Non-destructive age estimation of biological fluid stains: An integrated analytical strategy based on near-infrared hyperspectral imaging and multivariate regression

Talanta

Non-destructive age estimation of biological fluid stains: an integrated analytical strategy based on near-infrared hyperspectral imaging and multivariate regression

--Manuscript Draft--

Manuscript Number:	
Article Type:	Research Paper
Section/Category:	Molecular Spectroscopy and Optical Probes
Keywords:	near infrared (NIR) spectroscopy; hyperspectral imaging (HSI); partial least squares (PLS) regression; biological fluid; aging estimation; forensic analysis
Corresponding Author:	Paolo Oliveri, PhD University of Genova Genoa, ITALY
First Author:	Cristina MANIS, MSc
Order of Authors:	Cristina MANIS, MSc
	Cristina Malegori, PhD
	Eugenio ALLADIO, PhD
	Marco VINCENTI, PhD
	Paolo GAROFANO, PhD
	Filippo BARNI, PhD
	Andrea BERTI, PhD
	Paolo Oliveri, PhD
Abstract:	From a criminalistic point of view, the accurate dating of biological traces found at the crime scene, together with its compatibility with the estimated crime perpetration timeframe, enables to limit the number of suspects by assessing their alibis and clarifying the sequence of events. The present study delineates, for the first time, the possibility of dating biological fluids such as semen and urine, as well as blood traces, by using a novel non-destructive analytical strategy based on hyperspectral imaging in the near infared region (HSI-NIR), coupled with multivariate regression methods. Investigated aspects of the present study include not only the progressive degradation of the biological trace itself, but also the effects of its interactions with the support on which it is absorbed, in particular the hydrophilic vs. hydrophobic character of fabric tissues. Results are critically discussed, highlighting potential and limitations of the proposed approach for a practical implementation.
Suggested Reviewers:	Markus Baumgartner, PhD Prof., University of Zurich: Universitat Zurich Markus.Baumgartner@irm.uzh.ch Expert on Forensic Analytical Chemistry and spectral imaging
	Simon Lewis, PhD Prof., Curtin University - Perth Bentley Campus: Curtin University S.Lewis@curtin.edu.au Expert on Forensic Analytical Chemistry and spectral imaging
	Grzegorz Zadora, PhD Prof., University of Silesia: Uniwersytet Slaski w Katowicach greg_zadora@yahoo.com Expert on Forensic Analytical Chemistry and Statistical Analysis in Forensic Sciences
	Gerda Edelman, PhD Dr., Netherlands Forensic Institute: Nederlands Forensisch Instituut g.edelman@nfi.minvenj.nl Expert on Crime Scene investigations, spectral imaging and multivariate data analysis
Opposed Reviewers:	

Università di Genova | DIFAR DIPARTIMENTO DI FARMACIA

Genova, March 2nd, 2022

Dear Editor,

We are pleased to submit the manuscript entitled "**Non-destructive age estimation of biological fluid stains: an integrated analytical strategy based on near-infrared hyperspectral imaging and multivariate regression**", for your consideration for publication in *Talanta*.

The study arises from a collaboration between two academic research groups (at the universities of Genova and Turin, respectively), the Regional Antidoping Center & Regional Toxicology Laboratory "A. Bertinaria" (Orbassano), and the Scientific Investigation Section (Reparto Investigazioni Scientifiche) of Carabinieri (Rome), and represents the evolution of a former study published in *Talanta* (C. Malegori, E. Alladio, P. Oliveri, C. Manis, M. Vincenti, P. Garofano, F. Barni, A. Berti, *Identification of invisible biological traces in forensic evidences by hyperspectral NIR imaging combined with chemometrics*, *Talanta*, Volume 215, 1 August 2020, 120911, https://doi.org/10.1016/j.talanta.2020.120911). The previous study was devoted to identification of biological stains, while the current one is focused on the estimation of time elapsed between the deposition of the biological fluid and its analysis.

We confirm that this manuscript is an original work; it has not been published elsewhere and is not under consideration by any other journal. We have no conflicts of interest to declare. As the Corresponding Authors, we confirm that the manuscript has been read and approved for submission by all the named authors.

We ensure our collaboration to further improve the manuscript, following the guidance and comments of the Reviewers that you might choose.

Yours sincerely,

Dr. Paolo Oliveri
Professor, Analytical Chemistry and Chemometrics
Vice-President
Ligurian Section of the Italian Chemical Society (SCI)
Coordinator
Divisional Group of Chemometrics
Department of Pharmacy (DIFAR), University of Genova, Italy
+39 010 353-2626; oliveri@difar.unige.it

Dr. Marco Vincenti
Full Professor, Analytical Chemistry
Scientific Director
Regional Antidoping Center & Regional Toxicology Laboratory "A. Bertinaria", Orbassano, Italy
Coordinator
Divisional Group of Forensic Analytical Chemistry
of the Italian Chemical Society (SCI)
Department of Chemistry, University of Turin, Italy
+39 011-670-5264; marco.vincenti@unito.it

Università di Genova | DIFAR DIPARTIMENTO DI FARMACIA

Novelty Statement:

A novel analytical strategy based on a cutting-edge technique (hyperspectral imaging in the near-infrared region) is presented, for the first time, as a powerful tool to simultaneously reveal, localize and date biological stains from different fluids (blood, semen, urine) on different supports (hydrophilic and hydrophobic fabrics). Time was introduced as a further dimension in hyperspectral imaging and, thanks to the development of appropriate multivariate regression models, it was possible to efficiently estimate the time (in h) elapsed between the deposition of the biological fluid and analysis. This was possible thanks to the capability of near-infrared spectroscopy to capture chemical modifications that occur during the ageing. Informative chemical maps were obtained, allowing the analyst to visualize the stain and to have a direct estimation of its age, in a completely non-destructive and non-invasive way. Prediction models were fully validated on separate test sets and the high-impact results are critically discussed.

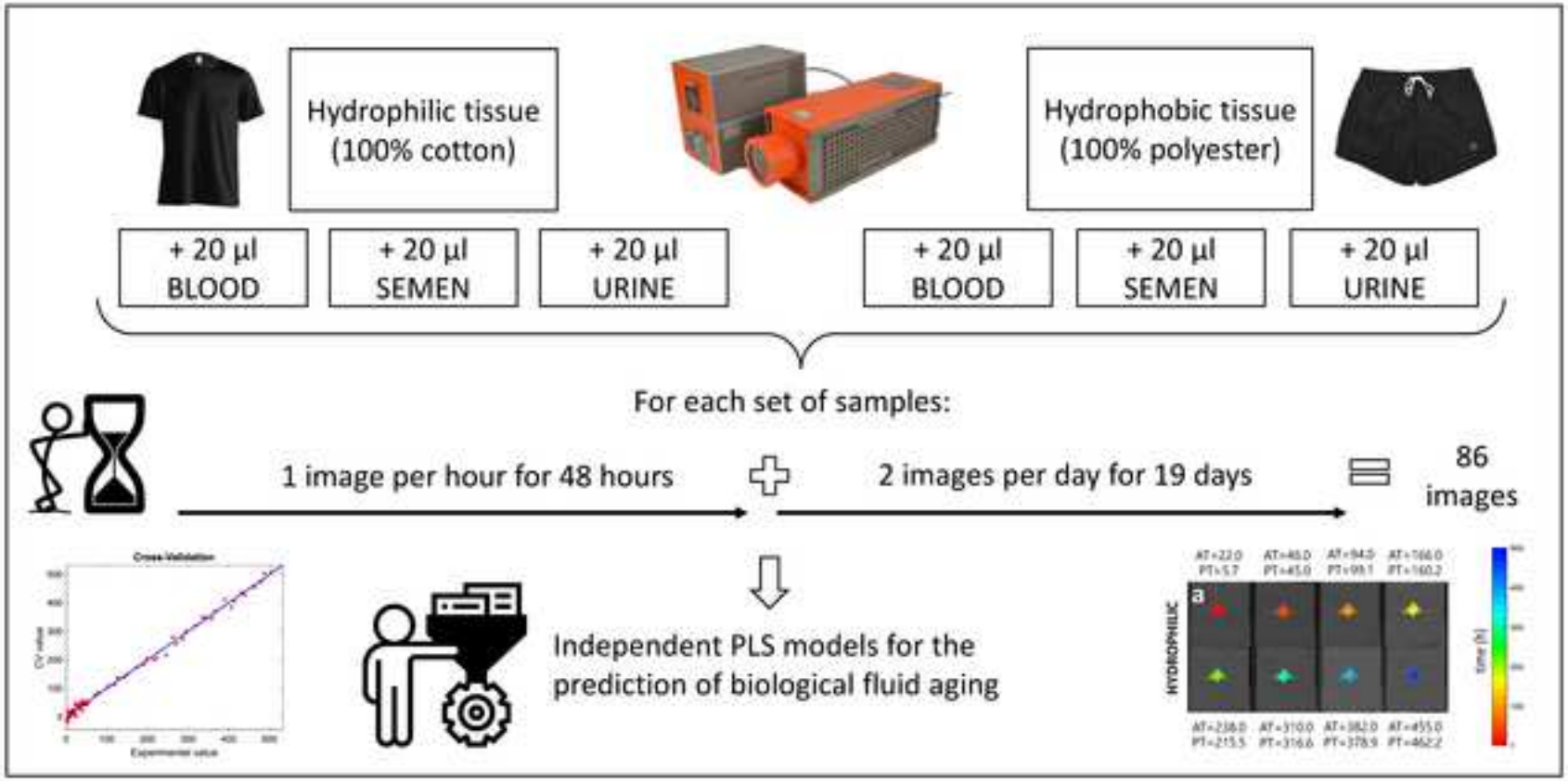
The innovative analytical solution presented in this study has a great impact not only in the forensic sector, but in the whole frame of analytical fields in which the ageing of samples is a key issue (from biological and clinical chemistry to shelf-life studies in the food sciences).

Dipartimento di Farmacia
difar.unige.it
P. IVA 00754150100

Viale Cembrano 4, 16148 Genova
oliveri@difar.unige.it

Highlights

- Development of a novel non-destructive analytical method for dating biological fluids
- Blood, semen and urine analysed on both hydrophilic and hydrophobic fabric tissues
- Near-infrared hyperspectral imaging (HSI-NIR) integrated with multivariate regression
- Degradation of biological traces and interactions with the support investigated
- Critical discussion of the proposed approach in view of a practical implementation



Non-destructive age estimation of biological fluid stains: an integrated analytical strategy based on near-infrared hyperspectral imaging and multivariate regression

Cristina MANIS[a], Cristina MALEGORI[b], Eugenio ALLADIO[a,c], Marco VINCENTI[a,c,*], Paolo GAROFANO[c], Filippo BARNI[d], Andrea BERTI[d], Paolo OLIVERI[b,*]

[a] Dipartimento di Chimica, Università degli Studi di Torino – Torino, Italy

[b] Dipartimento di Farmacia (DIFAR), Università degli Studi di Genova – Genova, Italy

[c] Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Laboratorio di Genetica Forense – Orbassano, Italy

[d] Reparto Carabinieri Investigazioni Scientifiche, Sezione di Biologia – Roma, Italy

*Corresponding authors:

Prof. Paolo Oliveri, PhD
Dipartimento di Farmacia (DIFAR)
Università degli Studi di Genova
Viale Cembrano, 4
16148 Genova (Italy)
Phone: +39 010 3532626
e-mail: oliveri@difar.unige.it

Prof. Marco Vincenti, PhD
Dipartimento di Chimica
Università degli Studi di Torino
Via P. Giuria, 7
10125 Torino (Italy)
Phone: +39 011 6705264
e-mail: marco.vincenti@unito.it

Abstract

From a criminalistic point of view, the accurate dating of biological traces found at the crime scene, together with its compatibility with the estimated crime perpetration timeframe, enables to limit the number of suspects by assessing their alibis and clarifying the sequence of events. The present study delineates, for the first time, the possibility of dating biological fluids such as semen and urine, as well as blood traces, by using a novel non-destructive analytical strategy based on hyperspectral imaging in the near infared region (HSI-NIR), coupled with multivariate regression methods. Investigated aspects of the present study include not only the progressive degradation of the biological trace itself, but also the effects of its interactions with the support on which it is absorbed, in particular the hydrophilic vs. hydrophobic character of fabric tissues. Results are critically discussed, highlighting potential and limitations of the proposed approach for a practical implementation.

Keywords: near infrared (NIR) spectroscopy; hyperspectral imaging (HSI); partial least squares (PLS) regression; biological fluid; aging estimation; forensic analysis.

1. Introduction

The recent advancements in DNA extraction and amplification awarded to biological evidences a crucial role in the forensic investigation. Their detection within the crime scene consistently provide the identification of a suspect or a victim, but also allows to exonerate innocent individuals and gather further knowledge to reconstruct the crime action (for example, by pattern analysis). Another part of information that biological traces may reveal is the alleged time frame when a crime was committed (Weyermann & Ribaux, 2012). This estimation substantially contributes to the activity level proposition assessment as it is proposed by the "hierarchy of propositions" theory (Cook et al. 1998). For these reasons, a procedure that allows to correlate the time of air exposure of a biological trace with its progressive degradation, is likely to provide considerable advantage in the resolution of violent criminal events. For estimation of the age of a bloodstain, several techniques have been investigated. Recent review by Bremmer (Bremmer, Bruin, Gemert, Leeuwen, & Aalders, 2012) and Zadora (Zadora & Menżyk, 2018) looked over the methods applied to this purpose, including high-performance liquid chromatography, electron paramagnetic resonance, atomic force microscopy, RNA degradation measurements, infrared and Raman spectrocopies. However, none of these methods is implemented in the forensic practice yet, because they are subjected to a variety of potential bias sources (Menżyk et al. 2020), require complex sample preparations and need to be performed in a laboratory setting (Edelman et al. 2012). An even more limited number of analytical techniques have been described for the age estimation of other biological traces.

The sole availability of latent traces, often in minimal amount, together with the presence of contaminants (Rosenblatt et al. 2019) and fluid mixtures, frequently makes the biological trace detection remarkably complex (Harbison & Fleming, 2016). While it is requested that the suspected biological evidences are promptly detected and identified, traditional body fluid identification and analysis unfortunately present severe limitations, essentially due to the critical and/or degraded conditions in which these evidences are often found (Bremmer, 2012). Moreover, their detection with traditional tests involves further degradation or even destruction of the sample (Virkler & Lednev, 2009). In particular, the use of chemical reagents makes the process of latent trace visualization destructive, preventing the repetition of the analysis possibly requested. On the other hand, the DNA profiling performed on a minimal amount of biological fluids requires its careful preservation. Indeed, both sample contamination by the use of presumptive chemical tests and degradation by confirmatory analysis represent substantial threats when only minimal/latent traces are available (Rankin-Turner et al. 2019). All these issues push research toward the development of new

techniques allowing to detect minute biological traces in a selective and non-destructive way, possibly using miniaturized and portable devices, enabling in-field activity.

The significant progress in the laser technology and the development of innovative detectors have greatly improved the spectroscopic methods for molecular characterization (Virkler & Lednev, 2009). In this regard, infrared (IR) radiation between 1000 and 2500 nm (referred to as the near infrared – NIR – region) represents a very powerful tool for forensic sciences, especially when it is combined with hyperspectral imaging techniques (HSI-NIR). The inherent investigation of the spatial dimension allowed by HSI-NIR makes it highly valuable in the analysis of heterogeneous surfaces. In detail, the HSI-NIR data are organized in a four-dimensional matrix, often called "hypercube", in which the x and y dimensions account for the spatial domain, while the z dimension represents the spectral one (reflectance values for each wavelength). The hypercube data processing using appropriate chemometric techniques allows the analyst to gain information on both the chemical nature of the substances present at the investigated surface and their location on it.

Recently, an analytical protocol based on HSI-NIR data was developed for the screening of various common substrates to highlight and identify various biological traces in the potential context of a crime scene (Malegori et al., 2020). In this study, a straightforward and reliable data processing for the extraction of the chemical information embodied within the spectra was proposed. This approach was based on calculating a normalized difference image (NDI), which provides a monolayer grayscale image maximizing the difference between the stain and the supporting material (Malegori & Oliveri, 2018).

In the present study, such an analytical protocol was applied to biological evidences enabling the evaluation of trace degradation processes, like oxidation or carboxylation, related to the period of air exposure. The application of multivariate regression to the acquired NIR spectra provided an estimate of the exposure period, corresponding to the time elapsed from the moment the crime was committed, based on the degradation level of the biological trace.

The choice of robust and straightforward chemometric techniques for data processing pursues the same aim. In the legal field, the traces should ideally be analysed and interpreted at the crime scene, in the original context, with no chance of external contamination. In contrast, the analytical procedures currently used to highlight and detect the biological traces exploit chemical and/or optical methods, both involving significant contact with the sample (Bremmer, Edelman, Vegter, Bijvoets, & Aalders, 2011). Hyperspectral imaging is particularly interesting within this context, as it is suited for non-contact identification of evidence, minimizing the risk of trace contamination and destruction. Further advantages associated with the spectroscopic techniques include the speed of analysis, absence of solvents and samples pre-treatments, ability to discriminate and identify several biological fluids. All these aspects make the HSI-NIR technique particularly appropriate and promising to pursue the goal of biological traces characterization.

2. Materials and Methods

2.1 Experimental protocol

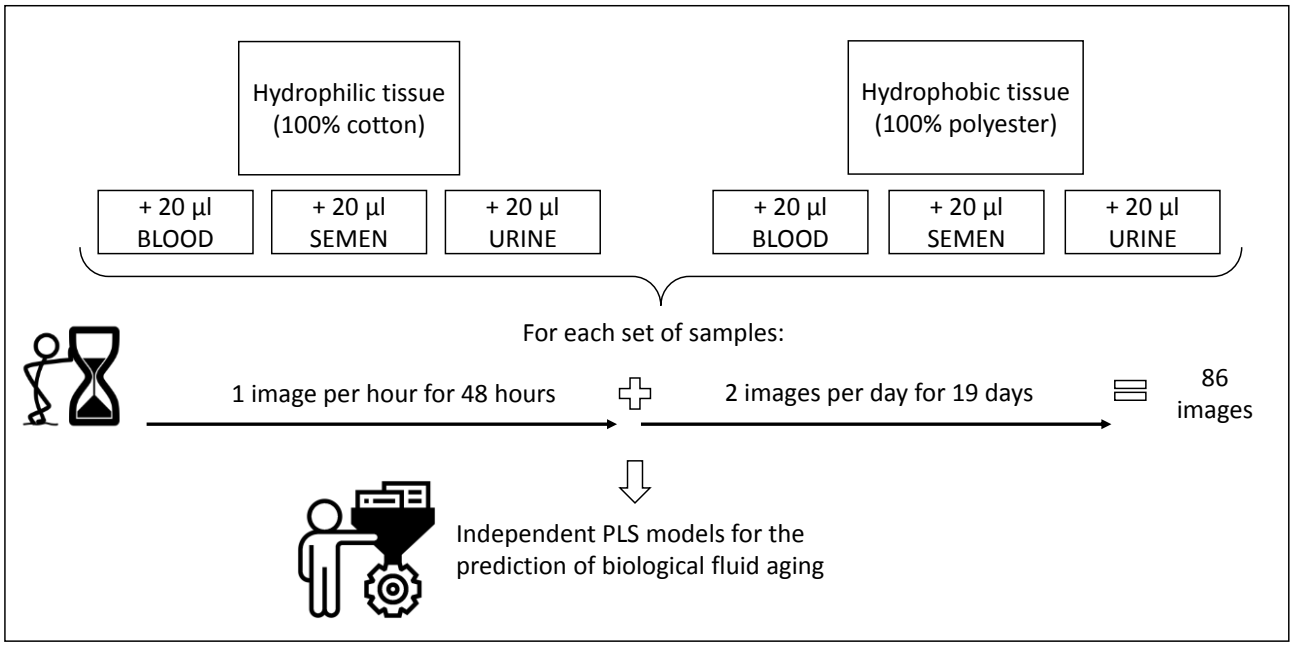


Figure 1: Scheme of the experimental protocol.

According to the scheme reported in Figure 1, two tissues with different hydrophilic vs. hydrophobic properties, namely black cotton and black polyester fabrics, were selected to investigate potentially distinct interactions occurring between the biological fluids and the supporting materials depending on their nature and chemical features.

Among the biological evidences most frequently found in crime scenes, three biological fluids were investigated: blood, urine, and semen. These fluids were provided by *Centro Regionale Antidoping e di Tossicologia "A. Bertinaria", Laboratorio di Genetica Forense* (Orbassano, TO - Italy) after collection from healthy volunteers. All samples were collected in plastic vials and stored at + 4 ° C for three days before application

The two fabrics were cut to produce three 5x5 cm cut-outs each to be used as a support for the deposition of the biological traces. The six supports plus three glass slides was stained with 20 µL of one of the fluids under examination: blood, urine, and seminal fluid. In the experimental protocol, the analysis of biological fluids on a glass support (transparent to NIR radiation) was included to confirm the characteristic NIR absorptions of each biological fluid.

The samples were analysed immediately after fluid deposition and the acquisitions were repeated every hour, for 48 hours, to evaluate the spectroscopic changes over time. At the end of the initial 48 hours, the hyperspectral images of the samples were acquired twice a day for 19 days at regular 12-hours intervals. Globally, each fabric scrap was subjected to 86 HSI-NIR analyses, spanning over a total of 504 hours. The acquisition frequency was chosen according to the known alteration kinetics of the main biological fluids constituents, typically occurring at high rate within the few initial hours and decreasing at longer time intervals. To simulate the conditions occurring at an actual crime scene, humidity and temperature were not set at controlled values, but both parameters were continuously monitored over the experimental periods. In detail, the average laboratory temperature was 25 °C (± 2 °C) and the average humidity was 60% (± 7 %). The room lighting followed regular circadian cycles. To best simulate typical conditions found at an actual crime scene, the tissue samples were exposed only to the light coming from the laboratory windows.

2.2 Hyperspectral imaging

For all types of samples, HSI-NIR data were acquired by a push-broom system composed of a SWIR3 hyperspectral camera operating in the 1000–2500 nm spectral range, at 5.6 nm resolution (Specim Ltd, Finland). The instrumental setting is characterised by three halogen lamps (35 W, 430 lm, 2900 K, each) as the illumination sources and a horizontal line scanner (40 × 20 cm LabScanner, Specim Ltd, Finland) on which samples are laid down. The system is controlled by the Lumo Scanner v. 2.6 software (Specim Ltd, Finland). Before each measurement, dark (closed shutter) and white (99% reflectance Spectralon® rod) images were automatically recorded and stored and were used to compute the spectral reflectance value (R) for each pixel and wavelength. For the acquisition, the scan parameters were set as follows: frame rate equal to 50.00 Hz and exposure time equal to 9.00 ms; manual focus was tuned before the scan. By scanning the entire surface of the sample, a complete three-dimensional hyperspectral image was created, where the first two dimensions represent the spatial information (coordinates of the planar space) while the third one is the spectral coordinate.

The spectra, recorded in the reflection mode (reflectance scale), were subsequently converted into pseudo-absorbance Log(1/R) values (Manley, 2014) to facilitate the direct understanding of the spectral characteristics (Oliveri, Malegori, Simonetti, & Casale, 2019).

2.3. Multivariate data processing

Images of the dehydrated biological fluids collected at different time points were processed to evaluate and confirm the characteristic NIR absorptions of each biological fluid. Specifically, their NIR images collected from glass supports were analysed by means of principal component analysis (PCA). To minimise unwanted systematic signal contributions due to physical irregularities of the sample surface, the NIR spectra were pre-processed by means of the standard normal variate transform (SNV) as a scatter-corrective pre-processing (Barnes et al. 1989), together with a Savitzky–Golay smoothing (15-datapoint window size, third-degree polynomial). The PCA outcomes confirmed the expected selection of the most informative wavelengths for each biological fluid, described in (Malegori et al. 2020). These characteristic wavelengths were employed to calculate a normalized difference image (NDI) which produces a monolayer grayscale image that maximizes the difference between the stain and support and highlights the biological trace (Malegori et al. 2020):

$$NDI = \frac{R_{\lambda s} - R_{\lambda f}}{R_{\lambda f}} \quad (1)$$

where $R_{\lambda s}$ is the reflectance value at the wavelength of the support (cellulose or polyester) and $R_{\lambda f}$ is the average reflectance value at the wavelength characteristic for each dehydrated biological fluid. In this way, the monolayer grayscale image is scaled between its maximum (forced to 255, white) and minimum (forced to 0, black) reflectance values.

The NDI image was calculated for each sample and time point. Afterwards, a masking procedure was applied to the images, with a threshold set on the corresponding NDI intensity histogram (Piarulli et al., 2020) that eliminates the pixels not matching the biological trace. The mean spectra were calculated by averaging the spectral profiles that are associated to the pixels constituting each of these crops. Then, they were used for building multivariate regression models by the Partial Least Squares (PLS) method, with the aim to evaluate possible correlations between the observed changes in the chemical content of the biological traces and their exposure time. In detail, six regression models were built: one for each type of biological fluid on both hydrophilic (cotton) and hydrophobic (polyester) supports. Each regression model was created using the NIR spectra as predictor variables and the exposure time (in hours) as the response variable. Spectra recorded at eight analysis times evenly distributed over the entire time range were selected as the test set for the models validation, while the training set was composed of 70 spectra. The data were pre-treated by column mean-centering. The optimal complexity (number of latent variables) of the PLS models was selected by applying a

cross-validation scheme on the training set data, with 5 cancellation groups, and choosing the number of latent variables (LV) corresponding to the model characterised by the lowest cross-validation error (RMSECV). In the subsequent validation phase, the optimised PLS model was applied to the test set and an algorithm was designed for creating false-colour maps of the biofluid stains, providing a visual and immediate understanding of trace aging time. The algorithm includes a first step of automatic segmentation of the image, based on a cut-off in the histogram of the NDI values, to automatically identify all the pixels corresponding to the presence of the biological trace on the support. By applying an object-based approach (Malegori & Oliveri, 2018), an average spectrum was calculated for each biofluid stain, on which the PLS prediction model of the time interval elapsed from the deposition of the biological trace was applied. Response values (hours from the deposition) computed by applying PLS model were codified according to a chromatic false-colour scale from red (minimum response value) to blue (maximum response value), used to color the selected pixels within the image.

3. Results and Discussion

3.1. Biological traces on glass

Prior to the development of regression models, to model the characteristic evolution of each biological traces during the exposure time in the spectral region of interest, the absorption patterns of dehydrated fluids on glass were assessed with the aim of confirming the distinctive wavelengths for each trace under study. The most informative wavelengths, confirmed in such a way, are reported in a previous study (Malegori et al. 2020). These wavelengths were used preliminarly to the subsequent development of regression models on the different tissues, to identify and segment the biological stains through the NDI calculation.

3.2. Biological traces on hydrophilic and hydrophobic tissues

PLS regression models computed on the mean spectra of each stain, automatically identified by the NDI procedure, were performed, initially, considering the whole time interval, for predicting the exposure time for the various biological traces, both on hydrophilic and hydrophobic tissues. Details of each model (number of latent variables, root mean square error in cross-validation – RMSECV, and root mean square error in prediction – RMSEP) are reported in Table 1. It should be remarked that root mean square errors have the statistical meaning of a standard deviation (dispersion parameter), expressed with the same measurement unit and scale as the response variable. The corresponding graphical representations of predicted vs. experimental time values are summarized in Figure 2. The blue line is the bisector of the orthogonal space and shows the ideal trend (predicted = experimental).

Table 1. Parameters associated to all of the PLS regression models built.

Model	Time range	LVs	RMSECV	RMSEP
Blood hydrophilic	Total	12	11.4	10.8
	First 48 h	10	4.4	---
	Following 19 days	14	11.2	---
Blood hydrophobic	Total	16	15.5	16.5
	First 48 h	14	4.3	---
	Following 19 days	9	16.1	---
Semen hydrophilic	Total	9	41.3	23.4
	First 48 h	7	6.2	---
	Following 19 days	11	40.8	---
Semen hydrophobic	Total	18	74.8	79.3
	First 48 h	1	14.0	---
	Following 19 days	8	92.4	---
Urine hydrophilic	Total	14	37.0	36.3
	First 48 h	7	6.4	---
	Following 19 days	8	62.9	---
Urine hydrophobic	Total	11	48.3	34.5
	First 48 h	10	7.4	---
	Following 19 days	12	66.0	---

In Figure 2*a*, the scatter plot of the predicted versus actual response (time elapsed from the deposition of the trace, expressed in hours) in cross-validation is shown for the blood trace deposited on the hydrophilic tissue. In this case, the PLS model was performed with all observation times of the training set and calculated using 12 latent variables, as suggested by the lowest root mean squared error in cross validation (RMSECV), which was equal to 11.4 h. The subplot 2*b* shows the corresponding graphical output for the test set, with an associated RMSEP equal to 10.8 h.

Figure 2*c* shows the scatter plot of the PLS model (predicted *vs.* actual response) in cross-validation for the blood trace laid on hydrophobic tissue. In this case, the model was calculated using 16 latent variables, as suggested by the lowest RMSECV, equal to 15.5 hours. The corresponding graph of the predicted *vs.* actual response calculated for the evaluation set is reported in Figure 2*d*, with an associated RMSEP equal to 16.5 h.

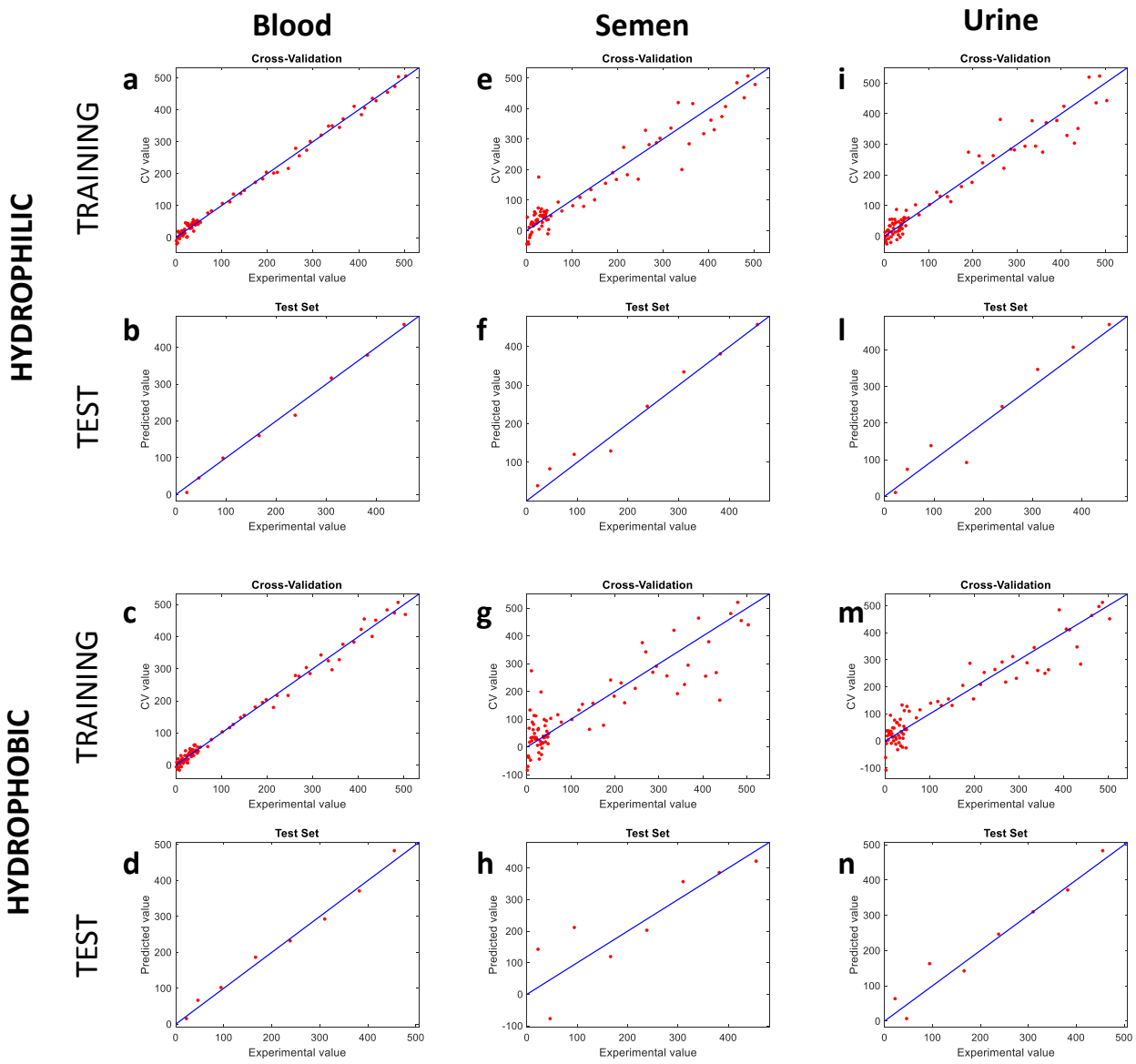


Figure 2: Predicted vs. experimental outcomes of the PLS models (both in CV and on the test set) for the different biological fluids and supports considered.

Analogous models were obtained for semen (Figure 2*e*-*h*) and urine (Figure 2*i*-*n*) traces deposited on hydrophilic and hydrophobic supports, respectively. The corresponding graphs, calculated both for the training set in cross-validation and for the test set respectively, are reported in the subplots, as indicated in the figure labels. In particular, the PLS models obtained on the training set for the trace of seminal fluid deposited on hydrophilic (Figure 2*e*) and hydrophobic (Figure 2*g*) tissues arose from 9 and 18 latent variables, corresponding to minimum RMSECV values of 41.3 hours (hydrophilic support) and 74.8 h (hydrophobic support), respectively. Results on the test set lead to RMSEP equal to 23.4 and 79.3, respectively (see Table 1). It is evident that larger RMSE values were observed for the seminal fluid with respect to blood traces, corresponding to broader scattering of data-points observed in Figures 2*e* and 2*g* compared with Figures 2*a* and 2*c*. This is not surprising, as the NIR spectral absorptions of semen traces are less intense and resolved with respect to the ones that characterise blood traces (Malegori et al. 2020).

Similar conclusions can be drawn from the regression models built that predict the time elapsed from deposition of the urine traces on hydrophilic and hydrophobic substrates. In particular, Figures 2*i* and 2*m*

show the predicted *vs.* actual delay time from deposition, for each substrate, as calculated from 14 and 11 latent variables, respectively. The corresponding lowest value of RMSECV was 37.0 h and 48.3 h for the two substrates. Results on the test set lead to RMSEP equal to 36.3 and 34.5, respectively.

From rough inspection of Figure 2 and Table 1, it is preliminarly inferred that the predictions appear quite effective for blood and appreciably less accurate for urine and semen, and that the hydrophilic tissue produce more predictable results than the hydrophobic tissue.

It is common opinion in forensic science that the biological traces may undergo major changes during the initial hours/days after deposition, while the degradation kinetics slows down afterwards (Zadora & Menżyk, 2018). For example, it has been demonstrated that clear distinction can be made between relatively fresh bloodstains and days-old ones (Doty, McLaughlin & Lednev, 2016) and between the latter and years-old traces (Doty, Muro & Lednev, 2017), but accurate discrimination within each of these broad timeframes are extremely complicated. Our experimental scheme followed a setting coherent with the expected degradation kinetics, with frequent HSI-NIR data acquisition in the first 48-hours (one every hour) and less frequent data collection in the following 19 days (twice per day). From this concept, it stems that separate PLS models for the two time frames may possibly fit better the data and provide more accurate prediction of the deposition time than a single model built on all data points. The outcome of these processing are depicted in Figure 3, where each dataset (and the corresponding PLS model) was separated into two: one related to the spectra collected within the first 48 hours and the second related to the following 19 days. The characteristics of all the PLS models obtained according to such a scheme are reported in Table 1.

From an overall inspection of Figure 3, it can be noted that the separate models follow comparable trends, with the notable exception of the model corresponding to the seminal fluid deposition on hydrophobic fabric. It is possible to conclude that, in practical forensic caseworks, it is not too detrimental to process the data by a single linear PLS regression, but the prediction accuracy improves with separate models, especially for acquisition times subsequent to the initial 48 hours, as it clearly appears from a comparison of the related RMSECV values reported in Table 1. As an example, RMSECV of the global PLS model for blood deposited on hydrophobic fabric is equal to 15.5 h, while the corresponding RMSECV of the PLS model limited to the first 48 h is equal to 4.3 h (less than one third).

A peculiar behaviour is observed for the traces of seminal fluid on hydrophobic fabric, showing almost no change in the first 48 hours, as a result of very limited evolution of the NIR spectral tracing, which is reflected in a flat-line PLS model and an almost constant predicted value (Figure 3*f*). In other words, no significant changes in the chemical characteristics of the sample occurred in the first 48 hours. The reason for such an odd observation relies in the chemical composition of the seminal fluid, consisting of a liquid component, called seminal plasma, and a cellular part, the spermatozoa. Seminal plasma is an aqueous solution of organic and inorganic substances, that serves as a natural reservoir of nutrients such as sugars, lipids and proteins as well as antioxidants (Sharma et al., 2013). The aqueous substrate of the seminal plasma is likely to prevent the soaking of the highly hydrophobic tissue, leaving the seminal plasma and the antioxidants contained in it at the surface of the fabric. According to this hypothesis, the initial oxidation of the semen trace involves the lipids, not the proteins, *i.e.* the substances monitored by the selected NIR wavelengths. After 48 hours, the protein degradation process starts, yielding to NIR absorption changes and, consequently, to an interpretable model (Figure 3*h*). When laid on hydrophilic tissue, such as cotton, the semen trace apparently undergoes

regular degradation kinetics of the protein component yielding to a linear model with predictable deposition times already in the initial 48 hours (Figure 3*e*).

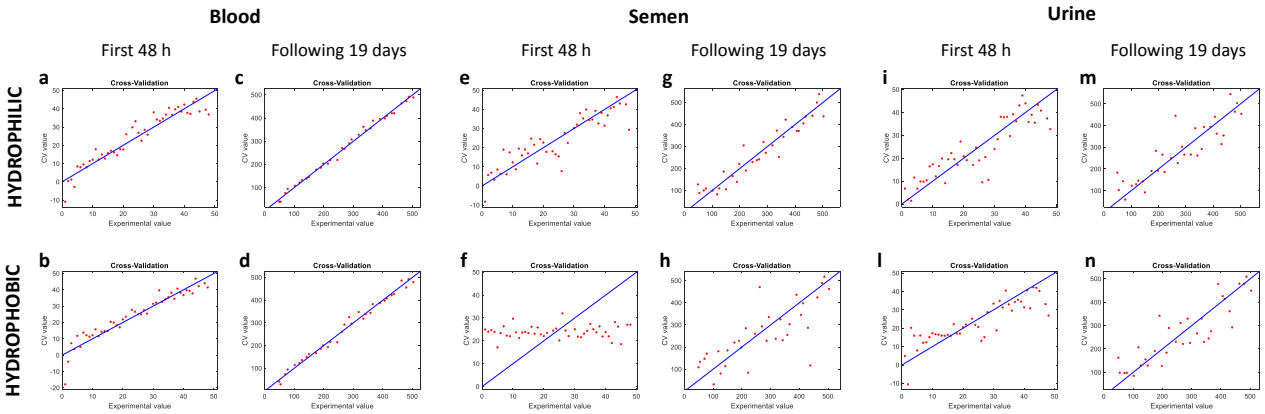


Figure 3: Separation of global time models into two distinct models: observations in the first 48 hours and observations in the following15 days. Blood (a-d), semen (e-h), and urine (i-n), for the hydrophylic (upper part) and for the hydrophobic fabric (lower part)

The blood evidence laid on hydrophilic and hydrophobic fabrics gave rise to PLS models characterized by a similar trend Figure 3*a-d*. Also the RMSECV values obtained from models that considers all data and those relative to the last 19 days are comparable (see Table 1). The degradation kinetics observed in the first 48 hours slightly differs from that obtained in the following 19 days, and logically correspond to lower RMSECV values (4.4 and 4.3 h for the first 48 h and 11.2 and 16.1 h for the subsequent 19 days). This result is in accordance with the studies reported in the literature regarding the blood oxidation processes, which is typically biphasic (Mie Tsuruga et al. 2001). In more detail, upon blood exposure to air, hemoglobin (HbO_2) rapidly (< 48 hours) interacts with oxygen in the ambient environment and undergoes auto-oxidation to methemoglobin (met-Hb). Once hemoglobin is auto-oxidized to met-Hb, it is slowly (hours or days, depending on the humidity and temperature of the surrounding environment) denatured to hemichrome (HC). The entire process is outlined below:

$$HbO_2 \xrightarrow{k_1} metHb \xrightarrow{k_2} HC \qquad (2)$$

It was later established how the oxidation of oxyhemoglobin in bloodstains follows a biphasic decay and that the rates can be described by first-order reaction kinetics. Besides, it was shown that all reaction rates exhibit a positive correlation with temperature and that the transition of met-Hb to HC also depends on humidity (Bremmer, de Bruin, et al., 2011).

3.3 Time trends displayed by colour maps

False-color prediction maps obtained for the test set samples provide the operator with a visual and immediate understanding of the exposure time of the biological trace to atmospheric air. Response values (in hours) are coded using a chromatic scale from red to blue, for the graphical representation in the image.

Figure *4* shows the maps resulting from applying this strategy to the eight samples forming the test set for each fluid-support combination.

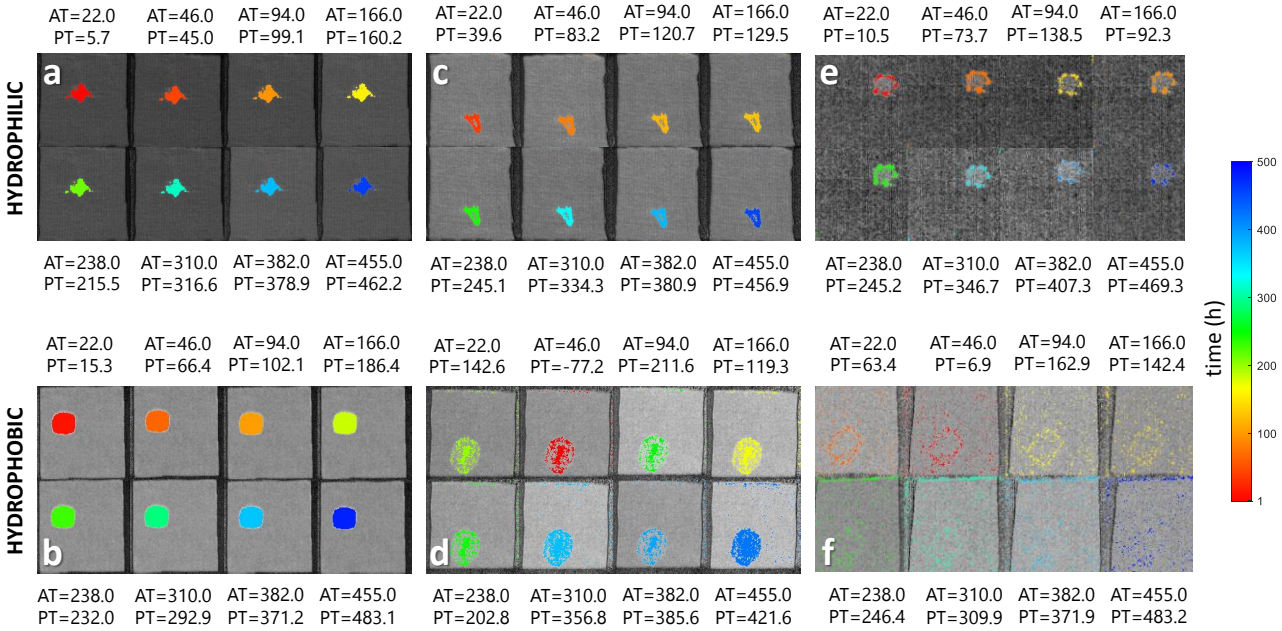


Figure 4: Comparison of the color maps related to the different fluids on the two supports. AT = actual time (hours); PT = predicted time (hours) obtained by application of PLS regression.

With the only exception of the first acquisition time, it is evident that the model can predict the exposure time of almost any trace with adequate degree of accuracy. The captions inserted in Figure 4 show the comparison between the real exposure time (AT) and the one predicted by the model (PT). The larger prediction uncertainty associated with the first acquisition time is likely due to non-linear dehydration kinetics found in the first 48 hours. On the other hand, the most striking outcome resulting from the data presented in Figure 4 is that the prediction accuracy increases with deposition time, particularly for blood traces. For example, the average absolute difference between actual and predicted deposition time in the timeframe 94-455 hours (approximately 4-19 days) is as low as 3.8% for blood deposited on the hydrophilic substrate and 6.6% for blood deposited on the hydrophobic tissue. This exceptional accuracy in prediction overturns the former studies' conclusions that evidenced a progressive decay of the distinctive spectral differences for the data collected after long deposition periods, as a result of the gradual completion of the blood degradation pathway (Menżyk et al. 2020). From a forensic point of view, it is particularly important to get reliable preditions from relatively old biological traces, because the elapse of a significant delay between the crime occurrence and the opening of forensic investigations represents a recurrent situation. On the other hand, the longer time elapses since the crime event, the more relevant become the random environmental factors that modify the trace degradation kinetics (Bremmer, 2012).

The prediction of the PLS model relative to the urine stains (subplots Figure 4*e* and 4*f*) is distinctly worse than those obtained for blood and semen. In this case, the models can still provide a satisfactory prediction of the exposure time for the urine trace, especially for longer exposure periods, but with lower accuracy than that obtained for the other biological fluids. This effect is more evident on the map related to the urine traces laid on the hydrophobic tissue (subplot 4*f*). The lower protein content present in urine with respect to blood and semen results in both a more difficult identification of the biological traces and a larger uncertainty in estimating the exposure time. Overall, lower prediction efficiencies corresponds to higher water content of the biological fluids under examination and longer persistence of significant amount of water on the

hydrophobic tissues. As a matter of fact, the higher water content alone does not explain the significant variation in the model predictive capacity. Further reason to be accounted for these differences is represented by the chemical nature of the matrices themselves. Indeed, both semen and blood are rich in proteins, while urine typically has low protein and high mineral salts content, the latter not presenting active groups in the near-infrared region. Consequently, the NIR signals recorded for dehydrated urine is not as informative as those acquired from the other two biological traces.

3. Conclusions

Several techniques have been explored during the last century to address the age determination of biological fluid stains. However, most of these studies were addressed to the aging of blood traces, leaving little room for the aging study of other biological matrices, such as dehydrated urine and seminal fluid, both being frequently found in the context of violent crimes. For the first time, this study systematically investigated the aging of various biological fluids – including blood, semen, and urine – using a state-of-the-art NIR imaging technique combined with chemometrics. Attention was not only devoted to the denaturation of the biological sample, but also the main interactions occurring between the biological trace and its supporting material were taken into account. In particular, highly hydrophilic and highly hydrophobic fabrics were considered. These experiments highlighted the major contribution of the aqueous fraction in the stability of the biological sample and its consequent impact on the predictive ability of the regression model. From a practical point of view, this information indicates that a different interpretation of the NIR response is needed, depending on the support hosting the biological trace.

Currently, all techniques used for biological trace aging studies are still developing and provide large standard deviations and partially inaccurate age estimates. For example, in a study by Li et al., a blood age prediction of the test set is reported with an overall average error of ±1.17 days (Li, Beveridge, O'Hare, & Islam, 2013). Reducing the standard deviation of predictions represents a fundamental challenge toward its implementation into forensic practice and eventually in court (Bremmer et al., 2012). The outcomes of the present study highlights a significant reduction of the model prediction uncertainty, which for aging bloodstains was equal to 3.8% (10.8 hours RMSEP value) in hydrophilic tissue and 6.6% (16.5 hours RMSEP value) in hydrophobic one. This improvement in estimating the trace age makes the procedure encouraging in the perspective of its application in a real crime scene. In particular, the present procedure should be combined with an on-site experimental set-up (Menżyk 2021) and a likelihood ratio-based approach (Menżyk et al. 2020) in order to address the role of atmospheric factors, such as humidity and temperature, contaminants and other interferences (Rosenblatt et al. 2019), which may play a key role in the trace degradation kinetics.

Another advantage of using NIR hyperspectral imaging for biological trace age estimation is its speed of acquisition. Actually, the time needed for the analysis depends on the crime scene size and the calculation power. In the specific case of the present study, the application of basic chemometric techniques combined with the speed of NIR image acquisition led to obtain the final answers in less than 5 minutes. Considering the practical potential of the protocol proposed in this study and its simplicity of use, the opportunity to implement it with portable instrumentation is likely to represent a strongly innovative turning point in crime scene investigation. Portable NIR hyperspectral imaging instrumentation would be used by forensic investigators to identify latent traces (Malegori et al. 2020), clarify the circumstances surrounding a violent crime, and establish any connection between victim, suspect, and crime scene. Lastly, the HSI-NIR analysis is completely non-destructive and non invasive making it particularly suitable for the direct detection of biological traces in the original context, without any damaging contact and preserving the DNA content of the trace.

Funding

Financial support provided by Università degli Studi di Genova (Research Project Curiosity Driven 2020: "3Depth – From 2D to 3D hyperspectral imaging exploiting the penetration depth of near-infrared radiation", CUP: D34G20000100005) is gratefully acknowledged.

References

Rankin-Turner, S., Turner, M. A., Kelly, P. F., King, R. S. P., Reynolds, J. C. (2019). Transforming presumptive forensic testing : in situ identification and age estimation of human bodily fluids. *Chemical Science*, 10, 1064–1069. https://doi.org/10.1039/c8sc04133d

Barnes, R. J., Dhanoa, M. S., & Lister, S. J. (1989). Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. *Applied Spectroscopy*, *43*(5), 772–777. https://doi.org/10.1366/0003702894202201

Bremmer, R. H., Bruin, K. G. De, Gemert, M. J. C. Van, Leeuwen, T. G. Van, & Aalders, M. C. G. (2012). Forensic quest for age determination of bloodstains. *Forensic Science International*, *216*(1–3), 1–11. https://doi.org/10.1016/j.forsciint.2011.07.027

Bremmer R. H., de Bruin D. M., de Joode M., Buma W. J., van Leeuwen T. G., Aalders M. C. G. (2011). Biphasic oxidation of Oxy-Hemoglobin in bloodstains. *PLoS ONE*, *6*(7), e21845. https://doi.org/10.1371/journal.pone.0021845

Bremmer, R. H., Edelman, G., Vegter, T. D., Bijvoets, T., & Aalders, M. C. G. (2011). Remote Spectroscopic Identification of Bloodstains. *Journal of Forensic Sciences*, *56*(6), 1471–1475. https://doi.org/10.1111/j.1556-4029.2011.01868.x

Cook, R., Evett, I.W., Jackson, G., Jones, P.J., & Lambert, J.A. (1998). A hierarchy of propositions: Deciding which level to address in casework. *Science & Justice* 38(4) (1998) 231-240. https:// doi.org/10.1016/S1355-0306(98)72117-3

Doty K.C., McLaughlin G., Lednev I.K. (2016). A Raman "spectroscopic clock" for bloodstain age determination: the first week after deposition, *Anal. Bioanal. Chem.* 408 (2016) 3993–4001. https://doi.org/10.1007/s00216-016-9486-z

Doty K.C., Muro C.K., Lednev I.K. (2017). Predicting the time of the crime: bloodstain aging estimation for up to two years, *Forensic Chem.* 5 (2017) 1–47. https://doi.org/10.1016/j.forc.2017.05.002

Edelman, G., Leeuwen, T. G. Van, & Aalders, M. C. G. (2012). *Hyperspectral imaging for the age estimation of blood stains at the crime scene*. *223*, 72–77. https://doi.org/10.1016/j.forsciint.2012.08.003

Harbison, S., & Fleming, R. (2016). Forensic body fluid identification: state of the art. *Research and Reports in Forensic Medical Science*, 11. https://doi.org/10.2147/rrfms.s57994

Jerry Workman Jr., L. W. (2012). *Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy*.

Kissi, N., Curran, K., Mogire, C. V., Fearn, T., & Mccullough, L. (2017). Developing a non‐invasive

tool to assess the impact of oxidation on the structural integrity of historic wool in Tudor tapestries. *Heritage Science*, 1–13. https://doi.org/10.1186/s40494-017-0162-1

Li, B., Beveridge, P., O'Hare, W. T., & Islam, M. (2013). The age estimation of blood stains up to 30 days old using visible wavelength hyperspectral image analysis and linear discriminant analysis. *Science and Justice*, *53*(3), 270–277. https://doi.org/10.1016/j.scijus.2013.04.004

Malegori C., Alladio E., Oliveri P., Manis C., Vincenti M., Garofano P., Barni F., Berti A. (2020). Identification of invisible biological traces in forensic evidences by hyperspectral NIR imaging combined with chemometrics. *Talanta*, *215*, 120911. https://doi.org/10.1016/j.talanta.2020.120911

Malegori, C., & Oliveri, P. (2018). *Principal Component Analysis*, in: M.K. N.C. Basantia, L.M.L. Nollet (Ed.), Hyperspectral Imaging Anal. Appl. Food Qual, CRC Press, Boca Raton, FL, USA..

Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chem. Soc. Rev.*, *43*(24), 8200–8214. https://doi.org/10.1039/C4CS00062E

Menżyk, A., Damin, A., Martyna, A., Alladio, E., Vincenti, M., Martra, G., Zadora, G. (2020). Toward a novel framework for bloodstains dating by Raman spectroscopy: How to avoid sample photodamage and subsampling errors. *Talanta*, 209, 120565. https://doi.org/10.1016/j.talanta.2019.120565

Menżyk, A. (2021). Toward a spectroscopy-based approach for estimating time elapsed since bloodstains deposition – Development of a novel framework for blood evidence evaluation. Ph.D. Thesis, University of Silesia in Katowice, Faculty of Science and Technology, Institute of Chemistry.

Mie Tsuruga, Ariki Matsuoka, Akira Hachimori‡, Yoshiaki Sugawara§, and K. S. (2001). The Molecular Mechanism of Autoxidation for Myoglobin and Hemoglobin. *Seibutsu Butsuri*, *41*(2), 74–79. https://doi.org/10.2142/biophys.41.74

Oliveri, P., Malegori, C., Simonetti, R., & Casale, M. (2019). The impact of signal pre-processing on the final interpretation of analytical outcomes – A tutorial. *Analytica Chimica Acta*, *1058*, 9–17. https://doi.org/10.1016/j.aca.2018.10.055

Piarulli, S., Sciutto, G., Oliveri, P., Malegori, C., Prati, S., Mazzeo, R., & Airoldi, L. (2020). Rapid and direct detection of small microplastics in aquatic samples by a new near infrared hyperspectral imaging (NIR-HSI) method. *Chemosphere*, *260*, 127655. https://doi.org/10.1016/j.chemosphere.2020.127655

Rosenblatt R., Halámková L., Doty K.C., de Oliveira Jr E.A.C.., Lednev I.K., Raman spectroscopy for forensic bloodstain identification: Method validation vs. environmental interferences, *Forensic Chem.* 16 (2019) 100175. https://doi.org/10.1016/j.forc.2019.100175

Sharma, R., Agarwal, A., Mohanty, G., Du Plessis, S. S., Gopalan, B., Willard, B., … Sabanegh, E. (2013). Proteomic analysis of seminal fluid from men exhibiting oxidative stress. *Reproductive Biology and Endocrinology*, *11*(1), 1. https://doi.org/10.1186/1477-7827-11-85

Virkler, K., & Lednev, I. K. (2009). Analysis of body fluids for forensic purposes: From laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Science International*, *188*(1–3), 1–17. https://doi.org/10.1016/J.FORSCIINT.2009.02.013

Yan, H., Xu, Y., Siesler, H. W., Han, B., & Zhang, G. (2019). *Hand-Held Near-Infrared Spectroscopy for Authentication of Fengdous and Quantitative Analysis of Mulberry Fruits*. *10*(November), 1–15. https://doi.org/10.3389/fpls.2019.01548

Weyermann C., Ribaux O. (2012). Situating forensic traces in time. *Sci. Justice* 52 (2012) 68–75. https://doi.org/10.1016/j.scijus.2011.09.003

Zadora G. & Menżyk A. (2018) In the pursuit of the holy grail of forensic science – Spectroscopic studies on the estimation of time since deposition of bloodstains. TrAC - Trends in Analytical Chemistry, 105, 137 - 165. https://doi.org/10.1016/j.trac.2018.04.009

Check-list

Paolo Oliveri and Marco Vincenti have been designated as the corresponding authors with contact details:
• E-mail address ✓
• Full postal address ✓

All necessary files have been uploaded:
Manuscript:
• Including keywords ✓
• All figures (include relevant captions) ✓
• All tables (including titles, description, footnotes) ✓
• Ensure all figure and table citations in the text match the files provided ✓
• Indicate clearly if color should be used for any figures in print ✓ No, thanks
Graphical Abstracts ✓
Highlights files (where applicable) ✓
Supplemental files (where applicable) ✓ Not present

Further considerations
• Manuscript has been 'spell checked' and 'grammar checked' ✓
• All references mentioned in the Reference List are cited in the text, and vice versa
• Permission has been obtained for use of copyrighted material from other sources (including the Internet) ✓
• A competing interests statement is provided, even if the authors have no competing interests to declare ✓
• Journal policies detailed in this guide have been reviewed ✓
• Referee suggestions and contact details provided, based on journal requirements ✓

List of Three Potential Reviewers

Prof. Markus Baumgartner
University of Zurich
Institute of Forensic Medicine
Expert on Forensic Analytical Chemistry and spectral imaging
Markus.Baumgartner@irm.uzh.ch

Prof. Simon Lewis
Curtin University, Perth, Australia
Faculty of Science and Engineering
Expert on Forensic Analytical Chemistry and spectral imaging
S.Lewis@curtin.edu.au

Prof. Grzegorz Zadora
University of Silesia in Katowice
Faculty of Science and Technology
Institute of Forensic Research, Krakow, PL
Expert on Forensic Analytical Chemistry and Statistical Analysis in Forensic Sciences
greg_zadora@yahoo.com
grzegorz.zadora@us.edu.pl

Dr. Gerda Edelman
Netherlands Forensic Institute, The Hague, Netherlands
Department of Crime Scene Investigation
Expert on Crime Scene investigations, spectral imaging and multivariate data analysis
g.edelman@nfi.minvenj.nl

Declaration of interests

☐The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☒The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Paolo Oliveri reports financial support was provided by University of Genoa.