Hierarchical classification models and Handheld NIR spectrometer to human blood stains identification on different floor tiles

Fonseca, Aline C.S.; Pereira, José F.Q.; Honorato, Ricardo S.; Bro, Rasmus; Pimentel, Maria Fernanda

Published in:
Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy

DOI:
10.1016/j.saa.2021.120533

Publication date:
2022

Document version
Peer reviewed version

Citation for published version (APA):
Hierarchical Classification Models and Handheld NIR Spectrometer to Human Blood Stains Identification on Different Floor Tiles

Aline C.S. Fonseca\textsuperscript{a}, José F.Q Pereira\textsuperscript{a,b,*}, Ricardo S. Honorato\textsuperscript{c}, Rasmus Bro\textsuperscript{d}, Maria Fernanda Pimentel\textsuperscript{e}

\textsuperscript{a} Federal University of Pernambuco, Department of Fundamental Chemistry, Av, Jornalista Aníbal Fernandes, 50.740-560, Cidade Universitária, Recife, Brazil.
\textsuperscript{b} State University of Campinas, Institute of Chemistry, Campinas, PO Box 6154, 13083-970, Brazil.
\textsuperscript{c} Federal Police, Cais do Apolo, 321, 50030-907, Recife, PE, Brazil.
\textsuperscript{d} University of Copenhagen, Department of Food Science, Rolighedsvej 26, DK-1958 Frederiksberg, Denmark
\textsuperscript{e} Federal University of Pernambuco, Department of Chemical Engineering, Av. dos Economistas, Cidade Universitária, s/n, 50.740-590, Recife, PE, Brazil

ABSTRACT

One of the most important types of evidence in certain criminal investigations is traces of human blood. For a detailed investigation, blood samples must be identified and collected at the crime scene. The present study aimed to evaluate the potential of the identification of human blood in stains deposited on different types of floor tiles (five types of ceramics and four types of porcelain tiles) using a portable NIR instrument. Hierarchical models were developed by combining multivariate analysis techniques capable of identifying traces of human blood (HB), animal blood (AB) and common false positives (CFP). The spectra of the dried stains were obtained using a portable MicroNIR spectrometer (Viavi). The hierarchical models used two decision rules, the first to separate CFP and the second to discriminate HB from AB. The first decision rule, used to separate the CFP, was based on the Q-Residual criterion considering a PCA model. For the second rule, used to discriminate HB and AB, the Q-Residual criterion...
were tested as obtained from a PCA model, a One-Class SIMCA model, and a PLS-DA model.
The best results of sensitivity and specificity, both equal to 100%, were obtained when a PLS-DA model was employed as the second decision rule. The hierarchical classification models built for these same training sets using a PCA or SIMCA model also obtained excellent sensitivity results for HB classification, with values above 94% and 78% of specificity. No CFP samples were misclassified. Hierarchical models represent a significant advance as a methodology for the identification of human blood stains at crime scenes.

Keywords: Near Infrared, Hierarchical Model, Handheld Spectrometer, Human Blood Stains, Crime Scenes, Classification

1. Introduction

Blood is one of the most valuable traces that can be recovered at certain types of crime scenes. To elucidate the events of the crime, using blood traces, criminal investigators obtain different types of information such as shape, dispersion, quantity, and position of the blood stains. This information enables identification of the lawbreaker and suggests how the crime was committed, helping the reconstruction of the crime scene [1,2]. For this evidence to be investigated and subjected to further DNA analysis, however, blood stains must be properly identified and collected at the crime scene. Traces of blood may be visually identified, but their reddish color, can be confused with other substances that have similar colors, the so-called common false positives [3].

When apparently bloody stains are found in a crime scene, they are subjected to tests capable of determining whether they are truly blood. These tests generally involve catalytical chemical reactions and the use of oxidizing reagents and indicators that change color when the oxidation reaction is catalyzed by hemoglobin happens. The most used presumptive tests are based on colorimetric reactions such as the Kastle-Meyer and Hemastix test [4], or on the
emission of light due to chemiluminescence or fluorescence when reacted with luminol [5,6]. These methods have some disadvantages, e.g., false positive results due to environmental contamination [7] and may be destructive, thus hindering posterior DNA analysis. The use of luminol needs special conditions so that its properties are not altered, for example, the need to be refrigerated and used in a dark environment. Another important factor is that most presumptive tests are not specific for human blood, meaning that traces of animal blood can be confused with human blood leading to the need of confirmatory tests [7-11].

Different studies show that an alternative to confirm the presence of human blood is by the use of spectroscopic techniques [9-18]. Some advantages of these techniques are that they are not destructive (allowing complementary analyses of the evidence) and, since they may be used with portable equipment, they can be taken to the crime scene itself [19-21]. For the identification of blood traces, the advantages mentioned are of great interest, since blood is a complex material and after the sample has been identified, it may be necessary to perform further analyses, such as DNA.

The popularization of portable equipment makes the application of these techniques even more manageable for forensic purposes, since the instruments are lightweight, have a relatively low cost, do not require unnecessary reagents, and can be carried to the site of a field analysis, allowing the investigator to save time [20-22]. In addition to the natural complexity involving the interpretation of spectra obtained in the near infrared region, the use of portable instruments at crime scenes makes the analysis even more complex as the traces may be found in different substrates, contexts, and conditions.

Chemometric techniques enable forensic specialists to extract information of interest in a more accurate, fast and comprehensive way by using chemical statistical methods. Currently,
there are several chemometric methods that can be applied to assist in the identification of blood samples, as well as to estimate the time of their deposition [23-25].

Edelman et al. [3] published a study with the objective of identifying and dating bloodstains. Spectra in the NIR region (800–2778 nm) of bloodstains and other substances with reddish color were collected on cotton fabrics and Partial Least Squares (PLS) regression models were constructed. The models obtained 100% sensitivity and specificity for the classification of blood samples. It is important to highlight that this work is one of the pioneering studies in the use of NIR spectroscopy for blood identification and has shown the advantages of dating blood samples as soon as possible after they have been deposited. Regression models using PLS were created individually for each colored background, which also would be difficult to do in a real situation in a criminal context. Other works have been developed using infrared spectroscopy for blood identification purposes, such as the study by Li and coworkers [26] who used diffuse reflectance spectra in the visible and near infrared regions to classify the blood of five species (dog, goat, rhesus monkey, rat and human). Spectra acquisition were made on 1200 samples deposited in test tubes with 5 ml of blood in the presence of an anticoagulant. An Artificial Neural Network (ANN) classification algorithm was used to build the models and 20% of the training data was used as a validation set. The results showed that the best values in terms of robustness and precision for species recognition were achieved when a combination of spectra in the visible and in the NIR were employed for model’s construction.

Spectra of ATR FT-IR were used by Lin et al. [14] to estimate the age of blood stains in indoor and outdoor environments that had been deposited for up to 107 days. The samples were placed on glass plates and nineteen collection points were defined over time (between 0 and 107 days). No on-site analyses were performed, and each bloodstain sample was collected
in an Eppendorf tube and mixed with 10 µL saline before being analyzed using ATR. The constructed PLS model was able to estimate the age of the blood stains, but the approach was more useful for samples collected in the long term (7-85 days), regardless of whether they were found in an internal or external environment. Sharma and coworkers [27] conducted a study in which ATR FT-IR spectroscopy was used to distinguish menstrual blood from peripheral blood, vaginal fluid, seminal fluid and other non-biological substances. The samples were deposited on different surfaces (for example: glass, plastic, paper, wood, cotton, and paper) and dried for 3 days, before obtaining the spectra. Classification and regression models were constructed using PCA-LDA, and PLS-R, respectively. The PCA-LDA and PLS-DA model showed a correct classification rate of 100% for menstrual and peripheral blood. This study is interesting because it includes another type of blood sample that may be involved in crime scenes involving sexual violence. Although they used different substrates, the samples had been removed from the surface and placed in an Eppendorf tube, and then the spectra were obtained. So, this analysis could not be carried out without altering the environment of the evidence. This is less likely to happen with portable equipment.

In an earlier study, our research group used a portable near-infrared spectrometer to identify blood stains on different substrates, where different classification models specific to each substrate (ceramic, porcelain tiles, metal, and glass) were built separately [20]. Several chemometric techniques were used to build the classification models: SIMCA, Linear Discriminant Analysis using Genetic Algorithm (GA-LDA), Linear Discriminant analysis using Successive Projections Algorithm SPA-LDA, and PLS-DA. These models were built individually for each type of substrate in the data set, and the best results were obtained with the models built using PLS-DA and GA-LDA whose sensitivity and specificity values were equal to one for all substrates [20]. The work has the merit of highlighting the potential of portable equipment for the identification of human blood; as mentioned, however, the models
were built separately for each substrate and, therefore, not validated for use in real cases. In addition, models constructed using Discriminant Analysis for human blood classification, such as PLS-DA, have difficulty in correctly classifying any false positives that were not used to build the model as boundaries between classes.

Seeking to solve the problem of combining different substrates, a more recent work developed by Pereira et al. [28] analyzed suspicious stains on cotton and synthetic fabrics with different colors and patterns. This studied associated NIR hyperspectral images with hierarchical models by a fusion of chemometric techniques. The main improvement in chemometrics was the use of hierarchical modeling to combine the techniques that made it possible to peel off layers of complexity by first classifying simpler and clearer subproblems. The hierarchical classification models built obtained excellent results and proved to be quite efficient. The sensitivity and specificity results were all above 95%, and for the samples prepared with the synthetic fabric, 100% sensitive for the identification of human blood. Hierarchical Models are a strategy to execute chemometric models in sequence through operations in a decision tree format. When applied for the purpose of classification, they are able to solve problems where only one of the classification techniques such as PLS-DA or SIMCA is not sufficient or appropriate to answer the problem, so more than one method can be used to establish the classification criterion [28].

The main purpose of the present work was to continue both pioneer studies mentioned above, aiming to create a unified and robust approach capable of identifying human blood stains directly deposited on tile floors with different compositions. Human blood (20 donors), animal blood (4 animals) and common false positive substances (7 commercial products) were deposited on five types of ceramics and four types of porcelain tiles, with different colors and
textures. The collected infrared spectra were used to develop a hierarchical model with two decision rules to first separate the blood samples and then identify the human blood samples.

2. Materials and Method

2.1 Sample Preparation

The blood samples were collected from the volunteers following the operational procedure described by Pereira et al [20]. Human blood (HB) samples were collected using individual, sterilized needles directly into the capillaries of the fingers of 22 donors (12 women and 10 men). Animal blood samples (AB) from five animals (3 sheep and 2 horses) were acquired, supplied by the veterinary hospital of the Federal Rural University of Pernambuco (UFRPE). Seven commercial products (balsamic vinegar, red lipstick, red wine, red bell pepper, jam, ketchup and soy sauce) of reddish color that could be visually confused with blood stains (here called common false positive - CFP) were selected from among the products tested by Edelman and coworkers [3]. The blood and CFP samples were deposited directly on the substrates (Fig. 1a) or using a Pasteur pipette (2 drops). The substrates were 9 different types of floor: 5 ceramic and 4 porcelain tiles (Fig. 1b). There was no control over the volume of blood deposited on each tile and the number of drops deposited on each substrate was different due to the amount of material collected from each donor. All donors signed a free consent form according to the requirements of the Research Ethics Committee involving Human Beings at the Federal University of Pernambuco (nº 1.059.225).
Fig. 1. (a) Human blood stains encoded and dried on one of the floors; (b) Image of the cleaned substrates (floors tiles).

2.2 Instrumentation and Spectra Acquisition

The spectra of blood and CFS samples were collected directly on the stains deposited on the substrate (tiles). Spectral measurements were performed using a MicroNir 1700 spectrometer manufactured by VIAVI with dimensions of 45 mm in diameter and 42 mm in height and 60g weight. A linear variable filter (LVF) was attached to an array of Indium Gallium and Arsenic detectors (InGaAs). Two Tungsten lamps were used as the sources of infrared radiation. The spectral range was from 908 nm to 1676 nm and the optical resolution was 1.25% of the central wavelength. Spectra were obtained with 64 scans and an integration time of 5 milliseconds. The operational procedure performed for the acquisition of the spectra followed a methodology similar to that described by [20]. Before spectra collection, the stains were allowed to dry for six days at a temperature of around 25 °C. In order to contemplate the maximum variability of each sample, four spectra were acquired from different locations of each stain and 3 spectra were also collected directly from the clean surface of the substrates. The data set consisted of 986 spectra, including 506 spectra of human blood, 200 spectra of animal blood and 280 spectra of common false positives. Table 1 details the number of spectra for each of the classes per substrate (tiles) contained in the database.
Table 1 – Number of spectra collected on each type of substrate.

<table>
<thead>
<tr>
<th>Substrates</th>
<th>Human Blood (HB)</th>
<th>Animal Blood (AB)</th>
<th>Common False Positive (CFP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PO1</td>
<td>65</td>
<td>32</td>
<td>-</td>
</tr>
<tr>
<td>PO3</td>
<td>60</td>
<td>8</td>
<td>56</td>
</tr>
<tr>
<td>PO4</td>
<td>40</td>
<td>32</td>
<td>56</td>
</tr>
<tr>
<td>PO5</td>
<td>75</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CE1</td>
<td>69</td>
<td>32</td>
<td>-</td>
</tr>
<tr>
<td>CE3</td>
<td>64</td>
<td>32</td>
<td>56</td>
</tr>
<tr>
<td>CE4</td>
<td>-</td>
<td>32</td>
<td>-</td>
</tr>
<tr>
<td>CE5</td>
<td>67</td>
<td>32</td>
<td>56</td>
</tr>
<tr>
<td>CE6</td>
<td>66</td>
<td>-</td>
<td>56</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>506</strong></td>
<td><strong>200</strong></td>
<td><strong>280</strong></td>
</tr>
</tbody>
</table>

2.3 Data Analysis

Data treatment and statistical analysis were performed using Matlab software (MATLAB® R2010a 7.10.0.499, MathWorks) and the hierarchical models, as well as the PCA, SIMCA and PLS-DA models were performed using PLS_Toolbox (Eigenvector Research, Inc).

2.3.1 Data Evaluation and Pre-Processing.

Two main sources of variations are presented in the data set. Variations related to the types of stain are divided into three classes (HB, AB and CFP); variations among the substrates (floor tiles) are divided into 9 classes (CE1, CE3, CE4, CE5, CE6, PO1, PO3, PO4 and PO5). The original data were evaluated initially according to the set of substrates through the plots of the original, pre-processed spectra and scores/ loadings of the PCA models. Different pre-processing techniques were used to correct scattering of radiation and minimize significant differences between samples prepared on different substrates. The following techniques were
tested: Standard Normal Variate (SNV), Savitzky-Golay smoothing filter, and Savitzky-Golay 1st and 2nd derivatives (5 to 15 point window). The best combination of pre-processing techniques was chosen based on visual inspection of the spectra correction and on the best results obtained by scores and loadings plots obtained from a Principal Component Analysis (PCA) using all datasets. The cross validation was performed using random samples with ten splits and five iterations.

2.3.2 Definition of Training and Prediction Sets

In order to evaluate the influence of the effect of the differences of type of floor tile, three training/prediction sets were employed to build the models. The first prediction set was selected with samples from a floor tile that had only human blood samples (PO5), another tile with only animal blood samples (CE4) and finally, a tile that had common false positive samples (CE6). The first training set was formed from the remaining samples. Other combinations of floor tile classes were used to form two new training and prediction sets. The selection of the tiles to compose these two new sets was based on the visualization of the scores of the PCA built with all samples deposited on all floor tiles. The criterion used to choose the training and prediction sets was the similarity among the tiles in the score plot (This discussion will be further detailed in section 3.2). The floor tiles used in all the prediction sets were totally independent from the floor tiles used in the respective training set.

3. Results and Discussion

3.1 Spectral Analysis and Pre-Processing

Initially, the spectra of the samples deposited on the substrate (floor tiles) were individually evaluated. As high noise and scattering effects were observed at the beginning and end wavelengths, 15 points at the beginning and 10 points at the end of each spectra were removed.
Different pre-processing techniques were tested to reduce noise and scattering effects, such as SNV, 1st Derivative and 2nd Derivative with different smoothing windows (from 7 to 15 points). Better correction of unwanted effects, a homogeneous grouping of samples in scores, and clear loadings profiles were achieved using the 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial.

The original (Fig. 2a) and pre-processed (Fig. 2c) spectra of the dataset can be seen in Fig. 2. A considerable difference between the intensity of the CFP spectra compared with the HB and AB spectra can be seen in Fig. 2b and Fig. 2d, where the mean original spectra and the mean pre-processed spectra are shown. CFP spectra show more intense bands in the regions between 1150 nm to 1250 nm, and between 1400 and 1550 nm. The band in ~1460 nm can be attributed to the first hydroxyl stretching overtone (-OH), probably due to the presence of alcohol and carboxylic acids [29] such as the citric acid used in foods such as jellies and ketchup. In the supplementary material (Figure S1), the original and preprocessed spectra of animal blood (AB) and human blood (HB), with the colors representing the different floor tiles on which they were deposited.
Fig. 2. a) Original spectra of the stains deposited on the floor tiles.; b) mean of original spectra; c) pre-processed spectra and d) mean of pre-processed spectra with 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial; 15 points at the beginning and 10 points at the end of the spectra were removed;

A PCA model (Fig. 3) was built using the dataset pre-processed with a 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial, 15 points at the beginning and 10 points at the end of the spectra were removed. The PC1 explains 96.09% of the data variance and it corresponds mainly to variations among the common false positive spectra collected from seven different commercial substances. The CFP scores closest to the blood scores represent the red wine samples. PC2 (Fig. 3a) explains 1.84% of the variance. PC1 loadings (Fig. 3b) show that the variables with the greatest weight are in the range between 1300 and 1500 nm, which demonstrates a strong influence of CFP spectra in this component. This behavior was expected due to the wide range of substances used in the CFP and the scores distribution in this PC.
The region between 1440 and 1485 nm may be associated with the first stretching overtones of the -OH bond present in water, alcohol and carboxylic acids. Between 1347 and 1367 nm occurs the spectral absorption band associated with combinations of the second overtone of the -CH bond in methyl groups (-CH₃) that can be associated with hemoglobin molecules. Located around 1511 nm is the first over-stretching of the -NH bond belonging to proteins. At the beginning of the spectrum, at 1007 nm occurs the stretch band of the second overtone of the N-H bond. Between 1191 and 1194 nm occurs absorption of the second stretching overtone of the -CH bond [29].

**Fig. 3.** PCA model built using full dataset pre-processed with 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial. a) PC1xPC2 score plot; b) PCA loading plot.

The scores of the HB and AB classes have similar coordinates in PC1, while in PC2 the AB scores tend to be slightly more negative than the HB scores. To visualize these differences between human and animal blood with more accuracy, the common false positives samples were removed and another PCA model using only the blood spectra was built (Fig. 4). PC1xPC2 scores for the HB and AB classes can be seen in Fig. 4a, where a tendency of separation between the two classes in PC1 can be observed. Here the HB scores have more positive coordinates indicating a stronger contribution of variables in the range between 1000 and 1200 nm (Fig. 4b) while the variables in the range between 1300 and 1500 nm show higher for the AB samples.
Fig. 4. PCA model of blood spectra pre-processed with 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial. a) PC1xPC2 score plot; b) Loadings plot.

3.2 Training and Prediction Sets

The first external prediction set was defined based on the idea of removing the greatest possible number of floor tiles without compromising the training set variability. The first strategy was to keep out of the training set those floor tiles that had only a single class of samples deposited on their surface. This was the case of the PO5 floor tile, which had only 75 Human blood (HB) spectra, and the CE4 floor tile, which had only 32 Animal blood (AB) spectra. Common false positive (CFP) samples are present in all the other classes of floor tiles, so, the strategy was to keep out the floor that did not have AB samples (as this class is the one with the lower number of spectra). Thus, floor CE6 was removed (So the PS1 had 229 spectra and TS1 had 757 spectra). These removed samples are totally independent from those used in training set, and they were only used for the external prediction set. Prediction Set 1 (PS1) thus was formed by three substrates (PO5, CE4 and CE6), corresponding to 23.3% of the dataset. The Training Set 1 (TS1) was composed of the six remaining substrates (CE1, CE3, CE5, PO1, PO3 and PO4), totaling 757 spectra.

The criterion used to choose the other training/prediction sets was the difference or similarity between the floor tiles, visually observing the score plots of the PCA performed with
all samples. Figure 5 (a and b) shows the score plots, colored according to the type of substrates and Fig. 5c the loading plots (Fig. 5).

The PC1 versus PC3 scores can be seen in Fig. 5b. PC3 (6.51% of the explained variance) shows a cluster with more negative scores, mainly influenced by the band between 1200 and 1300 nm (Fig. 5c). PC3 highlights the difference between a cluster of scores with more negative coordinates (CE5 samples) and a cluster with positive coordinates (PO1 samples) (Fig 5b).

Fig. 5. PCA model of blood spectra stains pre-processed with 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial and colored according to the substrates. a) PC1xPC2 score plot; b) PC1xPC3 score plot; c) Loadings plot.

Prediction Set 2 (PS2) was built selecting samples from one substrate (floor tile) class with scores located more internally in the PCA score plot; then, another class of substrate was selected with scores located more externally: the PO4 class (dark blue), and the CE5 (purple), respectively. Prediction Set 3 (PS3) was built using samples from two classes of substrates with scores located more internally in the PCA model (Fig. 5): CE1 (red) and the substrate PO4 (dark blue). Table 2 details how the spectra of the different tiles types (substrates) were distributed in these three training/prediction sets.
Table 2 — Distribution of the types of spectra and floor tiles in the different training/prediction sets.

<table>
<thead>
<tr>
<th>Training set</th>
<th>Floor Tiles</th>
<th>Nº. Spectra</th>
<th>Prediction set</th>
<th>Floor Tiles</th>
<th>Nº. Spectra</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>HB  AB  CFP</td>
<td>PS1</td>
<td>CE4,PO5,CE6</td>
<td>141 32 56</td>
</tr>
<tr>
<td>TS1</td>
<td>CE1,CE3,CE5, PO1,PO3,PO4</td>
<td>365 168 224</td>
<td>PS1</td>
<td>CE4,PO5,CE6</td>
<td>141 32 56</td>
</tr>
<tr>
<td>TS2</td>
<td>CE1,CE3,CE4,CE6, PO1,PO3,PO5</td>
<td>375 136 168</td>
<td>PS2</td>
<td>CE5,PO4</td>
<td>131 64 112</td>
</tr>
<tr>
<td>TS3</td>
<td>CE3,CE4,CE5,CE6, PO1,PO3,PO5</td>
<td>397 136 224</td>
<td>PS3</td>
<td>CE1,PO4</td>
<td>109 64 56</td>
</tr>
</tbody>
</table>

To assess more accurately the difference between the floor tiles, spectra of floor tiles with no substance deposited on their surfaces were collected and evaluated as well. A PCA model (Fig. 6) was performed. In Fig. 6a, from the PC2 scores, some substrates have closer coordinates, such as CE1, CE4, CE6, PO3, PO4 and PO5; PC1 scores are more similar for CE1, CE4 and PO3, and CE5, CE6, PO4 and PO1. Looking at PC3 (Fig. 6b), it is possible to identify that PO4 and PO5 have similar scores for this component.

Fig. 6. PCA model of cleaned substrate (floor tile) spectra pre-processed with 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial. a) PC1xPC2 score plot; b) PC1xPC3 score plot; c) Loadings plot.

3.3 Hierarchical Model by Fusion of Chemometric Techniques
Based on the initial analyses, hierarchical models with two rules were developed. The first rule consisted of a PCA model responsible for identifying the differences among the CFP spectra; and the second rule was responsible for identifying the different types of blood (human and animal blood). Three different hierarchical models were built, adjusting the classification model to assess the second decision (human or non-human blood): the first model used a PCA model, the second used a SIMCA model and the third used a PLS-DA model. These three different versions of the hierarchical model were built for each of the three training sets described previously. The procedure described in more details below for TS1/PS1 was performed for the other two training/prediction sets as well.

3.3.1 Separating CFP from human/animal blood

The first decision in the hierarchical model was to identify whether the sample was blood or a common false positive and/or a possible outlier. A PCA model was built using only the blood spectra of the training set, leaving the CFP spectra out of the model. Fig. 7 shows the PC1xPC2 (Fig. 7a) and PC1xPC3 (Fig. 7b) scores of the blood spectra present in Training set 1, where the HB scores have more positive coordinates in PC1 in relation to the AB scores resembling the behavior of the scores in Fig. 4.
Fig. 7. PCA model of TS1 blood spectra pre-processed with 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial. a) PC1 x PC2 score plot; b) PC1 x PC3 score plot; c) Loadings plot; d) Residues plot of blood data, as well as the projection of the CFP samples (previously removed from the PCA model).

One of the graphs obtained from the construction of PCA models is the residual graph or influence plot. The spectra of CFP samples from the training set (previously removed) were then projected onto the PCA model, and their residue obtained, as shown in Fig. 7d. This plot was used to establish the threshold to be applied as the first decision rule in the hierarchical model. Q-Residuals is a parameter that indicates whether a sample is well-fit to the constructed model. Samples with a high Q-Residual value indicate that there is a great difference between their actual value and their projection in the model, which can be a possible outlier [30]. The red line in Fig. 7d illustrates the limit chosen to separate samples with Q-Residuals >1.5, which were considered non-blood samples (as in the case of the CFP samples). This strategy for defining classification limits or criteria for deciding whether a sample belongs or not to a class in the PCA model, based on the values of Q-Residuals, is also considered as a single criterion.
SIMCA [31]. Samples with Q-Residuals <1.5 were considered blood samples and were submitted to the next rule of the hierarchical model, as described below.

3.3.2 Identifying Human Blood

Three different methods were used to establish the second decision rule: the first was a PCA model, the second used a SIMCA model and the last, a PLS-DA model. Fig. 8 represents a schematic of the three hierarchical models evaluated.

![Scheme of the Hierarchical Model](image)

**Fig. 8.** Scheme of the Hierarchical Model

3.3.2.1 PCA rule (Hierarchical Model 1 – HM1)

A PCA model was built with only the human blood spectra from the training set. Three components were used to define the PCA model (92.5% variance). A tendency of separation between the HB samples can be observed in PC1 score plots (Fig. 9a). The most positive scores in PC1 belong to spectra of human blood samples deposited on tiles CE1, CE3 and PO3; the group with the most negative scores corresponds to the spectra of human blood samples deposited on tiles PO1, PO4 and CE5. The loadings of PC1, PC2 and PC3 show that there is some influence of noise in the initial variables, which coincides with the spectral range most affected by the substrate information. Similar to the procedure described in item 3.3.1, the animal blood spectra from the training set were projected onto the PCA model (built with only
human blood samples) and the influence plot (Fig. 9d) was used to establish the decision rule (Q-Residual <1.85, as illustrated by the red line).

Fig. 9. PCA model of TS1 human blood stains spectra pre-processed with 1st Savitzky-Golay derivative technique with 15 smoothing points and a 2nd degree polynomial. a) PC1xPC2 score plot; b) PC1xPC3 score plot; c) Loadings plot; d) Residues plot.

The sensitivity and specificity values of the internal validation obtained through the projection of Training set 1 (TS1) samples in this first hierarchical model (HM1) were 0.99 and 0.88, respectively. These values demonstrate a satisfactory fit of the model. As expected, the specificity value is lower than the sensitivity value; this occurred because some animal blood samples were classified as human blood (Fig. 9d). The first prediction set samples were then submitted to this HM1 (here called HM1.1), and the sensitivity and specificity values were calculated as equal to 1.0 and 0.78, respectively. All human blood and common false positive spectra were classified correctly, however, 69% of the animal blood spectra were classified as human blood.
The same procedure was used to build the hierarchical models for the other training sets (TS2 and TS3). The results of sensitivity (Sn) and specificity (Sp) for each model can be compared in Table 3. The figures of merit obtained from the internal validation for all models show sensitivity above 0.91 and specificity above 0.84. These values indicate a satisfactory result for the classification models.

The respective prediction sets (PS2 and PS3) were submitted to each of their hierarchical models, called here HM1.2 and HM1.3, respectively. For the second prediction set, the sensitivity and specificity values were 0.33 and 0.7, respectively. These results were expected, due to the selection of samples of this prediction set, where the scores of one of the floor tiles (CE5) had more external coordinates than the others, as seen in Figs. 5 and 6. The best results for the HM1 were obtained using the third prediction set, with sensitivity equal to 1.0 and specificity equal to 0.94. The third prediction set was created by using samples from two floor tiles whose scores were located more internally in the PCA model. However, some animal blood samples were classified as CFP, probably due to the lower number of samples of this class in the third training set.

3.3.2.2 SIMCA rule (Hierarchical Model 2 – HM2)

A One-class SIMCA model was built with human blood to compose the second hierarchical model (HM2). The final answer obtained in this model was the probability that the sample belonged to human blood class, with 95% confidence. Thus, the samples that obtained a probability value above 50% were classified as belonging to class HB. The sensitivity values obtained for internal validation were lower in all hierarchical models using SIMCA as a second rule (HM2) in relation to the model which used PCA as the second decision rule (HM1). SIMCA is based on the values of Q-Statistical and Hotelling’s $T^2$, which further restricts the classification limits. The sensitivity and specificity values for HM2.1 were equal to 0.92 and 0.85, which represents an increase in specificity in relation to the values obtained with the
HM1.1. There is a decrease in sensitivity from 1.0 to 0.92, however, representing a minor decrease in the correct classification of HB samples. When compared to the previous model (HM1.2), the HM2.2 built with the second training set also had a decrease in sensitivity from 0.33 to 0.03, while the specificity did not undergo any relevant statistical change, remaining at approximately 0.7. For the HM2.3 built with third training set, specificity increased to 1.0, but there was a small decrease in sensitivity once again, in relation to previous HM1.3 from 1.0 to 0.94.

3.3.2.3 PLS-DA rule (Hierarchical Model 3 – HM3)

The last approach used to differentiate human blood from animal blood employed PLS-DA. Human and animal blood spectra were considered for the construction of the discrimination boundaries between the two classes. Five latent variables were used to build the model. Internal validation was performed through cross-validation using random blocks. Fig.10a shows the human and animal blood score plot. All samples were classified correctly, with the specificity and sensitivity of the model equal to 1. From the VIP score plot (Fig. 10b), can be seen the variables which had the strongest influence on the discrimination between the two classes. The variables above the significance limit (red line) belong approximately to the region between 1370 nm and 1510 nm. This spectral range includes the regions of the first over-stretching of the -OH bond from the water at 1440-1485 nm and the first over-stretching of the -NH bond present in proteins (1511 nm) [29].

The output response of the hierarchical model (HM3.1, HM3.2 and HM3.3) was the probability of samples belonging to the human blood class. Regarding the specificity and sensitivity referring to internal validation, an increase was observed in relation to the models built using PCA (HM1) or SIMCA (HM2) (Table 3). The sensitivity and specificity values for
the prediction sets of the HM3.1 were equal to 1.0. Higher values were also obtained for HM3.3 built with TS3, with sensitivity equal to 0.97 and specificity equal to 1.

**Table 3** – Resume of the results of the internal validation and prediction sets for the three different hierarchical models built with spectra in the range from 1000 to 1614 nm.

<table>
<thead>
<tr>
<th>Hierarchical Model</th>
<th>Second Rule</th>
<th>Training Set</th>
<th>Internal Validation</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>HM1.1</td>
<td>PCA</td>
<td>TS1</td>
<td>0.98</td>
<td>0.88</td>
</tr>
<tr>
<td>HM1.2</td>
<td>PCA</td>
<td>TS2</td>
<td>0.98</td>
<td>0.84</td>
</tr>
<tr>
<td>HM1.3</td>
<td>PCA</td>
<td>TS3</td>
<td>0.98</td>
<td>0.94</td>
</tr>
<tr>
<td>HM2.1</td>
<td>SIMCA</td>
<td>TS1</td>
<td>0.91</td>
<td>0.93</td>
</tr>
<tr>
<td>HM2.2</td>
<td>SIMCA</td>
<td>TS2</td>
<td>0.96</td>
<td>0.91</td>
</tr>
<tr>
<td>HM2.3</td>
<td>SIMCA</td>
<td>TS3</td>
<td>0.99</td>
<td>0.92</td>
</tr>
<tr>
<td>HM3.1</td>
<td>PLS-DA</td>
<td>TS1</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>HM3.2</td>
<td>PLS-DA</td>
<td>TS2</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>HM3.3</td>
<td>PLS-DA</td>
<td>TS3</td>
<td>0.99</td>
<td>1.00</td>
</tr>
</tbody>
</table>

The first number of the models refers to the technique used as second rule (HM1 – PCA, HM2 – SIMCA and HM3 – PLS-DA). The second number refers to the training/prediction set used (HM*.1 – TS1/PS1; HM*.2 – TS2/PS2; HM*.3 – TS3/PS3).

### 3.3 Models built using a reduced spectral range

It was observed that at the beginning of the spectrum, in the range between 1100 and 1300 nm, there is more information about the floor tiles. Another way to identify the spectral region of greatest interest in relation to blood information is to analyze the Variable Importance in Projection (VIP) scores. This strategy is also widely used to perform variable selection.

Through this graph it is possible to observe which variables are most important in a PLS model [32]. The graph of VIP scores in Fig. 10b indicates that the spectrum region that is best able to discriminate between human and animal blood is the range after 1300 nm. Based on these observations, in an attempt to improve the classification of human blood as well as to minimize the influence of the type of substrate, we reduced the spectral range (from the 1304 nm to 1614 nm) and built new hierarchical models (HM4, HM5 and HM6). The substrate classes of the training sets were maintained, and the pre-processing chosen was the 1st derivative with a 2nd degree polynomial and smoothing of 7 points. Table 4 shows the results of the internal
validation and prediction sets for the different hierarchical models built with reduced spectra (from 1300 to 1614 nm).

Fig. 10. PLS-DA model built with TS1. a) HB and AB classification scores plot; b) VIP scores for the discrimination of the HB and AB.

Comparing the results for all HM using the reduced spectral range (Table 4), similar to the observed for the HM1, HM2 and HM3, it is evident that the worst results were also obtained using the second training set. However, a comparison of the values of Table 3 and 4 suggests that the models built with the training set 2 using a reduced spectral range (HM4.2, HM5.2 and HM6.2) were more robust in relation to the type of floor where the stains are deposited. The sensitivities for the prediction set were 1.0, 0.61 and 0.63 for HM4.2, HM5.2 and HM6.2 models, respectively; whereas for the HM1.2, HM2.2 and HM3.2 the values were 0.33, 0.03, and 0.22., respectively. There is an improvement in the sensitivity values compared to previous models, especially when considering training set 2 (TS2), indicating that the reduced spectral range models were less influenced by the floor tiles.

The best results were obtained using the first and the third training sets and the PLS-DA as second rule, which correctly classified 100% of the samples of each class and obtained results of sensitivity and specificity equal to 1. The hierarchical models built using the third training set also obtained excellent results for the classification of human blood even in the models that
used a PCA model built only with human blood as a decision rule (all the results of sensitivity and specificity were above 0.94); nonetheless, some animal samples were mistaken with common false positives.

Table 4 – Resume of the results of the internal validation and prediction sets for the three different hierarchical models built with spectra in the range from 1300 to 1614 nm

<table>
<thead>
<tr>
<th>Hierarchical Model</th>
<th>Second Rule</th>
<th>Training Set</th>
<th>Internal Validation</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>HM4.1</td>
<td>PCA</td>
<td>TS1</td>
<td>0.98 0.91</td>
<td>1.00 0.85</td>
</tr>
<tr>
<td>HM4.2</td>
<td>PCA</td>
<td>TS2</td>
<td>1.00 0.82</td>
<td>1.00 0.78</td>
</tr>
<tr>
<td>HM4.3</td>
<td>PCA</td>
<td>TS3</td>
<td>0.99 0.92</td>
<td>1.00 0.76</td>
</tr>
<tr>
<td>HM5.1</td>
<td>SIMCA</td>
<td>TS1</td>
<td>0.97 0.94</td>
<td>1.00 0.90</td>
</tr>
<tr>
<td>HM5.2</td>
<td>SIMCA</td>
<td>TS2</td>
<td>0.94 0.94</td>
<td>0.61 0.84</td>
</tr>
<tr>
<td>HM5.3</td>
<td>SIMCA</td>
<td>TS3</td>
<td>0.97 0.95</td>
<td>0.62 0.90</td>
</tr>
<tr>
<td>HM6.1</td>
<td>PLS-DA</td>
<td>TS1</td>
<td>0.99 1.00</td>
<td>1.00 1.00</td>
</tr>
<tr>
<td>HM6.2</td>
<td>PLS-DA</td>
<td>TS2</td>
<td>0.99 1.00</td>
<td>0.63 0.84</td>
</tr>
<tr>
<td>HM6.3</td>
<td>PLS-DA</td>
<td>TS3</td>
<td>0.99 1.00</td>
<td>1.00 0.94</td>
</tr>
</tbody>
</table>

The first number of the models refers to the technique used as second rule (HM1 – PCA, HM2 – SIMCA and HM3 – PLS-DA). The second number refers to the training/prediction set used (HM*.1 – TS1/PS1; HM*.2 – TS2/PS2; HM*.3 – TS3/PS3).

Except for the hierarchical models built with the second training set, all models correctly classified human blood better than 90%. When evaluating the models that used a PCA model or a PLS-DA model as a second decision rule, 100% of the human blood samples were classified correctly, that is, no false negative results were obtained. The PLS-DA models identified animal blood more accurately than the others. This was expected because this technique is based on the use of both the human and animal blood classes to delimit the boundaries between them. It should be noted that the animal blood class was formed not only by a smaller number of spectra, but also by samples from only two species of animals.

Regarding the results reported in the literature, no work has studied the identification of blood on different substrates (different types of floor tiles) using the same classification model for them all, simultaneously. Thus, the results presented here can be compared with the results obtained by Pereira et al. [20] in which the classification models were built individually for each type of substrate. The models built by Pereira et al. [20] to classify samples of HB, AB
and CFP deposited in a type of porcelain tile obtained sensitivity and specificity equal to 1, correctly classifying 100% of the samples using the techniques of SPA-LDA, GA-LDA and PLS-DA. Using the SIMCA technique, all of the human blood and common false positives samples were correctly classified. That study, however, was carried out with a single type of porcelain tile and a single type of ceramic. The results obtained in our study are similar to those of Pereira and coworkers [20], however, the present work developed more robust models and applied the tests under more real conditions, considering also the differences among the substrates (the floor tiles). Furthermore, the variability of CFP was not included in the construction of the hierarchical classification models, which makes it possible to reject different false positives in the first decision rule.

3.4 Floor Tile Evaluation Protocol

As discussed previously, it is important to avoid prediction of samples deposited on floor tiles (here also called substrates) that are significantly different from the training set. Hence, a protocol was developed for prior assessment of the floor tile. This protocol followed a methodology similar to that proposed for the construction of hierarchical models for blood identification. A hierarchical single-node model (called here HMFT) was built using a PCA model constructed only with cleaned floor tile spectra used on the training set. Three hierarchical models were built (HMFT.1, HMFT.2 and HMFT.3): one for each set of tiles represented in the three training sets (FTTS.1, FTTS.2 and FTTS.3). These models have a single node where the decision rule was the Q-Residuals provided through the PCA built using the same substrate spectra used in each training set.

The hierarchical models created for this protocol were able to identify whether the floor tile spectra of the prediction set could be reliably predicted considering the substrates (floor tiles) represented in the training set. For example, by projecting the spectra of the CE5 floor
tiles in the PCA model built with the floor tiles spectra of the same classes as the second training set (FTTS.2), the hierarchical model (HMFT.2) was able to identify that the CE5 spectra were not represented by this set. As can be seen in Fig. 11, the scores of the second prediction set of floor tile (FTPS.2) are located more externally in PC1xPC3 (Fig. 11a) and PC3xPC2 (Fig. 11b) when projected on the PCA model built with the FTTS.2 floor samples. Figure 11c shows that the residues from the spectra of the CE5 and PO4 floor tiles. As can be seen in the residuals graph both have high values of Q-Residuals,

In a similar way, the spectra of CE4, CE6, and PO5 were submitted to the floor tile evaluation protocol built with the first training set of floor tiles (FTTS.1), and the response obtained was that all floor tiles were similar to the training set. A strategy to increase reliability in the use of the hierarchical model for the identification of human blood is to introduce this protocol as a previous step to evaluate if the floor tile where the trace was found has the variability represented by the hierarchical model built to identify human blood.

Fig. 11. PCA built with TS2 floor tile data a) PC1xPC3; b) PC2xPC3; c) Residues plot.

4. Conclusion

Hierarchical models by fusion of different chemometric techniques represent a significant advance for the development of a robust and confirmatory methodology to identify human blood stains at crime scenes, providing minimal interference with the evidence and using
an objective and technical approach. The hierarchical models constructed from training sets
with a more comprehensive variability (TS1 and TS3) and using a PLS-DA model as the second
decision rule to classify human blood presented the best results of sensitivity and specificity,
both equal to one. The same values were obtained by the models built for the sets with a reduced
spectral range. It should be noted, however, that the PLS-DA model is a discriminant model; to
create the boundary of the human blood class, it uses the information from animal blood. For
more accurate use of PLS-DA, it would be important to create a more complete database of
animal blood. More samples of blood from different species of animals would provide a wider
data base against which future models could be built.

It is relevant to highlight that the hierarchical models built using PCA or SIMCA to
discriminate between human and animal blood samples, also obtained excellent sensitivity
results for human blood classification, with values between 0.94, 0.97 and 1. Unlike the PLS-
DA, these models used only information from the class of interest, which in this case was the
HB.

In a forensic context, when a possible trace of blood is found, it is important to identify
and collect the stain for further tests, even if it is animal blood. In the case of our study, the
specificity for human blood identification in relation to animal blood was not a crucial problem,
since all models obtained 100% correct classification for the common false positive samples.
A protocol for prior evaluation of the substrate was also proposed, in order to guarantee that the
substrate where the stain is deposited is well represented by the training set. This methodology
proved to be quite effective and enabled the analyst to assess the reliability of the result provided
by the model built according to the representativeness of the type of floor tile on which the trace
was deposited.
Acknowledgements

The authors would like to acknowledge the Núcleo de Química Analítica Avançada de Pernambuco - NUQAAPE (FACEPE grants APQ-0346-1.06/14), Núcleo de Estudos em Química Forense - NEQUIFOR (CAPES grants AUXPE 3509/2014), CNPq (grants 428891/2018-7), Instituto Nacional de Tecnologias Analíticas Avançadas - INCTAA (CNPq grants 573894/2008-6 and 465768/2014-8 and FAPESP grants 2008/57808-1 and 2014/50951-4).

The English text of this paper has been revised by Sidney Pratt, Canadian, MAT (The Johns Hopkins University), RSAdip - TESL (Cambridge University).

REFERENCES


A.V. Morillas, J. Gooch, N. Frascione, Feasibility of a handheld near infrared device for


