

Diffuse Reflectance FTIR of Latent Fingerprints and Discriminant Analysis for Sex Identification in Humans

Marco Antonio de Souza,^{1b} ^{a,b,c} Alexandre S. Santos,^d Sebastião William da Silva,^d
Jez Willian B. Braga^{1b} ^{*,b,e} and Marcelo Henrique Sousa^c

^aPolícia Federal, 70770-917 Brasília-DF, Brazil

^bInstituto de Química, Universidade de Brasília, CP 4478, 70910-900 Brasília-DF, Brazil

^cGrupo de Nanotecnologia Verde, Faculdade de Ceilândia, Universidade de Brasília,
72220-140 Brasília-DF, Brazil

^dLaboratório de Espectroscopia Óptica, Instituto de Física, Universidade de Brasília,
70910-900 Brasília-DF, Brazil

^eInstituto Nacional de Ciência e Tecnologia em Bioanalítica (INCT Bio),
13083-970 Campinas-SP, Brazil

Latent fingerprint is an important crime scene evidence, but it is not always recoverable or technically suitable for analysis with fingerprint patterns. Forensic science has shown that other information can be explored from traces using chemical compounds. Infrared spectroscopy is a nondestructive technique that is widely applied to a variety of forensic evidence. In this work, infrared spectroscopy and partial least square discriminant analysis were used to determine the human sex based on latent fingerprint analysis. Fingerprint samples were taken from 42 male and female donors, then kept in either dark or light storage conditions, and the Fourier transform infrared (FTIR) spectra were measured considering a period of up to 30 days from collection. The regions from 3000 to 2800 cm^{-1} and 1790 to 1150 cm^{-1} presented the greatest differences in the peak intensities among the two sex groups. The results showed a correct discrimination rate higher than 80%.

Keywords: sex determination, latent fingerprints, fingerprints, chemometric analysis, FTIR

Introduction

Fingerprint analysis has always been of great importance in establishing the authorship of a crime. It is one of the most common traces found at a crime scene and, in this case, it is known as a latent fingerprint.^{1,2} Traditional techniques for print development, such as the use of cyanoacrylate, ninhydrin and other chemical developers, always work towards producing a better contrast between the fingerprint and the background where it was deposited.¹ Thus, the number of details and their characteristics that appear after their development individualize the fingerprint and enable the comparison with fingerprint patterns to be carried out, but the number of minutiae is often not enough to carry out a comparison.³

Since the 1990s, there have been a number of studies to extract other information from fingerprints, such as differentiating adults from children,⁴ finding traces of illicit substances and explosives,^{5,6} dating^{7,8} and, more recently, the discrimination of male and female subjects.^{3,9,10}

The understanding of the chemical substances that form fingerprints is essential for these studies, since it is from the understanding of the degradation kinetics, concentration and verification of the existence of a given substance that these new results can be achieved. The chemical composition of fingerprints originates mainly from the eccrine and sebaceous glands. Eccrine sweat is mainly composed of water (98%), and the rest of its content is either organic material (e.g., proteins, amino acids, and lactate) or inorganic (e.g., Na^+ , K^+ , Cl^- and other metal ions). Similarly, squalene, cholesterol, glycerides, fatty acids, and a range of lipid esters principally made up the sebaceous secretions. Contaminants detected in these

*e-mail: jez@unb.br

Editor handled this article: Ivo M. Raimundo Jr. (Associate)

substances also include cosmetics, hair-care products and medications.¹¹

Recent studies involving vibrational spectroscopy have shown that it is feasible to determine the sex of an individual for forensic purposes, using other traces that can be found at a crime scene. In our previous work,⁹ to determine sex, fingerprints from males and females were taken, kept under either dark or light conditions and then studied seven days later. The analysis used Raman spectroscopy with partial least square discriminant analysis (PLS-DA) and support vector machine discriminant analysis (SVM-DA), and the results of discrimination were approximately 80-93%. Huynh *et al.*³ carried out a biocatalytic method to identify sex using amino acids extracted from the fingerprint using ultraviolet and visible (UV-Vis) spectroscopy. In other studies to determine the sex of human subjects, Widjaja *et al.*¹² used Raman spectroscopy to analyze nail clippings, and Lednev and co-workers¹³ used Raman spectroscopy to analyze saliva samples.

Recently, Sharma *et al.*¹⁴ used attenuated total reflectance Fourier transform infrared (ATR-FTIR) with principal component analysis (PCA) and PLS-DA methods to classify and predict the sex of male and female nail clippings. The classification rate of the normalized derived data was found to be 100 and 90% for women and men, respectively. Another study¹⁵ demonstrated that it is possible to establish statistically significant differences between male and female groups when analyzing the absorption bands of proteins and lipids in saliva samples by infrared spectrometry. The authors demonstrated that the absorbance of the bands attributed to proteins and nucleic acids is greater for men, while the absorbance of the bands attributed to lipids is greater for women.

As already mentioned, fingerprints can provide other information besides authorship. Considering the advances in Fourier transform infrared (FTIR) and chemometric methods, the aim of this work is to evaluate the use of FTIR spectroscopy and the supervised method PLS-DA for discriminating the sex of human subjects based on latent fingerprint analysis. FTIR spectroscopy has not been previously reported to identify sex in human subjects using latent fingerprints. In addition, the low cost and speed of the analyses, as well as the preservation of the sample, can be pinpointed as advantages of using this method.

Experimental

Ethics committee

The Ethics and Human Research Committee of the Faculty of Health Sciences of the University of Brasília

approved this research (protocol 42304220.0.0000.0030), following resolution 466/12 of the National Health Council (CNS).

Samples acquisition and data measurements

To evaluate the feasibility of this technique for sex discrimination, this first study was conducted adopting a standard procedure followed by each donor. Firstly, the donors were instructed not to use cosmetics for a period of 24 h before the fingerprint collection procedure. To carry out the collection, each donor needed to wash his/her hands with neutral liquid soap, rinsing until the soap was completely removed and waiting 10 min for the hands to dry without touching anything. To produce a sebum-rich fingerprint, the donor was required to press the right thumb on the forehead for 3 s and thus to produce two fingerprints. Each fingerprint was produced on a glass slide, one square inch in size, covered with aluminum foil. FTIR measurements were performed with all samples on the day of collection (D0), 7 days later (D7) and after 30 days (D30). After the D0 measurements, the two fingerprints collected on D0 from each donor were divided into two groups, where the first group was stored under light conditions and the other group under dark conditions.

The fingerprint samples were obtained from 21 women (17 Caucasian and 4 Black) and 21 men (17 Caucasian and 4 Black), aged between 25 and 65 years old. Considering that two fingerprints were obtained from each donor, which were analyzed at three different time periods after collection, a total of 252 standard samples were obtained. For different regions of each fingerprint, three spectra were acquired at the previously determined time periods (D0, D7 and D30). The average of these three spectra was calculated and used for further chemometric analysis. The spectra measurements were performed in the Bruker Vertex 70 equipment using the 40° angle reflectance method with the manual reflection unit for tensor (model A513/Q, Ettlingen, Germany), with 64 s and 64 scans, in the spectral region from 400 to 4000 cm⁻¹. Due to problems observed in some spectra during acquisition, 246 samples were selected from the 252 standard samples.

Chemometric analysis

PLS-DA was carried out using appropriate functions from the PLS-toolbox (version 8.81, Eigenvector, Wenatchee, WA)¹⁶ and MATLAB R2020b (The Mathworks Inc., Natick, MA)¹⁷ or data modeling and multivariate analysis. The average of the spectra of each sample were used and all conditions (different days and light exposure) were considered.

FTIR spectra from female and male samples were arranged in a single matrix $\mathbf{X}(m,n)$, where m is the number of training spectra and n the number of wavenumbers (cm^{-1}). The PLS-DA algorithm correlates the spectral data \mathbf{X} with a $\mathbf{y}(n,1)$ vector that contains the information if the sample came from the class/sex female ($y = 1$) or male ($y = 0$). Thus, the female sample constituted a true positive result ($y = 1$), while the male was considered a true negative ($y = 0$). PLS-DA had performed a binary discrimination through the use of the distributions of the class values (\hat{y}) predicted for both classes in the training step; this made it possible to keep at a minimum level the number of false positive and negative errors, in consonance with the Bayes theory.¹⁸ Further details of the PLS-DA algorithm can be obtained in specific references.^{19,20}

An outlier is a result of a sample presenting distinct characteristics from the other spectra of the same class or even the entire training set. These abnormal spectra can occur due to changes in the chemical composition of the fingerprint or the instrumental measurements. The spectra measured for both sexes, considering D0, D7 and D30, were joined in the same matrix and the outliers were excluded based on Hotelling T^2 and Q residuals of a PCA modeling using cross validation, with a significance level of 0.01. Next, the remaining spectra were divided using the Kennard-Stone algorithm; 33% of the samples were chosen for the test set, while the other 67% were attributed to the training set, following the standard settings of PLS-Toolbox and previous studies undertaken by this research group.²⁰⁻²²

The Kennard-Stone algorithm starts by selecting the two most dissimilar samples using the Euclidean distance. In each following iteration, the algorithm singles out the sample showing the greatest distance from the other samples already selected. This procedure occurs repeatedly until the acquisition of the selected number of training samples.²³

Before discrimination modeling took place, the preprocessing of spectra was performed. An evaluation of preprocessing methods, such as smoothing, normalizing, derivative and mean center, enabled us to obtain the best performance using the model optimizer tool, available in the PLS-Toolbox.¹⁶ The preprocessing method was chosen after calculation of the root mean square error of cross validation (RMSECV).

The choice for the latent variables for PLS-DA modeling were obtained by venetian blinds cross-validation, using the lowest value of the RMSECV or after this parameter reaches a plateau. The discrimination models calculated using the different preprocessing methods and latent variables were ranked as false negative,

false positive and efficiency rate, calculated according to equations 1 to 3.^{19,24}

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \times 100 \quad (1)$$

$$\text{FNR} = \frac{\text{FN}}{\text{FN} + \text{TN}} \times 100 \quad (2)$$

$$\text{EFR} = \sqrt{\frac{\text{TN} \times \text{TP}}{(\text{TN} + \text{FP}) \times (\text{TP} + \text{FN})}} \times 100 \quad (3)$$

where FN is the number of samples predicted as false negative, FP as false positive, TN is the number of samples predicted as true negative, TP as true positive. FNR is the false negative rate, FPR is the false positive rate and EFR is efficiency rate. Sensitivity (SEN) and specificity (SPEC) values were determined by 100 minus the respective values of FNR or FPR.

Results and Discussion

The same samples evaluated in this study was used in our previous work,⁹ applying Raman microspectroscopy and a different chemometric approach. However, as the FTIR spectra used in this study were obtained with a conventional reflectance accessory, the method development was more challenging since most of the sample surfaces were not covered by the fingerprint. For this reason, the average of the replicates was calculated for each sample. In addition, since no apparent signal or tendency was observed related to the time or storage conditions, all the conditions were considered in an attempt to develop a more robust PLS-DA model.

The FTIR spectra used for the model training, after baseline correction and normalization, are presented in Figure 1a. Even after preprocessing, these spectra reveal significant noise and intensity variation at the beginning (500 to 400 cm^{-1}) and at end of the middle infrared (IR) region (4000 to 3500 cm^{-1}). Nevertheless, some differences in the peak intensities between the two sex groups can be seen from 3000 to 2800 and 1790 to 1150 cm^{-1} , which are more visible on the average spectra presented in Figure 1b.

The FTIR spectra of the fingerprint in Figure 1 present features suggestive of lipid, carotenoid, and protein bands. The spectral regions 1000-1850 and 2700-3600 cm^{-1} were the most informative and are attributed to molecular vibrations of eccrine and sebaceous material (Table 1). The hydrocarbon chains are found at 3000 cm^{-1} , which correspond to the C-H stretching mode. At 1739 cm^{-1} there is an absorption band corresponding to a carbonyl stretching mode and the

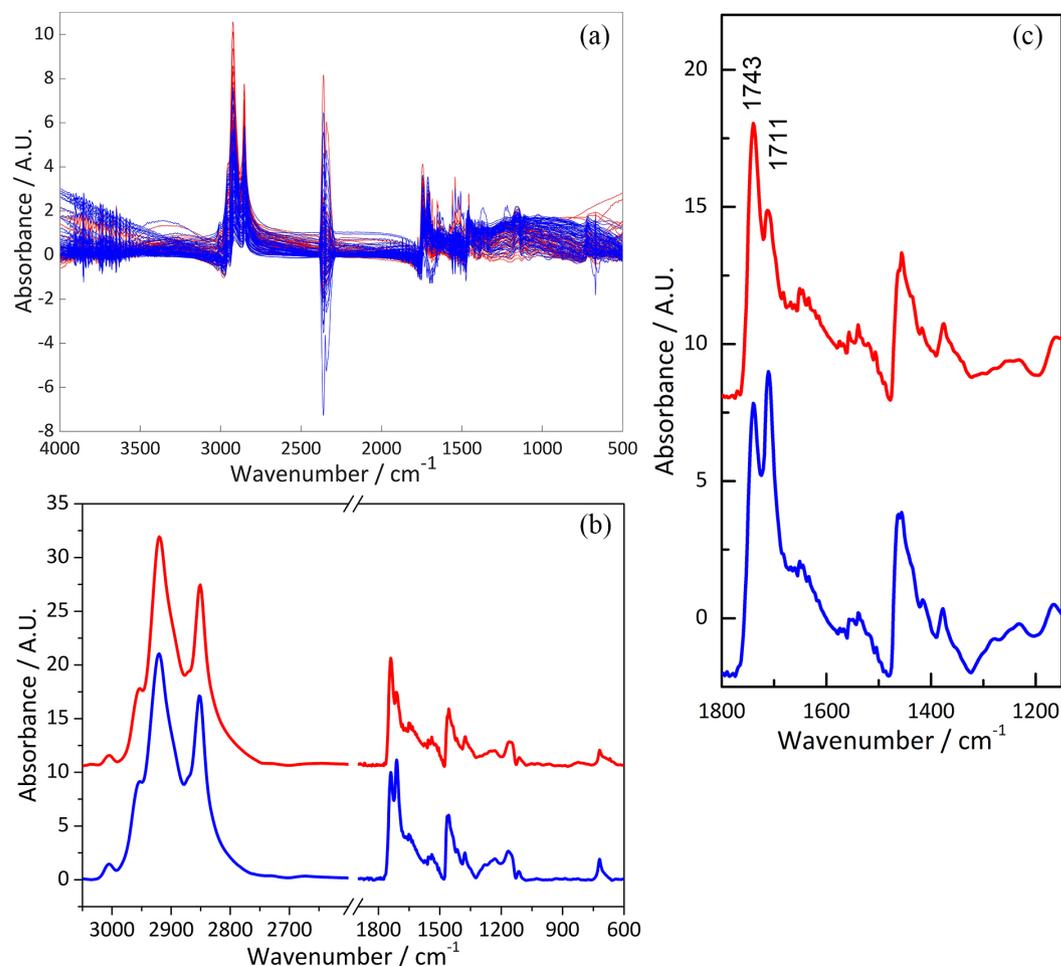


Figure 1. (a) Preprocessed reflectance FTIR spectra of latent fingerprint samples used in the training set. (b) Average spectrum of latent fingerprints of female (red) and male (blue) classes. (c) Close view of female and male mean spectra from 1150 to 1800 cm^{-1} , where the main spectral difference between them (1711 and 1743 cm^{-1}) is highlighted.

shoulder at 1711 cm^{-1} is attributed to a second carbonyl stretching; these suggest the presence of triglycerides and/or phospholipids and fatty acids, respectively. Additionally, bands related with proteins (secondary amide) can be observed at 1645, 1536 and 1229 cm^{-1} .²⁵⁻²⁷

The mean spectral measurements suggest that subtle FTIR spectral differences, which specifically relate to determining sex, are present. These signals are predominantly seen in the wavenumbers at 1739 and 1711 cm^{-1} . The changes noted in the intensities of these absorption signals may arise from the chemical, biological or physical processes that take place in fingerprints from its deposition until the carrying out of the FTIR measures. The initial composition of fingerprints changes by processes including degradation, drying, oxidation or polymerization. Some studies^{3,28} have pinpointed differences in the chemical composition of male and female fingerprints, especially related to the fatty acid content.

In some criminal cases, the latent fingerprint is not collected on the same day as the crime took place, as

Table 1. Major vibrational bands obtained from a particle in fingerprint deposit of an adult female²⁵⁻²⁷

Band / cm^{-1}	Vibration
2915	CH_3 symmetric stretch (long aliphatic C chains)
2847	CH_2 symmetric stretch (long aliphatic C chains)
1739	C=O stretch (saturated esters)
1711	C=O stretch (fatty acids)
1645	C=O stretch (secondary amide)
1536	N-H in-plane bend and C-N stretch (secondary amide)
1451	CH_3 asymmetric bend and CH_2 symmetric bend (aliphatic C chains)
1375	CH_3 symmetric bend (aliphatic C chains)
1229	C-N stretch (secondary amide)
1152	C-C-O stretch (saturated ester)
1108	O-C-C stretch (saturated ester)

mentioned above. Therefore, the samples used to build the models were collected from 42 donors and the spectra were acquired on the day of the fingerprint deposition, and

at seven and thirty days later. After the model calculations using the optimizer tool, Baseline (Automatic Weighted Least Squares, order = 5), Normalize (1-Norm, Area = 1) and Mean Centering were the methods chosen to develop the model.

The initial data analysis to identify extreme outliers was performed with PCA models developed in each class using the preprocessed data and the entire FTIR spectra. After outlier exclusion by T^2 and Q residuals, the training set for model development used 157 spectra from females and males, while 62 spectra were used for the test set. Then, a new model was calculated with this new training set using the entire spectral region, and the vector importance projection (VIP) scores were used for variable selection (Figure 2). The VIP scores show that the regions from 3000 to 2800 and 1790 to 1150 cm^{-1} seem to be the most suitable for sex estimation, which agrees with most of the bands highlighted in Figure 1 and Table 1. Three independent models were developed with each region and the combination of them (model 1: 3000 to 2800 cm^{-1} ; model 2: 1790 to 1150 cm^{-1} and model 3: 3000 to 2800 and 1790 to 1150 cm^{-1}), wherein the region 1790 to 1150 cm^{-1} occurs, highlighted in Figure 2, presented the best results.

The results for both training and test set for the estimation of class values, after outlier exclusion and variable selection, with the PLS-DA model developed using 14 latent variables, are shown in Figure 3. The high number of latent variables may be a result of the different sources of variation in the data (differences between the individuals of the same sex, the sample conditions and in the instrumental measurements). A few spectra still showed values of T^2 and Q residuals that were higher than the 0.99 confidence limits (Figure 3a). Nevertheless, since only one exclusion step was performed during the model development and validation,

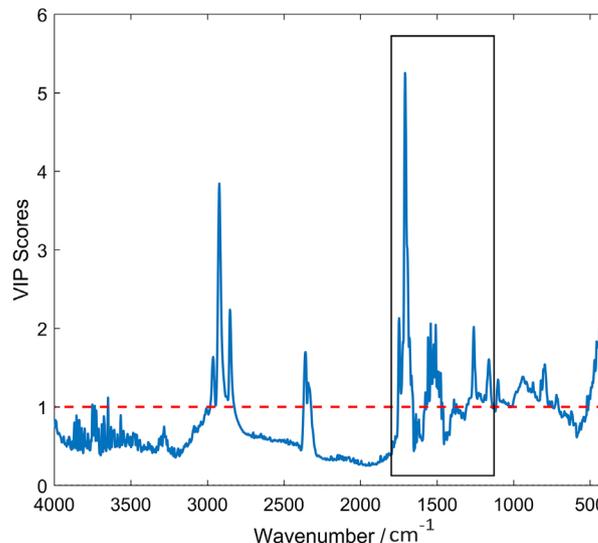


Figure 2. Vector importance projection (VIP) scores for the PLS-DA model developed with 14 latent variables and the full FTIR spectra. The delimited area is the region of the spectrum used for the PLS-DA model.

these samples were not removed from the datasets. The decision to use only one outlier exclusion step arose from the restricted number of samples available for this work. In our opinion, the outlier samples present in Figure 3a may be a reflection of the small number of samples used to model the differences between light exposure and acquisition on different days. Most of the samples excluded in the FTIR data were different from the ones excluded in our previous study using Raman microspectroscopy, which may be a result of the significant differences in the experimental measurements and data analysis.⁹

The dispersion of the class values acquired using the PLS-DA model is shown in Figure 3b. These values showed notable variation, illustrating the difficulty of sex identification using latent fingerprints. However, the

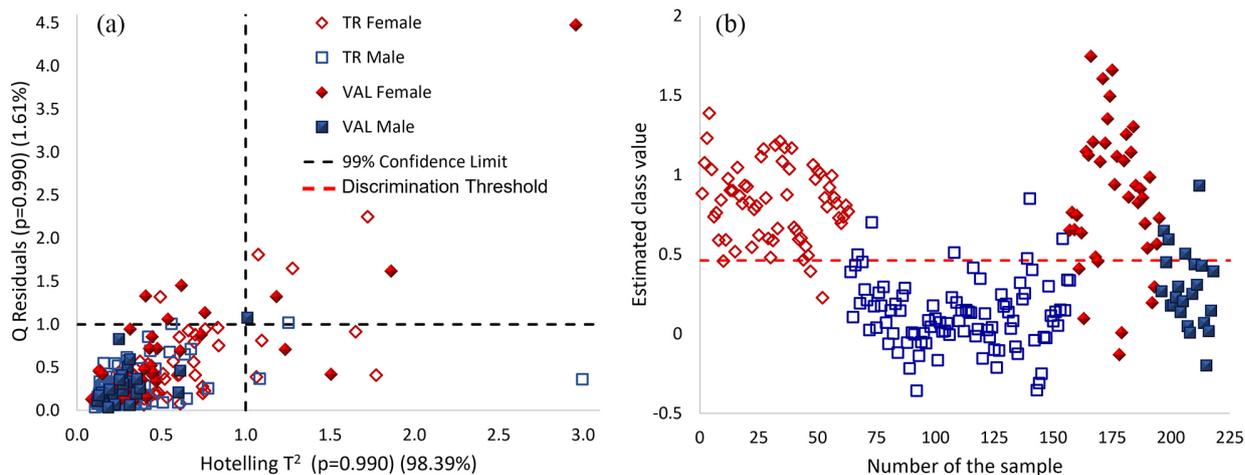


Figure 3. (a) Distribution of the Hotelling T^2 versus Q residuals and (b) predicted class for the training and test sets using the PLS-DA model developed in the 1494-1768 cm^{-1} region. (---) 99% confidence limits, (\diamond) training female (TR female), (\square) training male (TR male), (\blacklozenge) test female (VAL female), (\blacksquare) test male (VAL male).

performance of the method can be better judged using its figures of merit (Table 2). The SEN and SPEC values showed that the model presented approximately the same discrimination rate for female and male samples, wherein approximately 95 and 82% were obtained for the training and test sets, respectively. The EFR encompasses the contribution of the other validation parameters, being the average/global parameter to judge the overall performance of the method. Although the estimated class values were significantly spread, good EFRs were acquired. The EFR obtained in the test set was 82.3%. Greater variability of data was observed in female samples, but it was not possible to establish any specific cause. However, hormonal variations and application of cosmetics should be considered in complementary follow-ups to this study. As in our previous study, SVM-DA models were also carried out in the FTIR data. However, the efficiency rate of these models was lower than the one observed with PLS-DA (< 60%). Therefore, these results were not included in this work.

Table 2. Estimated figures of merit for the proposed method

PLS-DA (14)	FTIR	
	TR	VAL
FN	3	7
FP	3	4
TP	60	32
TN	91	19
SEN/TPR	95.2	82.1
SPEC/TNR	96.8	82.6
FNR	4.8	17.9
FPR	3.2	17.4
EFR	96.0	82.3

The number of latent variables of the PLS-DA model is indicated in parentheses. PLS-DA: partial least square discriminant analysis; FTIR: Fourier transform infrared spectroscopy; TR: training; VAL: validation; FN: false negative; FP: false positive; TP: true positive; TN: true negative; SEN: sensibility; TPR: true positive rate; TNR: true negative rate; SPEC: specificity; FNR: false negative rate; FPR: false positive rate; EFR: efficiency rate.

It is important to highlight that this study aimed to evaluate the use of fingerprints as a tool for determining the sex of the subject. From the forensic point of view, this information is important, and it can be part of the first step in revealing the identity of the subject, especially when the latent fingerprint is not suitable for fingerprint identification, either due to poor deposition or a lack of database matches.

Conclusions

Fingerprints are traces of great importance to forensic

science, as they contribute to the identification of the author or even the victim of a crime. However, these traces collected at a crime scene sometimes lack the minutiae to be compared to a standard fingerprint. In this context, the chemical information extracted by instrumental analysis methods can help to reduce the list of suspects. This study moves in that direction and presents a methodology for classifying the sex of individuals based on latent fingerprints. Within a period of up to 30 days from the acquisition of the fingerprint, correct discrimination of the sex of subjects were obtained at a rate of over 80%, indicating that the method can determine the sex of the subject with a high correct identification rate.

The number of samples used in the training set of this study was relatively small, but despite this limitation, it was possible to confirm the possibility of determining the sex of the subject by means of the chemical information found in FTIR spectra from latent fingerprints, providing valuable forensic information.

In addition, other challenges arise related to class discrimination based on latent fingerprint analysis, such as ethnicity and age groups, which can still be explored with the FTIR technique, resulting in new methodologies for forensic applications.

Acknowledgments

The authors thank INCTBio (CNPq grant 465389/2014-7), CNPq process 304870/2018-8, CAPES, FAPDF and Brazilian Federal Police.

References

- Amin, M. O.; Al-Hetlani, E.; Lednev, I. K.; *TrAC, Trends Anal. Chem.* **2021**, *143*, 116341. [Crossref]
- INTERPOL, <https://www.interpol.int/How-we-work/Forensics/Fingerprints>, accessed in November 2022.
- Huynh, C.; Brunelle, E.; Halámková, L.; Agudelo, J.; Halánek, J.; *Anal. Chem.* **2015**, *87*, 11531. [Crossref]
- DeLorenzo, R.; Kimbrough, D. R.; *J. Chem. Educ.* **1998**, *75*, 1300. [Crossref]
- Yang, T.; Guo, X.; Wang, H.; Fu, S.; wen, Y.; Yang, H.; *Biosens. Bioelectron.* **2015**, *68*, 350. [Crossref]
- López-López, M.; García-Ruiz, C.; *TrAC, Trends Anal. Chem.* **2014**, *54*, 36. [Crossref]
- Andersson, P. O.; Lejon, C.; Mikaelsson, T.; Landström, L.; *ChemistryOpen* **2017**, *6*, 706. [Crossref]
- González, M.; Mariotti, K. C.; Gomes, A. A.; Ferrão, M. F.; Limberger, R. P.; *Braz. J. Anal. Chem.* **2020**, *8*, 141. [Crossref]
- Souza, M. A.; Santos, A. S.; da Silva, S. W.; Braga, J. W. B.; Sousa, M. H.; *Forensic Chem.* **2022**, *27*, 100395. [Crossref]

10. Carvalho, D. S.; Alecrim, M. M. V.; de Sousa Jr., R. T.; Ribeiro Jr., L. A.; *Sci. Justice* **2022**, *62*, 181. [Crossref]
11. Bleay, S. M.; Bailey, M. J.; Croxton, R. S.; Francese, S.; *WIREs Forensic Sci.* **2021**, *3*, 1. [Crossref]
12. Widjaja, E.; Lim, G. H.; An, A.; *Analyst* **2008**, *133*, 493. [Crossref]
13. Muro, C. K.; Fernandes, L. S.; Lednev, I. K.; *Anal. Chem.* **2016**, *88*, 12489. [Crossref]
14. Sharma, A.; Verma, R.; Kumar, R.; Chauhan, R.; Sharma, V.; *Microchem. J.* **2020**, *159*, 105504. [Crossref]
15. Bel'skaya, L. V.; Sarf, E. A.; Solomatin, D. V.; *Appl. Spectrosc.* **2020**, *74*, 536. [Crossref]
16. *PLS_Toolbox 8.8.1*, Eigenvector Research, Inc., M.; Wenatchee, USA, 2020.
17. *Matlab*[®], version 9.9 (R2020b); The MathWorks Inc.; Natick, MA, USA, 2020.
18. Brereton, R. G.; Lloyd, G. R.; *J. Chemom.* **2014**, *28*, 213. [Crossref]
19. Botelho, B. G.; Reis, N.; Oliveira, L. S.; Sena, M. M.; *Food Chem.* **2015**, *181*, 31. [Crossref]
20. Custódio, M. F.; Magalhães, L. O.; Arantes, L. C.; Braga, J. W. B.; *J. Braz. Chem. Soc.* **2021**, *32*, 513. [Crossref]
21. Magalhães, L. O.; Arantes, L. C.; Braga, J. W. B.; *Microchem. J.* **2019**, *144*, 151. [Crossref]
22. Snel, F. A.; Braga, J. W. B.; da Silva, D.; Wiedenhoeft, A. C.; Costa, A.; Soares, R.; Coradin, V. T. R.; Pastore, T. C. M.; *Wood Sci. Technol.* **2018**, *52*, 1411. [Crossref]
23. Westad, F.; Marini, F.; *Anal. Chim. Acta* **2015**, *893*, 14. [Crossref]
24. Oliveri, P.; Downey, G.; *TrAC, Trends Anal. Chem.* **2012**, *35*, 74. [Crossref]
25. Williams, D. K.; Schwartz, R. L.; Bartick, E. G.; *Appl. Spectrosc.* **2004**, *58*, 313. [Crossref]
26. Ricci, C.; Phiriyavityopas, P.; Curum, N.; Chan, K. L. A.; Jickells, S.; Kazarian, S. G.; *Appl. Spectrosc.* **2007**, *61*, 514. [Crossref]
27. Dorakumbura, B. N.; Boseley, R. E.; Becker, T.; Martin, D. E.; Richter, A.; Tobin, M. J.; van Bronswijk, W.; Vongsvivut, J.; Hackett, M. J.; Lewis, S. W.; *Analyst* **2018**, *143*, 4027. [Crossref]
28. van Dam, A.; van Beek, F. T.; Aalders, M. C. G.; van Leeuwen, T. G.; Lambrechts, S. A. G.; *Sci. Justice* **2016**, *56*, 143. [Crossref]

Submitted: June 21, 2022

Published online: December 2, 2022

