normal Variate (SNV) [32] and selection of specific regions to examine whether correct classification rates could be improved.

2.3.2. Exploratory data analysis and classification of the samples for the database

PCA was used for unsupervised data analysis to explore any potential pattern in the separation of the samples of different shades and/or brands and to reduce data dimensionality for subsequent LDA. As mentioned at the end of section 1.3.2, LDA does not work efficiently on a data set that has more variables than samples. By applying PCA, 1502 variables, representing wavenumbers from 4000 to 10,000 cm^{-1}, were transformed into seven variables representing scores for the first seven PCs. While all data analysis was carried out in the Unscrambler X, Microsoft Excel was used to plot PCA scores plots. When building LDA prediction and classification models, Mahalanobis method was used and the data from the 76 prepared samples was split into training and testing sets, 75% and 25% of the whole dataset respectively. Every fourth spectrum, starting from the first, was included in the testing set (540 spectra) and the rest were put in the training set (1620 spectra). The training set was used to train the LDA models and define categories that would be used in assigning classes to unknown samples from the testing set. LDA prediction and classification results can be displayed in a confusion matrix. Confusion matrix (Table 5) is a table recording correct and incorrect classifications, where each row represents the predicted class assigned to a sample and each column corresponds to the actual class of the samples. Model success was assessed by calculating performance evaluation measures [33] (Table 2), such as specificity (True Negative Rate), sensitivity (True Positive Rate), accuracy, and False Positive Rate (FPR). To minimise bias in PCA modelling, when selecting samples for the training/testing sets, selection intervals were started at different positions e.g. every fourth spectrum, starting from the second, third, and fourth sample. In addition, all classes have equal number of samples to ensure uniform representation of classes.

3. Results and discussion

3.1. NIR spectra and data Pre-treatment

Recorded NIR spectra (Fig. 1) illustrate notable differences between lighter and darker foundation shades. Peaks in the spectra of lighter shades (Table 1, samples 1–3, 7–9, 13–15) tended to be sharper when compared to those of the darker shades (Table 1, samples 4–6, 10–12, 16–18). This could be explained by the optical effect of black pigments (Black Iron Oxide Ci 77499), which is due to non-selective light absorption [34]. Since darker shades are likely to contain more black iron oxide pigment compared to lighter shades, this would cause the samples to absorb more light resulting in less intense peaks. Optical effect of white pigments, such as TiO_2, is due to non-selective light scattering [35] and TiO_2 also reflects the NIR portion of the solar spectrum [36], resulting in more prominent peaks. NIR light reflectance additionally depends on how the pigments interact with other ingredients and particle size [37]. Due to the nature on NIR the original spectra are very difficult to analyse, therefore, additional data analysis methods, such as chemometrics, are applied.

The authors experimented with applying SNV pre-treatment and focusing on specific regions to potentially improve classification model performance. Whilst SNV improved the NIR spectra visually, it did not have a positive effect on LDA classification results and in some cases had reduced sensitivity (Table 3). Specificity and accuracy remained high across all classes (above 96.5%) and sensitivity was even improved for some foundation shades, such as ENU and NYX G (from 93.3% to 100%). At the same time, with application of SNV, sensitivity has significantly decreased for darker shades, such as EBO (from 96.7% to 86.7%) and LEB (from 83.3% to 56.7%).

Typically, SNV pre-treatment removes the multiplicative interferences of scatter and particle size [32,38], which reduces natural physical variance [38] thus potentially improving prediction and classification models. In this work, however, we did not observe positive results and in some cases correct classification rates were compromised

Table 2
Performance evaluation measures and their respective formulas[33]. TP – True Positive, FP – False Positive, TN – True Negative, FN – False Positive.

Performance Evaluation Measure	Formula
Specificity (True Negative Rate)	$\frac{TN}{TN+FP}$
Sensitivity (True Positive Rate)	$\frac{TP}{TP+FN}$
Accuracy	$\frac{TP+TN}{TP+FP+TN+FN}$
False Positive Rate	$\frac{FP}{FP+TN}$

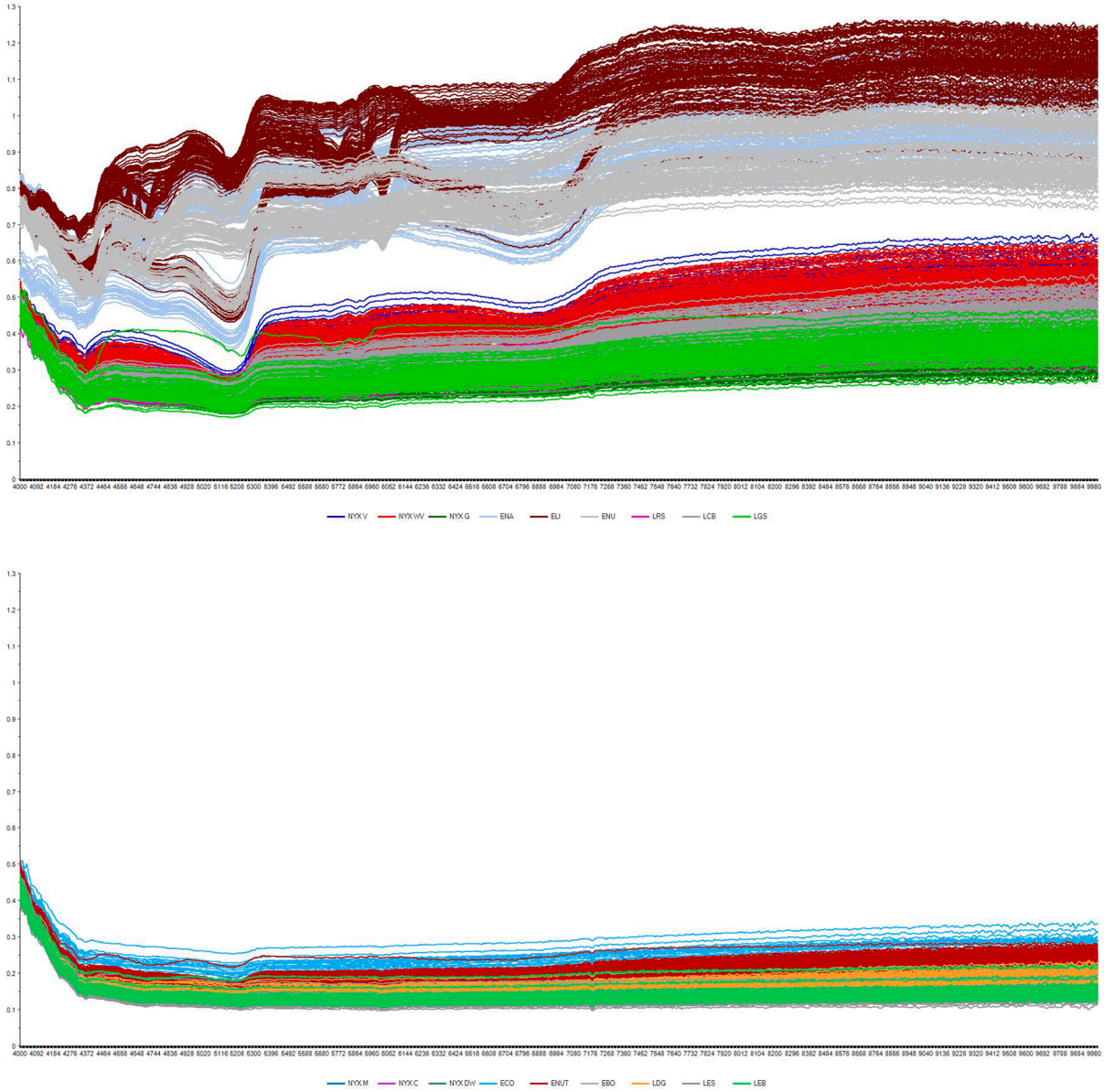

Fig. 1. NIR spectra of lighter (top) shades (Table 1, samples 1–3, 7–9, 13–15) and darker (bottom) shades (Table 1, samples 4–6, 10–12, 16–18) without any pre-treatments. Lighter shades, particularly e.l.f. samples appear to have more prominent peaks compared to darker shades.

by applying the pre-treatments mentioned above.

When focusing on selected wavenumber ranges with most prominent peaks, sensitivity was improved for darker shades but was significantly reduced for lighter shades, such as NYXV and LRS (data not shown). Due to this, a complete dataset with no data pre-treatments or changes to the wavenumber range was used when constructing prediction and classification models.

3.2. Exploratory data analysis

The PCA involved the testing of groups of samples. The greater the separation between the data clouds, the greater the likelihood that the models could distinguish between them. As per the method, the Unscrambler X was permitted to use up to seven PCs to create each model. Two separate PCA scores plots were plotted for two groups: lighter (Table 1, samples 1–3, 7–9, 13–15) and darker shades (Table 1, samples 4–6, 10–12, 16–18) to examine the separation of foundation samples of similar shade.

As seen in Fig. 2.1, all three e.l.f. foundation shades are clearly separated from the other two brands along PC1, which accounts for 99% of variation. Upon zooming in (Fig. 2.2) two overlapping clusters can be distinguished, where the lightest shades NYX V and NYX WV cluster together (Fig. 2.2, cluster A) and somewhat overlap with a slightly darker shade, LCB, from a different brand. The rest of the shades in this plot also cluster together (Fig. 2.2, cluster B) and are separated from the two NYX samples, NYXV and NYXWV, along PC1 and PC2 (1% of variation).

On the PCA scores plot for darker shades (Fig. 3), e.l.f. shades are separated from the other two brands along PC1 (99% variation), similar to Fig. 2.1, except for EBO samples. ECO and ENUT samples (Fig. 3 cluster B and A respectively) were also separated from each other along PC2 with some overlap between the two clusters. EBO samples clustered together with other darkest shades (Fig. 3, cluster D) away from samples from the other two brands, LES, LEB, and NYX DW. NYX M, NYX C, and LDG samples clustered together (Fig. 3, cluster C) away from the other two data clouds, but somewhat overlapping with the darkest shades (Fig. 3. cluster D). PC1 appears to contribute to cluster separation the most.

Table 3
LDA classification table comparing performance evaluation measures between two models. One based on dataset with no pre-treatment (raw) and the other based on dataset pre-treated with SNV (SNV).

Data pre-treatment	raw	SNV	raw	SNV	raw	SNV	raw	SNV
Performance evaluation measures (%)	**Specificity (TNR)**		**Sensitivity (TPR)**		**Accuracy**		**False Positive Rate**	
NYX V	99.4	99.6	83.3	80.0	98.5	98.5	0.6	0.4
NYX WV	99.0	98.8	90.0	90.0	98.5	98.3	1.0	1.2
NYX G	100.0	99.8	93.3	100.0	99.6	99.8	0.0	0.2
NYX M	99.4	100.0	90.0	96.7	98.9	99.8	0.6	0.0
NYX C	99.2	99.8	90.0	90.0	98.7	99.3	0.8	0.2
NYX DW	99.8	99.4	83.3	83.3	98.9	98.5	0.2	0.6
ENA	99.6	100.0	100.0	100.0	99.6	100.0	0.4	0.0
ELI	100.0	100.0	100.0	100.0	100.0	100.0	0.0	0.0
ENU	100.0	100.0	93.3	100.0	99.6	100.0	0.0	0.0
ECO	100.0	100.0	100.0	93.3	100.0	99.6	0.0	0.0
ENUT	99.6	99.4	100.0	100.0	99.6	99.4	0.4	0.6
EBO	99.8	98.2	96.7	86.7	99.6	97.6	0.2	1.8
LRS	100.0	100.0	96.7	96.7	99.8	99.8	0.0	0.0
LCB	100.0	99.8	100.0	100.0	100.0	99.8	0.0	0.2
LGS	99.6	100.0	100.0	100.0	99.6	100.0	0.4	0.0
LDG	99.8	99.6	96.7	96.7	99.6	99.4	0.2	0.4
LES	99.2	97.8	76.7	80.0	98.0	96.9	0.8	2.2
LEB	98.2	98.8	83.3	56.7	97.4	96.5	1.8	1.2
average	99.6	99.5	93.0	91.7	99.2	99.1	0.4	0.5
median	99.7	99.8	95.0	96.7	99.6	99.5	0.3	0.2

When all shades were plotted on a PCA scores plot (data not shown), the two groups (lighter and darker foundation shades, Table 1) were separated along PC1 (99% variance). Lighter shades were more positively correlated with PC1 than darker shades and PC2 did not seem to contribute to sample separation in a significant way. Upon zooming in on each group similar patterns of separation were observed, as described above in Fig. 2 and Fig. 3.

From these results it was expected that the darkest foundation shades (NYX DW, EBO, LES, LEB) were likely to be misclassified due to their overlap on the PCA scores plot. Similarly, since NYX V and NYX WV samples clustered together in Fig. 2.2 it was also predicted that these samples were likely to be misclassified as each other.

The PCA models for the overall database showed the most variance from principal components 1 and 2 and exhibited satisfactory separation in PC-1. The foundations are all reasonably similar in chemical composition, as described in section 1.2. Distinguishing between the samples, therefore, is expectedly more difficult. Notably, sunscreen also contains components similar to those found in cosmetic foundations, such as TiO_2, silica, and talc [13,20,23]. A study on investigation of sunscreens [23] has also observed close proximity of the samples in PCA scores plot despite significantly different PC values between the samples. Such proximity of samples on the scores plot was attributed to the large sample size (n = 109) and most samples containing similar base constituents.

Multiple factors may be contributing to separation of the liquid foundation samples on the PCA scores plot. e.l.f. foundations contain BN, magnesium stearate, magnesium palmitate, and magnesim sulfate, unlike the other two brands. Positioning of lighter e.l.f. shades (Fig. 20.1) and darker shades (Fig. 3) may suggest that the presence of aforementioned chemicals separates e.l.f. samples from the rest. Further separation of e.l.f. samples may be attributed to slightly greater amount of cyclohexasiloxane in ELI and ENU compared to ENA samples. Since NYX V and NYX WV are clustered away from NYX G and L'Oréal Paris samples, this separation maybe due to higher alcohol content in L'Oréal Paris samples compared to NYX and e.l.f. samples. l'Oréal Paris and NYX G samples further along the PC2 axis than NYX V and NYX WV, which may be due to higher content of disteardimonium hectorite or glycerine. NYX samples also contain significantly more TiO_2, which could have influenced the clustering of these samples. Study by Sharma and co-workers [20] showed similar results, where some foundation samples clustered according to amount of certain ingredients, such as alcohols and amides, silica and dimethicone.

A study performed by Wong, Sauzier, and Lewis [39] noted that degree of distinction of lipstick sample clusters on a PCA scores plot (PC1-3) partially depended on lipstick shade and degree of colour variation within individual colour groups [39]. Chophi, Sharma, and Singh [40] demonstrated that PC1 and PC4 could differentiate all lipstick samples, however, they only used red lipsticks, which were noted to have greater degree of colour variation [39] thus potentially making them easier to differentiate.

3.3. Final database construction

After the PCA scores plots were assessed and the number of variables was reduced by applying PCA as described in section 2.3.2, LDA prediction model was constructed, which was then used for classification of samples for the database from the test set. The prediction model was 93.45% accurate and confusion matrix for the model demonstrated where misclassifications were likely to occur in the test set. As predicted from assessment of PCA scores plots, in the training set, most of misclassifications happened with NYX V and NYX WV samples and LES and LEB samples. The two NYX shades were fair and similar in colour to each other but had different undertones. NYX V is described as "vanilla with neutral undertone", whereas NYX WV is "light with warm undertone" [8]. Likewise, LES and LEB samples are very similar in colour but have different undertones, cold (LES) and neutral (LEB) [11].

When this model was used for LDA classification of unknown samples, the following results were obtained (Tables 4 and 5). Specificity and accuracy (Table 4) were pretty high across all shades (above 97.4%), while False Positive Rate remained below 1% except for LEB samples, which had False Positive Rate of 1.8%. Notably, True Positive Rate (sensitivity) for darker (Table 1, samples 4–6, 10–12, 16–18) shades tended to be lower compared to lighter (Table 1, samples 1–3, 7–9, 13–15) shades, with values as low as 76.7% for LES samples. Median sensitivity value for lighter shades was 96.7% and 90.0% for darker shades.

Whilst in some cases sensitivity values were lower than desired (below 95%), looking at the confusion matrix (Table 5), the three pairs of shades with the lowest sensitivity values (NYXV and NYXWV, NYXC and NYXDW, LES and LEB) were very similar in colour to each other and were made by the same respective brands.

LDA is a form of hard analysis, where samples are assigned to the nearest class, even if the sample does not actually belong to that class rather than soft analysis where a sample may be assigned to multiple

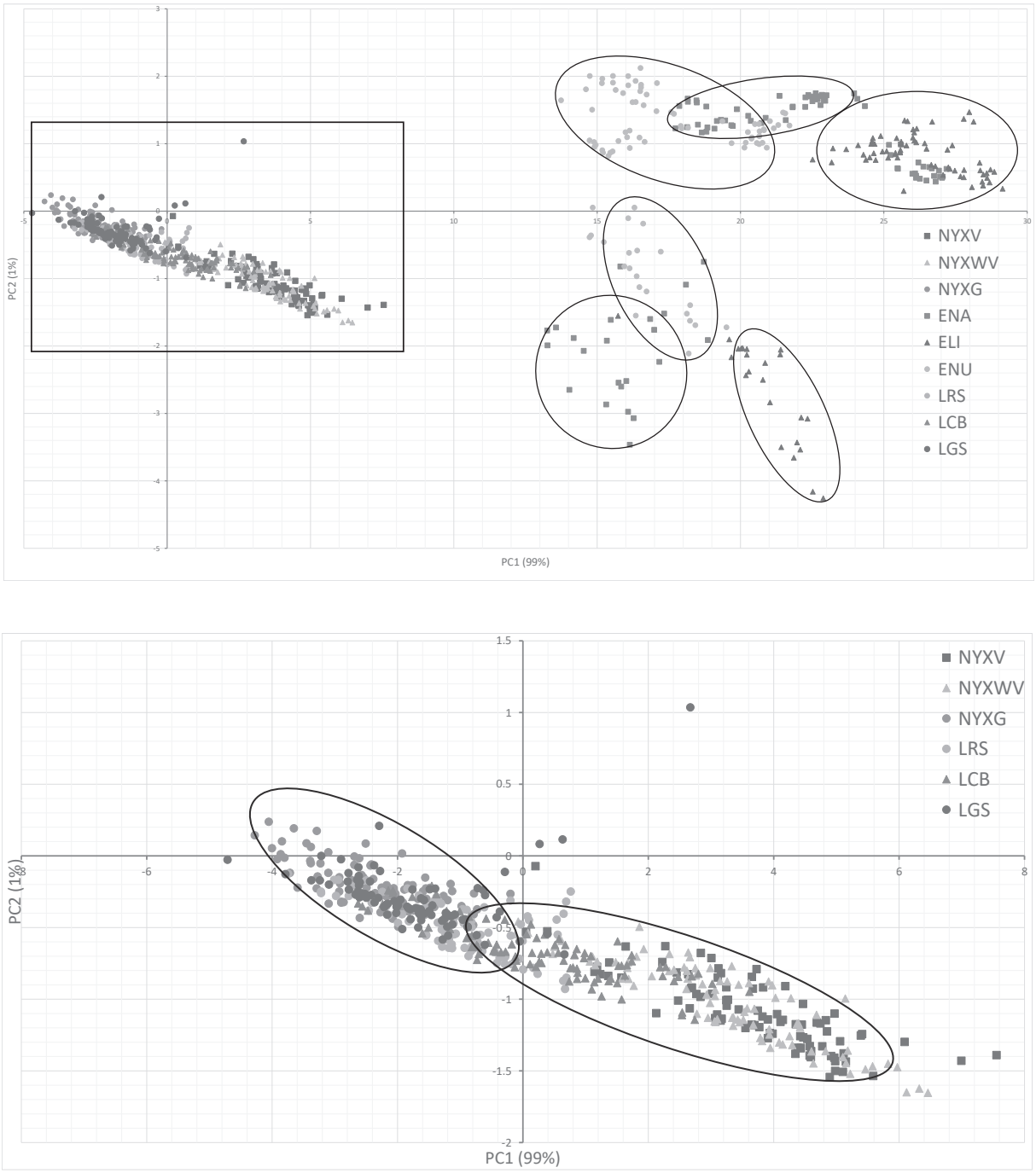

Fig. 2. 2D PCA scores plot of lighter foundation shades (Table 1, samples 1–3, 7–9, 13–15) all together (2.1) and zoomed in part of the scores plot (2.2). The first two Principal Components were used, both of which contributed to sample separation. Foundation sample clustering appears to be based on colour, for example, lighter shades NYX V and NYX WV group together.

classes or none [27]. While hard analysis may lead to misclassification of a sample, such risk could be reduced by expanding the cosmetics database.

Sharma and colleagues [20] constructed PCA-LDA classification model which could classify all samples with 100% accuracy, however, only fair, light, and medium shades were used, which could have contributed to high classification accuracy. As observed in section 3.2 and 3.3, lighter foundation shades tended to be classified more accurately and had lower FPR (median 0.0% and average 0.3%) compared to darker shades (median 0.5% and average 0.5%). Studies done on other types of cosmetics, such as lipsticks [39–41] and vermilion (*sindoor*) [42], similar level of accuracy was achieved. With lipsticks, samples that were very similar to each other in colour, were more likely to be misclassified [39], similar to what was observed with NYX V, NYX WV and LES, LEB samples. In addition, there seemed to be a variation among different batches produced by the same manufacturer [40], which led to lower classification accuracy (81.48%). Kaur and coworkers [41] achieved 100% classification accuracy with LDA when samples were analysed in three individual groups separated by colour (brown, pink, red). LDA classification accuracy and sensitivity could potentially be improved by separating the foundation database in groups by colour, similar to studies done on lipsticks [39,41].

Type of material used as a matrix for smears did not appear to affect classification results in a meaningful way, which was likely because the

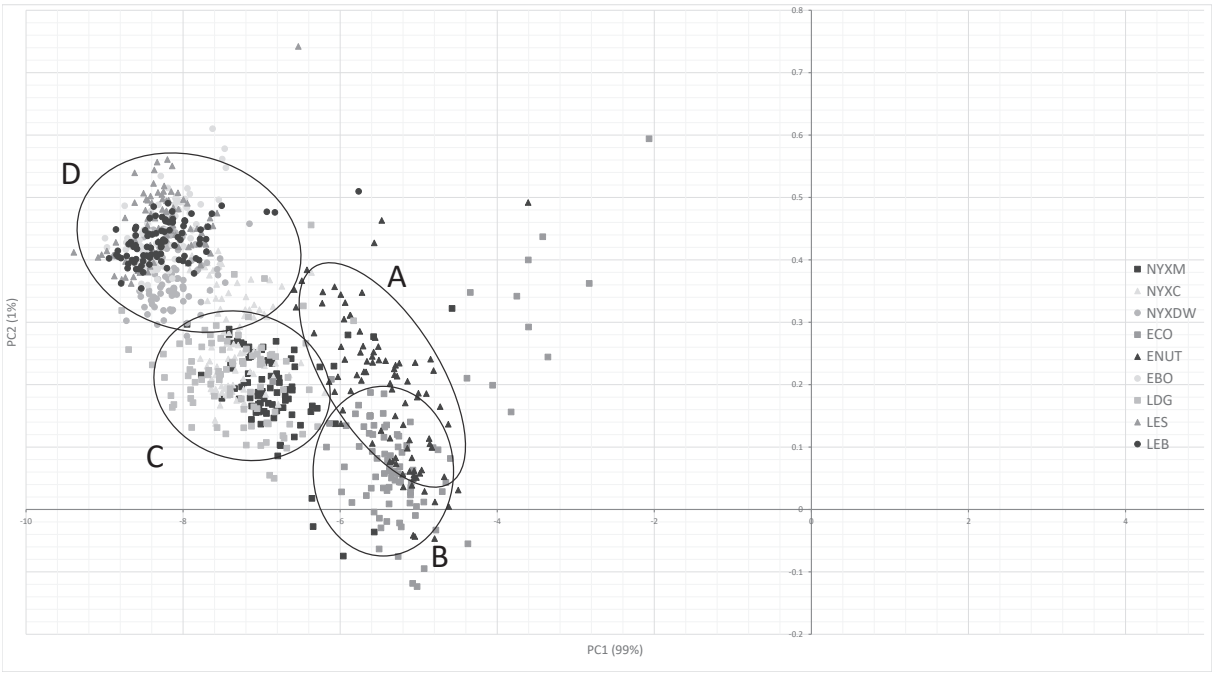

Fig. 3. PCA scores plot of darker foundation shades (Table 1, samples 4–6, 10–12, 16–18). Cluster A appears to include ENUT samples, cluster B – ECO samples, cluster C – NYX M, NYX C, LDG, and cluster D – NYX DW, EBO, LES, LEB.

Table 4
LDA classification table showing performance evaluation measures of LDA classification model based on raw data with no pre-treatment. Spec. – Specificity, Sens. – Sensitivity, FPR – False Positive Rate.

Samples	Spec. (%)	Sens. (%)	Accuracy (%)	FPR (%)
NYX V	99.4	83.3	98.5	0.6
NYX WV	99.0	90.0	98.5	1.0
NYX G	100.0	93.3	99.6	0.0
NYX M	99.4	90.0	98.9	0.6
NYX C	99.2	90.0	98.7	0.8
NYX DW	99.8	83.3	98.9	0.2
ENA	99.6	100.0	99.6	0.4
ELI	100.0	100.0	100.0	0.0
ENU	100.0	93.3	99.6	0.0
ECO	100.0	100.0	100.0	0.0
ENUT	99.6	100.0	99.6	0.4
EBO	99.8	96.7	99.6	0.2
LRS	100.0	96.7	99.8	0.0
LCB	100.0	100.0	100.0	0.0
LGS	99.6	100.0	99.6	0.4
LDG	99.8	96.7	99.6	0.2
LES	99.2	76.7	98.0	0.8
LEB	98.2	83.3	97.4	1.8
average	99.6	93.0	99.2	0.4
median	99.7	95.0	99.6	0.3

samples were not aged and did not have sufficient time to be absorbed into the material they were smeared on. As was observed by Sharma and colleagues [20], cream foundations tend to be absorbed into porous substrates, thus reducing the sample concentration on the substrate surface, which may lead to weaker peaks or appearance of additional peaks. These interferences could not be completely resolved by spectra subtraction. Similar matrix effects were noted by López-López, Özbek and García-Ruiz [12]; out of 49 different lipstick marks applied on 12 different surfaces, blue T-shirt (56% cotton, 44% viscose), lab coat (67% polyester, 33% cotton), paper cup, cigarette butt, and white plastic cup resulted in additional bands in the lipstick spectrum, hindering the identification of the lipstick samples. This was resolved by spectra subtraction of the surface, which resulted in a cleaner sample spectrum.

Very few papers on cosmetic foundation identification in a forensic context are available [7,20] and of those available, a relatively narrow range of foundation shades and substrate materials were used., and only one of them used chemometrics to classify cosmetic foundations quantitatively [7]. Using a wider range of materials and collecting more data to form each model would significantly broaden the prospect of the database and improve its accuracy of assignment. It is important to mention that addition of darker foundation shades is crucial in database construction. The Crime Survey for England and Wales (year ending March 2020) reported that those of Mixed and Black or Black British ethnicity were significantly more likely to be a victim of sexual assault as compared to White, Asian or Other ethnic groups (1.8, 2.6, 4.5 times for Mixed and 1.45, 2.1, 3.6 times for Black/Black British ethnicities) [14].

Currently, there is no database for cosmetic foundation, but such development will provide investigators with an invaluable tool to establish links quickly and efficiently; thus, providing an initial case lead in forensic investigations. This method can also be used for non-destructive preliminary analysis to identify the nature of the sample before any destructive techniques are employed for further identification.

4. Conclusion

A combination of NIR and PCA-LDA has proved to be a promising method as it has been sensitive enough to distinguish between the chemical compositions of most shades of foundation, with the percentage of correct assignment for the foundation at 93.0%. The remaining 27% does not indicate that the LDA was the origin of the discrepancies, as work on the other types of cosmetics has shown. The aforementioned studies [20,35–38] used ATR-FTIR for data collection, which might have aided a more accurate LDA classification.

It is important to note that owing to the limitations of the hardware available to us, larger smears have been used than what might reasonably be expected at a scene of a crime. In addition, NIR is susceptible to interferences from the materials themselves, which could prove to be an issue when expanding the database to a broader range of materials but the use of an NIR-microscope or FTIR may help overcome these issues. Chemometric statistical analysis also allows a critical evaluation of the results obtained and may help overcome matrix interferences. Other

Table 5
LDA confusion matrix showing correct and incorrect samples classification. Top raw represents actual class of each sample and the first raw represents classes that were assigned to said samples. Each group consisted of 30 NIR spectra.

Actual Class/ Predicted class	NYX V	NYX WV	NYX G	NYX M	NYX C	NYX DW	ENA	ELI	ENU	ECO	ENUT	EBO	LRS	LCB	LGS	LDG	LES	LEB
NYX V	25	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NYX WV	5	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NYX G	0	0	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NYX M	0	0	0	27	3	0	0	0	0	0	0	0	0	0	0	0	0	0
NYX C	0	0	0	3	27	1	0	0	0	0	0	0	0	0	0	0	0	0
NYX DW	0	0	0	0	0	25	0	0	0	0	0	1	0	0	0	0	0	0
ENA	0	0	0	0	0	0	30	0	2	0	0	0	0	0	0	0	0	0
ELI	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0	0	0
ENU	0	0	0	0	0	0	0	0	28	0	0	0	0	0	0	0	0	0
ECO	0	0	0	0	0	0	0	0	0	30	0	0	0	0	0	0	0	0
ENUT	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0	1	0	1
EBO	0	0	0	0	0	1	0	0	0	0	0	29	0	0	0	0	0	0
LRS	0	0	0	0	0	0	0	0	0	0	0	0	29	0	0	0	0	0
LCB	0	0	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0	0
LGS	0	0	2	0	0	0	0	0	0	0	0	0	0	0	30	0	0	0
LDG	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	29	0	0
LES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	23	4
LEB	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	7	25

chemometric methods, such as Soft Independent Modelling of Class Analogies (SIMCA) or Partial Least Squares Discriminant Analysis (PLS-DA) could be explored to identify the most suitable classification method that would be suitable for a wide range of foundation shades.

There is clearly extensive research left to be carried out surrounding foundations as trace evidence, the benefits of which will be extensive through the forensic community.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] C. Roux, Trace evidence: the silent witness, Australian Sci. 21 (6) (2000 Jul) 22.
[2] L.S. Castillo-Peinado, M.D. Luque de Castro, An overview on forensic analysis devoted to analytical chemists, Talanta 167 (2017) 181–192.
[3] A.R.W. Jackson, J.M. Jackson, Forensic Science, 4th ed., Pearson Education, Harlow, England; Boston, 2017.
[4] E. Mistek, M.A. Fikiet, S.R. Khandasammy, I.K. Lednev, Toward Locard's exchange principle: recent developments in forensic trace evidence analysis, Anal Chem. 91 (1) (2019) 637–654.
[5] M.A. Seguí, M.M. Feucht, A.C. Ponce, F.A.V. Pascual, Persistent lipsticks and their lip prints: new hidden evidence at the crime scene, Forensic Sci. Int. 112 (1) (2000) 41–47.
[6] M. Gładysz, M. Król, P. Kościelniak, Differentiation of red lipsticks using the attenuated total reflection technique supported by two chemometric methods, Forensic Sci. Int. 1 (280) (2017 Nov) 130–138.
[7] A. Gordon, S. Coulson, The evidential value of cosmetic foundation smears in forensic casework, J. Forensic Sci. 49 (6) (2004) 1–9.
[8] Y. Moriuchi, S. Tominaga, T. Horiuchi, Precise analysis of spectral reflectance properties of cosmetic foundation, in: A.B. Salberg, J.Y. Hardeberg, R. Jenssen (Eds.), Image Analysis, Berlin, Heidelberg, Springer, Berlin Heidelberg, 2009, pp. 138–148.
[9] BORN TO GLOW! NATURALLY RADIANT FOUNDATION [Internet]. NYX Professional Makeup. [cited 2021 Mar 10]. Available from: https://www.nyxcosmetics.co.uk/born-to-glow-naturally-radiant-foundation/NYX_749.html.
[10] FLAWLESS FINISH FOUNDATION [Internet]. e.l.f. [cited 2021 Mar 10]. Available from: https://www.elfcosmetics.co.uk/flawless-finish-foundation/300061.html?dwvar_300061_color=Natural%20%28Previously%20Porcelain%29&cgid=face-foundation#start=1.
[11] L'Oréal Paris True Match Liquid Foundation with SPF and Hyaluronic Acid 30ml [Internet]. L'Oréal Paris. [cited 2021 Mar 10]. Available from: https://www.loreal-paris.co.uk/true-match/foundation/1-n-ivory.
[12] M. López-López, N. Özbek, C. García-Ruiz, Confocal Raman spectroscopy to trace lipstick with their smudges on different surfaces, Talanta 1 (123) (2014) 135–139.
[13] S. Pasieczna-Patkowska, T. Olejnik, Analysis of cosmetic products using different IR spectroscopy techniques, Annales UMCS, Chemia. 68 (2014) 106–195.
[14] Stripe N. Sexual offences prevalence and victim characteristics, England and Wales [Internet]. Office for National Statistics. 2021 [cited 2021 Apr 10]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/datasets/sexualoffencesprevalenceandvictimcharacteristicsenglandandwales.
[15] R. Chophi, S. Sharma, S. Sharma, R. Singh, Trends in the forensic analysis of cosmetic evidence, Forensic Chem. 1 (14) (2019), 100165.
[16] E. Bergslien, An Introduction to Forensic Geoscience, 2nd ed., Wiley, Chichester, West Sussex; Hoboken, NJ, 2012, p. 482.
[17] R.B. Chavan, N. Bhargavi, A. Lodagekar, N.R. Shastri, Near infra red spectroscopy: a tool for solid state characterization, Drug Discovery Today 22 (12) (2017) 1835–1843.
[18] J. Dubois, J.-C. Wolff, J. Warrack, J. Schoppelrei, E. Lewis, NIR chemical imaging for counterfeit pharmaceutical product analysis, Spectroscopy (Santa Monica). 22 (2007) 40–50.
[19] M.J. Vredenbregt, L. Blok-Tip, R. Hoogerbrugge, D.M. Barends, D. de Kaste, Screening suspected counterfeit Viagra® and imitations of Viagra® with near-infrared spectroscopy, J. Pharm. Biomed. Anal. 40 (4) (2006) 840–849.
[20] S. Sharma, R. Chophi, H. Kaur, R. Singh, Differentiation of cosmetic foundation creams using attenuated total reflection fourier-transform infrared spectroscopy: a rapid and nondestructive approach in trace evidence analysis, J. Forensic Sci. 65 (3) (2020) 751–761.
[21] B. Lavine, J. Workman, Chemometrics, Anal. Chem. 78 (12) (2006) 4137–4145.
[22] G. Reich, Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications, Adv. Drug Deliv. Rev. 57 (8) (2005) 1109–1143.
[23] A. Angrish, R. Kumar, R. Chauhan, V. Sharma, On the IR spectroscopy and chemometric based rapid and non-destructive method for the investigation of sunscreen stains: application in forensic science, Spectrochim. Acta Part A Mol. Biomol. Spectrosc. 242 (2020), 118708.
[24] J.N. Miller, J.C. Miller, Statistics and Chemometrics for Analytical Chemistry, sixth ed., Prentice Hall, Harlow, England, 2010.
[25] L. Gál, M. Oravec, P. Gemeiner, M. Čeppan, Principal component analysis for the forensic discrimination of black inkjet inks based on the Vis–NIR fibre optics reflection spectra, Forensic Sci. Int. 257 (2015) 285–292.
[26] B. Stumpe, T. Engel, B. Steinweg, B. Marschner, Application of PCA and SIMCA statistical analysis of FT-IR spectra for the classification and identification of different slag types with environmental origin, Environ. Sci. Technol. 46 (7) (2012) 3964–3972.
[27] C.R. Rao, The utilization of multiple measurements in problems of biological classification, J. Roy. Stat. Soc.: Ser. B (Methodol.) 10 (2) (1948) 159–193.
[28] R.G. Brereton, G.R. Lloyd, Partial least squares discriminant analysis: taking the magic away: PLS-DA: taking the magic away, J Chemometrics. 28 (4) (2014) 213–225.
[29] E. Kaznowska, J. Depciuch, K. Szmuc, J. Cebulski, Use of FTIR spectroscopy and PCA-LDC analysis to identify cancerous lesions within the human colon, J. Pharm. Biomed. Anal. 134 (2017 Feb) 259–268.
[30] E. Kaznowska, J. Depciuch, K. Łach, M. Kołodziej, A. Koziorowska, J. Vongsvivut, et al., The classification of lung cancers and their degree of malignancy by FTIR, PCA-LDA analysis, and a physics-based computational model, Talanta 186 (2018) 337–345.
[31] M. Blanco, M. Alcalá, J. Planells, R. Mulero, Quality control of cosmetic mixtures by NIR spectroscopy, Anal Bioanal Chem. 389 (5) (2007) 1577–1583.
[32] R.J. Barnes, M.S. Dhanoa, S.J. Lister, Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra, Appl Spectrosc. 43 (5) (1989) 772–777.
[33] Sokolova M, Japkowicz N, Szpakowicz S. Beyond Accuracy, F-Score and ROC: A Family of Discriminant Measures for Performance Evaluation. In: Sattar A, Kang B, editors. AI 2006: Advances in Artificial Intelligence [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 2006 [cited 2021 Jul 7]. p. 1015–21. (Hutchison D, Kanade T, Kittler J, Kleinberg JM, Mattern F, Mitchell JC, et al., editors. Lecture Notes in Computer Science; vol. 4304). Available from: http://link.springer.com/10.1007/11941439_114.

[34] M. Khajeh Mehrizi, S.M. Mortazavi, S. Mallakpour, S.M. Bidoki, M. Vik, M. Vikova, Effect of carbon black nanoparticles on reflective behavior of printed cotton/nylon fabrics in visible/near infrared regions, Fibers Polym. 13 (4) (2012) 501–506.
[35] Q. Gao, X. Wu, Y. Fan, Solar spectral optical properties of rutile TiO2 coated mica–titania pigments, Dyes Pigm. 109 (2014) 90–95.
[36] E. Aranzabe, M.I. Arriortua, A. Larrañaga, A. Aranzabe, P.M. Villasante, R. March, Designing multifunctional pigments for an improved energy efficiency in buildings, Energy Build. 147 (2017) 9–13.
[37] V. Fang, J. Kennedy, J. Futter, J. Manning, A review of near infrared reflectance properties of metal oxide nanostructures [Internet]. 2013 [cited 2021 Jul 17] p. 23. Report No.: 2013/39. Available from: https://www.gns.cri.nz/static/pubs/2013/SR%202013-039.pdf.
[38] A. Candolfi, R. De Maesschalck, D. Jouan-Rimbaud, P.A. Hailey, D.L. Massart, The influence of data pre-processing in the pattern recognition of excipients near-infrared spectra, J. Pharm. Biomed. Anal. 21 (1) (1999) 115–132.
[39] J.X.W. Wong, G. Sauzier, S.W. Lewis, Forensic discrimination of lipsticks using visible and attenuated total reflectance infrared spectroscopy, Forensic Sci. Int. 298 (2019) 88–96.
[40] R. Chophi, S. Sharma, R. Singh, Forensic analysis of red lipsticks using ATR-FTIR spectroscopy and chemometrics, Forensic Chem. 17 (2020), 100209.
[41] K. Kaur, P.K. Yadav, G.S. Bumbrah, R.M. Sharma, Forensic classification of lipsticks using attenuated total reflectance – Fourier transform infrared (ATR-FTIR) spectroscopy, Vib. Spectrosc. 110 (2020), 103146.
[42] R. Chophi, S. Sharma, R. Singh, Discrimination of vermilion (*sindoor*) using attenuated total reflectance fourier transform infrared spectroscopy in combination with PCA and PCA-LDA, J Forensic Sci. 66 (2) (2021) 594–607.