

SCIENTIFIC REPORTS



OPEN

Estimation of the age of human bloodstains under the simulated indoor and outdoor crime scene conditions by ATR-FTIR spectroscopy

Hancheng Lin^{1,2}, Yinming Zhang¹, Qi Wang¹, Bing Li¹, Ping Huang² & Zhenyuan Wang¹

Estimation of the age of human bloodstains is of great importance in forensic practices, but it is a challenging task because of the lack of a well-accepted, reliable, and established method. Here, the attenuated total reflection (ATR)-Fourier transform infrared (FTIR) technique combined with advanced chemometric methods was utilized to determine the age of indoor and outdoor bloodstains up to 107 days. The bloodstain storage conditions mimicked crime scene scenarios as closely as possible. Two partial least squares regression models—indoor and outdoor models with 7–85 days—exhibited good performance for external validation, with low values of predictive root mean squared error (5.83 and 4.77) and high R^2 values (0.94 and 0.96) and residual predictive deviation (4.08 and 5.14), respectively. Two partial least squares–discriminant analysis classification models were built and demonstrated excellent distinction between fresh (age ≤ 1 d) and older (age > 1 d) bloodstains, which is highly valuable for forensic investigations. These findings demonstrate that ATR-FTIR spectroscopy coupled with advanced chemometric methods can be employed as a rapid and non-destructive tool for age estimation of bloodstains in real-world forensic investigation.

Bloodstain is one of the most frequently encountered biological evidences at crime scenes¹, especially those for violent crimes. Accurate estimation of the age of bloodstains can be a tremendous help for forensic investigators in the reconstruction of the event timeline, determination of the time of death or injury of the victim, and reduction of the pool of suspects².

Numerous methods have been investigated to determine the age of bloodstains, such as use of oxygen electrodes³, electron paramagnetic resonance (EPR)⁴, high-performance liquid chromatography (HPLC)⁵ and RNA degradation^{6,7}. Unfortunately, these approaches are not robust, limiting their forensic application. In recent years, Agudelo *et al.*^{8,9} investigated human serum for determining the age of bloodstains using bioaffinity-based and biocatalytic assays. However, this method was inappropriate for forensic practice because in actual cases it is difficult to extract an adequate amount of serum from dried bloodstains. Additionally, several novel techniques have been explored to determine the age of bloodstains, including fluorescence lifetime measurements^{10,11}, atomic force microscopy¹² and the use of smartphones for quantifiable colour change correlations¹³.

Vibrational spectroscopy techniques are becoming more and more popular in forensic science because of their non-destructive, rapid, quantitative, and confirmatory features¹⁴. The results of several noteworthy studies (2011 and later) investigating the use of spectroscopic methods to estimate bloodstain age are summarized in Table 1. These methods showed an ability to determine bloodstain age, especially when combined with chemometric methods. However, most of the studies were conducted using simulated bloodstain samples under ideal laboratory conditions. In real-world case work, varying ambient conditions will affect the process of bloodstain denaturation and aggregation and ultimately contribute to the complexity and difficulty of bloodstain age estimation. In the present work, an approach combining ATR-FTIR spectroscopy with chemometric methods was established

¹Department of Forensic Pathology, Xi'an Jiaotong University, Xi'an, 710061, P. R. China. ²Department of Forensic Pathology, Institute of Forensic Science, Ministry of Justice, Shanghai, 200063, P. R. China. Correspondence and requests for materials should be addressed to P.H. (email: huangp@ssfjd.cn) or Z.W. (email: wzy218@mail.xjtu.edu.cn)

Year	Methods	Storage environment	Chemometrics	Range of age	Error of age prediction	Ref
2011	Reflectance spectroscopy	laboratory conditions	—	0–60 d	—	38
2011	Reflectance spectroscopy	laboratory conditions	LDA	1–19 d	±0.71 d	46
2012	Hyperspectral imaging	simulated crime scene	—	0.1–200 d	13.4% of the actual age	39
2012	Near infrared spectroscopy	laboratory conditions	PLSR	0–28 d	8.9% of the actual age	47
2013	Hyperspectral imaging	laboratory conditions	LDA	0–7, 0–30 d	±0.27 and ±1.17 d, respectively	48
2016	Raman spectroscopy	laboratory conditions	PLSR	1–168 h	±2.19 h	41
2017	Visible reflectance spectroscopy	calorstats	PCA-SVMR	2 h–45 d	±42.79 h	42

Table 1. Summary of spectroscopy techniques proposed for the age determination of bloodstains.

for determining the age of bloodstains up to 107 days. The bloodstain samples were created and stored in indoor and outdoor environments, and the storage conditions closely simulated real crime scenes.

Chemometric methods are capable of extracting useful information from complex spectral datasets to yield more comprehensive and accurate results¹⁵. The development of reliable and robust chemometric framework to handle analytical data has been identified as an important step in the biological spectroscopy analyses^{16–18}. In our study, partial least square regression (PLSR) and partial least square -discrimination analysis (PLS-DA) were employed to analyse blood spectral dataset, since these two PLS approaches have experienced a broad acceptance in the spectral analysis with their powerful ability of dealing with noisy and collinear spectral variables. A good example of PLS approaches in clinical applications is outlined in the study by Khoshmanesh *et al.*¹⁹, in which, the aim was to detect early-stage malaria parasites in infected erythrocytes. In the field of forensic science, these two methods have also been used, such as in the identification of species' blood^{20,21} and bone²², investigation of burned bones²³, and profiling of cocaine in seizures²⁴. In our laboratory, ATR-FTIR in combination with PLSR has proven to be a good tool for the characterization of post-mortem biochemical changes in rabbit plasma^{25,26}. However, one main problem of PLS approaches is over-fitting. To tackle this problem in our study, the constructed PLSR and PLS-DA models based on the resulting spectral dataset were validated with two independent sample datasets, both of which originated from two volunteers whose samples were separate from those used to develop the models.

Materials and Methods

Sample preparation. This study was conducted in accordance with the guidelines of the National Institute of Health, China. The protocol was approved by the Ethical Committee of Xi'an Jiaotong University and informed written consent was obtained from all blood donors. Fresh whole-blood samples (without anticoagulants) were obtained from four healthy volunteers (two males and two females) and deposited immediately onto glass slides to form bloodstains. Nineteen time points were set: 0.25, 1, 2, 3, 4, 5, 6, 7, 9, 12, 15, 19, 24, 30, 40, 50, 65, 85, and 107 d. For each time point, 6 bloodstain samples per donor were prepared, of which 3 samples were stored in an indoor environment and 3 in an outdoor environment. It should be emphasized that the indoor condition was not specifically controlled and the samples were exposed to dim sunlight during the day and no light at night. Bloodstain samples placed in the outdoor environment were exposed to the light, heat, and humidity of the outdoor environment but not rain. A total of 556 bloodstain samples were ultimately collected, encompassing an indoor training group of 228 samples and an outdoor training group of 228 samples. These two groups were used for chemometric model constructions. Additionally, according to the aforementioned method, two bloodstain groups (indoor and outdoor groups; each group contained 114 samples with bloodstain age ranging from 0.25 to 107 d) from two other healthy individuals (one male and one female) were prepared. These two groups, called test groups, were employed to validate the constructed chemometric models.

Spectra collection and data preprocessing. Spectral acquisition was performed using a Nicolet iS 50 FTIR Spectrometer (Thermo Fisher Scientific, Waltham, WA, USA) equipped with an ATR accessory (Thermo Fisher Scientific, Waltham, WA, USA) containing a diamond crystal face approximately 2 mm in diameter. Before each measurement, the bloodstain sample was collected in an Eppendorf tube and mixed with 10 μ L of normal saline uniformly. Subsequently, 1 μ L of sample was deposited on the ATR crystal face and dried with an air dryer for approximately 4 min. The spectra were recorded in the range of 900–1800 cm^{-1} at a resolution of 4 cm^{-1} with 32 scans. The background spectra were subtracted automatically from the sample spectra. For each sample, 3 replicated spectra were collected and then averaged to form a single spectrum. The spectra were recorded with OMNIC software version 9.2 (Thermo Fisher Scientific, Waltham, WA, USA).

Next, baseline correction, unit vector normalization and multiplicative scatter correction (MSC) preprocessing methods were applied to the 1800–900 cm^{-1} region to eliminate baseline offsets, remove artefacts related to the analytical techniques and samples under study, and reduce the effects of light scattering²⁷. The data preprocessing was carried out by Unscrambler 9.7 (CAMO software, Oslo, Norway). The original and pre-processed raw spectra are shown in the Supporting Information.

Multivariable statistical analysis. PLSR is a multivariate regression method that can decompose the X-variable with the guidance of the Y-variable and find latent variables (LVs), which are linear combinations

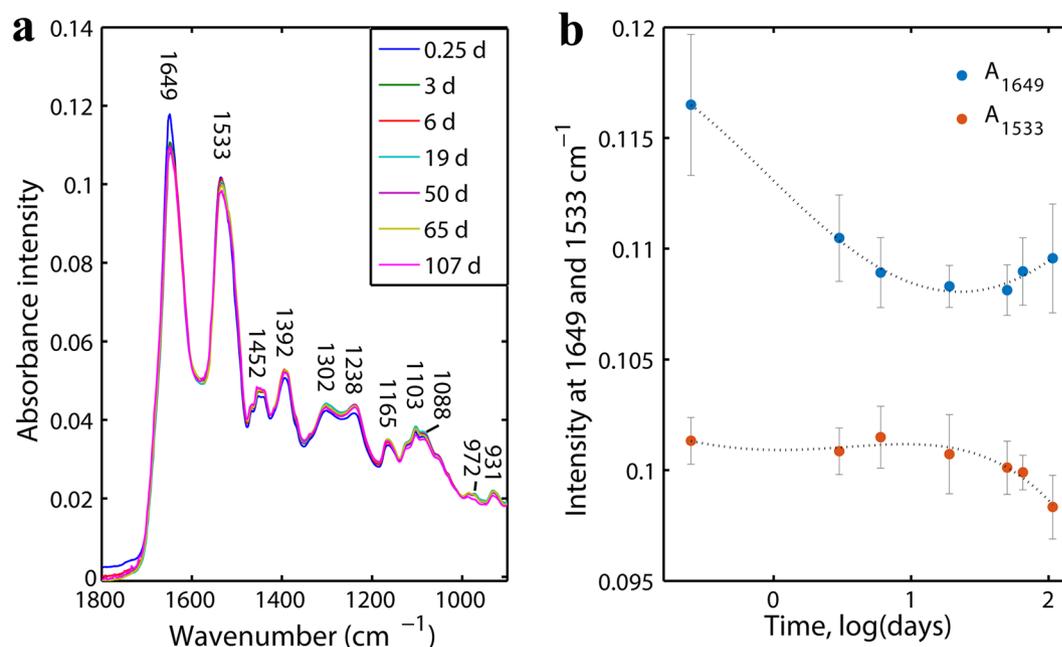


Figure 1. (a) FTIR averaged spectra of outdoor bloodstains at different time points in the range of 1800–900 cm^{-1} . (b) The trends of the intensities of the peaks at 1649 and 1533 cm^{-1} for all spectra over time with polynomial curve fit lines (model = 4).

of the original variables to maximize the co-variation between X and Y during regression²⁸. In this work, the X-variable corresponded to the matrix of spectral intensity and the responding Y-variable was associated with age values. PLS-DA is a classification method based on the PLS approach in which the Y-variable is chosen to represent the class membership²⁹. PLSR and PLS-DA were established using Matlab software version R2014a (MathWorks, Natick, MA, USA) equipped with PLS Toolbox 8.1.1 (Eigenvector Research, Manson, WA, USA).

To evaluate the stability and predictive ability of PLSR and PLS-DA models, both internal cross-validation (CV) and external validation were performed³⁰. In our study, the CV was performed using 10 folds with the Venetian blinds procedure. External validation was performed using constructed models to predict the test bloodstain samples that were stored in the same environment as the training bloodstain samples. Another test, called the “expanding test” (actually, this test is another type of external validation), was performed using constructed models to predict the test samples whose storage environment differed from that of the training samples. The purpose of the expanding test was to explore the predictive power of the models in estimating bloodstain age under multiple environmental conditions.

Root mean square error (RMSE), including calibrated RMSE (RMSEC), cross-validated RMSE (RMSECV) and predicted RMSE (RMSEP), and R^2 and residual predictive deviation (RPD), as the three main parameters of the model’s calibrated and predicted results, were used to evaluate the regression model reliabilities. High values of R^2 and RPD and a low value of RMSE demonstrate a well-established PLS regression model. Notably, an RPD value above 3 indicates that the model is very reliable for prediction purposes^{31,32}.

Data availability. All data generated or analysed during this study are available from the corresponding author on reasonable request.

Results and Discussion

In this work, the studied spectral range of 1800–900 cm^{-1} , also called the “biofingerprint region”, offers the most information on the chemical compounds of biological samples³³, including lipid esters (1800–1700 cm^{-1})³⁴; amide I, II and III proteins (1700–1500 cm^{-1} , 1350–1200 cm^{-1})^{34,35}; and nucleic acids and carbohydrates (1200–900 cm^{-1})³⁶. As for bloodstains, the corresponding infrared spectra provide detailed information regarding haemoglobin, which makes up 97% of the dry content of blood².

Figure 1a shows a comparison of the average spectra for the outdoor bloodstains with seven selected ages. The assignments of the main observed bands are tabulated in Table 2. As can be seen in Fig. 1b, the highly varied vibrational bands were at 1649 (corresponding to the α -helix structures of haemoglobin)³⁷ and 1533 cm^{-1} (representing amide II). The average absorbance intensity at 1649 cm^{-1} decreased at first, reached the minimum value at age = 19 d, and increased slowly thereafter. However, the average absorbance intensity at 1533 cm^{-1} did not show a decreasing trend until 19 days after deposition. These findings suggest that the secondary structures of haemoglobin changed constantly as age increased, and the most probable cause was associated with the kinetic efforts of haemoglobin ($\text{Hb} \rightarrow \text{HbO}_2 \rightarrow \text{met-Hb} \rightarrow \text{hemichrome}$) during bloodstain aging^{38–40}. When fresh blood was exposed to air, autoxidation of haemoglobin would occur immediately, followed by denaturation and aggregation

Frequency (cm ⁻¹)	Assignment
~931	Symmetric C-O stretching from carbohydrates
~972	Symmetric C-O stretching from carbohydrates
~1088	Symmetric vibration of PO ₂ ⁻
~1103	Symmetric C-O stretching from carbohydrates
~1165	C-O vibration
~1238	Asymmetric vibration of PO ₂ ⁻
~1302	Amide III band
~1392	Symmetric vibration of COO ⁻ of fatty acids and polysaccharides
~1452	C-H bending from CH ₃
~1533	Amide II band
~1649	α-Helical structures of proteins, amide I

Table 2. ATR FT-IR peak component assignment of bloodstains.

as time progressed⁴¹. These findings also helped to confirm that the process of bloodstain degradation started immediately and can be detected in a few hours and over a longer period of time.

Our results also demonstrated the capacity of the ATR-FTIR technique for detecting changes in minor components such as blood glucose (probable bands around 931, 972, and 1103 cm⁻¹), except the dominant spectral changes of haemoglobin and its derivative components of aging bloodstains. However, it is impracticable to estimate the age of a bloodstain with the selection of one or several absorption peaks by visualizing intensity changes because of the overlapping spectral features of bloodstain samples. Hence, in our next step, multivariate chemometric methods, which are powerful in the extraction and analyses of information-rich spectroscopic signals, were utilized to construct the model for bloodstain age estimation.

Age delimitation by the PLSR model. PLS regression analysis was performed with 12 and 14 LVs, to deliver satisfactory prediction performances and to build models for age estimation of indoor and outdoor bloodstains over the entire age period (0.25–107 d). Figure 2a and b illustrate the calibration results of the indoor and outdoor PLSR models, both of which exhibited good predictive ability as reflected by the R² of 0.96 and 0.98 and RMSEC of 5.97 and 4.73 d, respectively.

Internal cross-validation, as a routine method to determine the number of LVs, was also employed to evaluate the robustness of the calibration model. A stable prediction PLSR model is expected to have a high R² value and a low RMSECV value while achieving prediction performance comparable to that of the calibrated model.

The cross-validated results are presented in Table 3. The R² and RMSECV values were 0.94 and 7.51, respectively, for the indoor regression model and 0.96 and 6.31 for the outdoor regression model. Although the RMSECV values were slightly higher than those of RMSEC, the comparable values of R² demonstrated a good overall fit for the internal cross-validations; thus, our two PLSR models can be considered robust and reliable.

Due to the limited scale of training bloodstains and high intra-species biodiversity, external validation was subsequently conducted to assess the models' predictive power using bloodstain samples from two other donors (these donors were outside the training dataset). The external validation performances of the two PLSR models are summarized in Table 3. The high values of RPD and similar values of RMSE and R² of the external validations compared to those of the cross-validations indicated that our PLSR models could be considered very reliable for estimating bloodstain age under the same environmental conditions. Expanding tests were also performed to assess how well our established PLSR models predicted the age of bloodstains under multiple environmental conditions. However, the higher RMSEP values and poorer values of RPD and R² (see Table 3) indicated the unreliability of the established models to predict bloodstain age under various environmental conditions.

In the ideal linear regression with regard to Fig. 2, all spectra (symbols) should lie directly on the line of best fit (the green line), and the minimal spread should be within the symbols for each age point. However, it was observed by visualization that the bloodstain spectra in the “fresh” time period (0.25–7 d in indoor and outdoor PLSR models) exhibited relatively larger discrepancies than the spectral points at age onwards. In particular, the discrepancies were even larger at the 0.25-d and 1-d time points in the indoor PLSR model. It was also observed that the spectral points at 107 d were almost off the fitting line in the indoor PLSR model. This result indicated that the indoor PLSR model was not appropriate for estimating the age of bloodstains that were approaching 107 days old.

Given the great forensic importance of timely estimation of bloodstain age and the PLSR models' limited capacity for estimating the age of older bloodstains, in the next step we reconstructed indoor and outdoor PLSR models with split age periods (one age period from 0.25 to 7 days and one from 7 to 85 days before performing the external validation and expanding tests. The age period of the test samples was consistent with that of the training samples for calibration in this study. The calibrated and validated results of the models are presented in Fig. 3 and Table 3, respectively. As can be observed, the calibration and internal cross-validation statistical parameters for the outdoor model in the 7–85-d time period were slightly lower than those for the outdoor model in the entire time period. Nevertheless, the much better validated results—higher RPD value (5.14) and lower RMSEP value (4.77)—and the smaller number of LVs (10) demonstrated that our model was simple, robust and very reliable

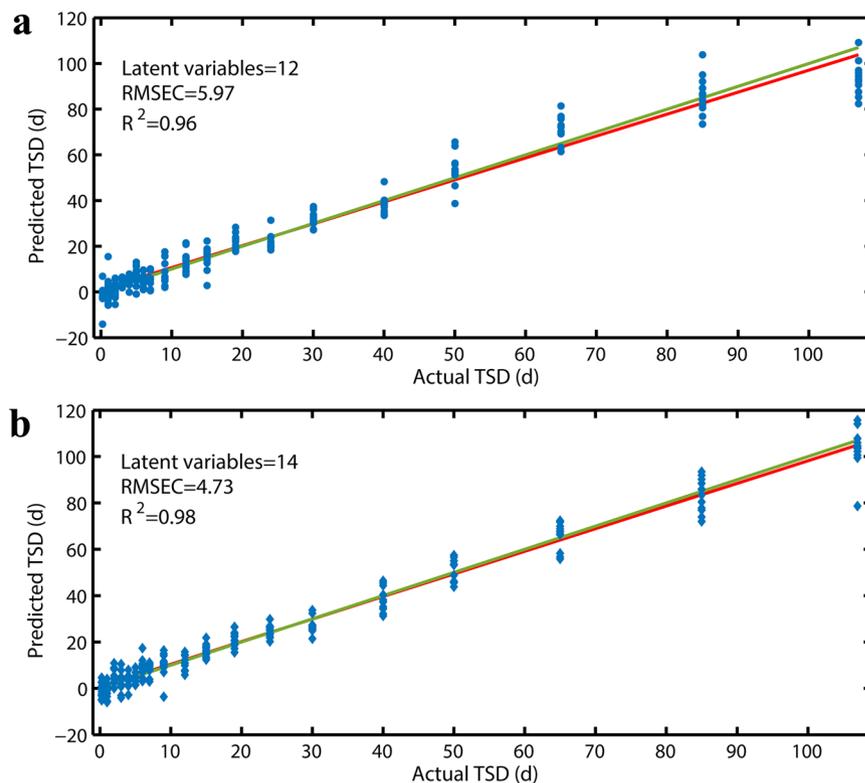


Figure 2. PLSR plots for (a) indoor and (b) outdoor bloodstain samples in the 0.25- to 107-d period showing the calibrated age predictions versus the actual age.

Time since deposition	Cross-validation		External validation			Expanding test		
	RMSECV	R ²	RMSEP	R ²	RPD	RMSEP	R ²	RPD
Indoor model								
0.25–7 d	1.20	0.72	1.18	0.72	1.90	1.53	0.53	1.19
7–85 d	5.88	0.94	5.83	0.94	4.08	12.93	0.71	1.73
0.25–107 d	7.51	0.94	7.24	0.94	4.20	13.34	0.80	2.24
Outdoor model								
0.25–7 d	0.91	0.84	1.10	0.76	2.09	2.15	0.08	0.99
7–85 d	6.35	0.93	4.77	0.96	5.14	19.99	0.31	0.70
0.25–107 d	6.31	0.96	6.43	0.95	4.42	23.61	0.38	0.77

Table 3. The validation results of PLSR models in the three time periods.

for prediction purposes. Better performance was also achieved with the indoor model in the 7- to 85-d time period; lower values of RMSEC (4.96), RMSECV (5.88), and RMSEP (5.83) indicated a higher accuracy of age predictions.

Our study also revealed that the indoor and outdoor PLSR models in the 0.25- to 7-d period were not appropriate for prediction of bloodstain age as reflected by the lower values of RPD (1.90 and 2.09, respectively). This apparent uncertainty in the early stage of the models was similar to that with the results published by Sun *et al.*⁴², who employed visible reflectance spectroscopy coupled with SVM to determine bloodstain age.

A possible explanation for the early-stage uncertainty is associated with the reaction kinetics of haemoglobin. According to the results obtained by Tsuruga *et al.*⁴³, the autoxidation process of HbO₂ can be divided into an initial fast decay and final slow decay. In one study by Bremmer *et al.*⁴⁰, the initial fast decay lasted a few hours and then transitioned to the slow decay. In another study by Bremmer *et al.*³⁸, the slow decay probably lasted ten days and entered into a slower decay phase subsequently. Additionally, the study results of the Bremmer research group demonstrated that oxidation rates of HbO₂ are strongly temperature-dependent and that the transition of met-Hb into hemichrome is strongly humidity-dependent⁴⁰. It is conceivable that the fluctuating temperature and humidity in both the indoor and outdoor environments where the bloodstains were stored resulted in the instability of the autoxidation process of HbO₂ and increased the complexity of the haemoglobin reaction kinetics in the 0.25- to 7-d period, which was probably corresponding to the early phase of slow decay. As a consequence, the variety and relative quantity of secondary structures of haemoglobin and its derivatives changed rapidly and irregularly,

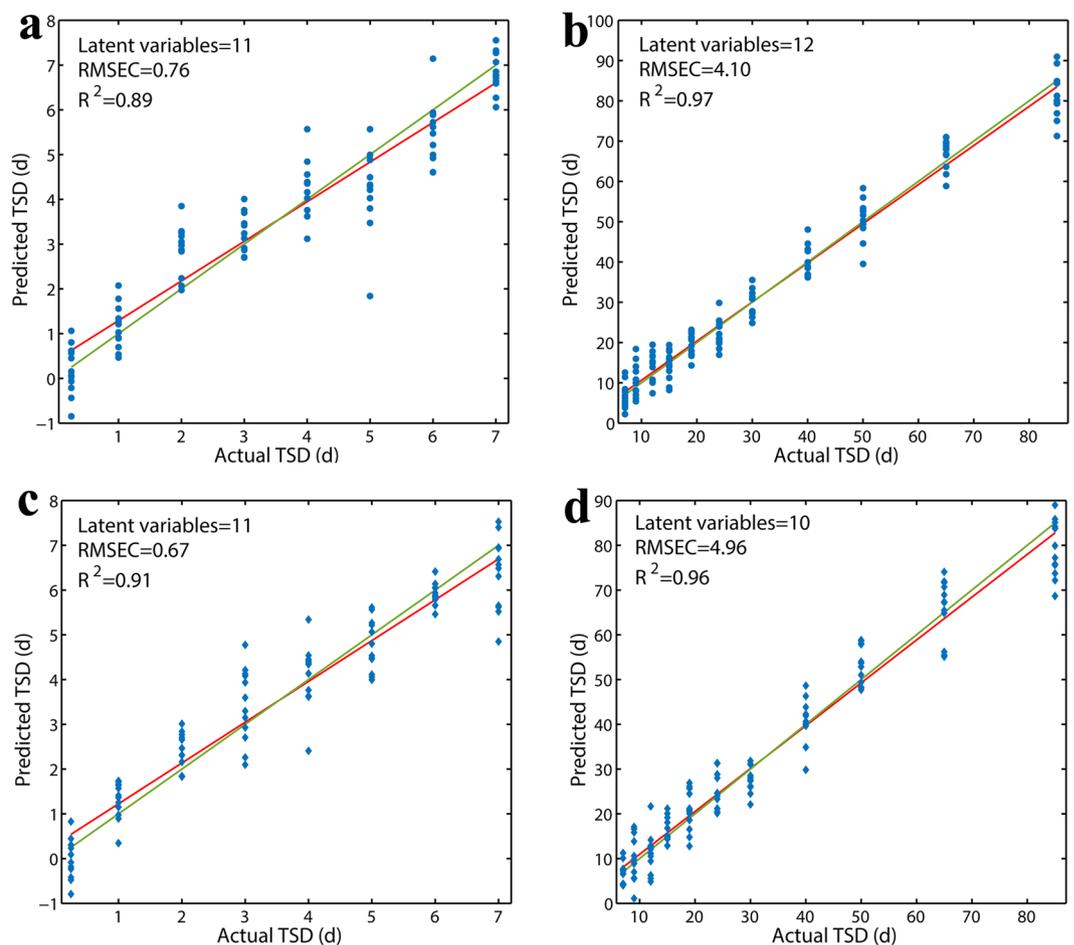


Figure 3. PLSR plots for indoor bloodstain samples in the (a) 0.25- to 7-d and (b) 7- to 85-d periods and outdoor bloodstain samples in the (c) 0.25- to 7-d and (d) 7- to 85-d periods showing the calibrated age predictions versus the actual age.

which in turn led to a relatively trendless variation of the corresponding spectral features (mainly amide regions; see Fig. 1) and resulted in a relatively large variation in the age prediction of bloodstains in the early time period (0.25-7 d). Additionally, the unreliability of all four reconstructed models (presented in Table 3), as well as that of the two previous PLSR models, in predicting bloodstain age under different environments in the entire time period showed that the contribution of environmental factors to the degradation of bloodstains was large.

Distinguishing between fresh and older bloodstains via PLS-DA. Two binary PLS-DA classification models (indoor and outdoor models) were developed using spectra originating from 228 indoor training samples and 228 outdoor training samples. Each spectrum was classified as either a fresh (age ≤ 1 d) or older (age > 1 d) bloodstain. The models were built with 8 and 9 LVs, respectively—the minimum numbers of LVs that delivered satisfactory classification. As seen from Fig. 4a and b, both models demonstrated good separation between these two classes. All the spectra belonging to the fresh bloodstains were classified as the fresh-bloodstain class, and spectra for only two indoor older bloodstains and four outdoor older bloodstains were misclassified in their respective model. The accuracies of the models were 0.99 and 0.98, respectively.

Receiver operating characteristic (ROC) analyses⁴⁴ were conducted to evaluate the discriminatory power of our PLS-DA classification models. The ROC curve was plotted as a function of sensitivity versus 1-specificity (see Fig. 4c and d). The area under the ROC curve (AUC)⁴⁴ was calculated to assess how well the classification model divided the samples to the positive class. AUC has values in the interval [0, 1], where a value of 0.5 means a random classification and 1 means perfect performance. For our two models developed to differentiate between fresh and older bloodstains under indoor and outdoor environments, the AUC values of ROC curves were 1 and 0.9996, respectively, which confirmed the classification capabilities of the models. Validations were also performed to evaluate the reliability and classification ability of the classification models in three different manners. The validated results are summarized in Table 4.

Classification performances of internal cross-validation and external validation were as perfect as that of the calibrated PLS-DA models. The classification parameters such as accuracy rate, class sensitivity and specificity were close to 1, similar to those obtained in the calibrated models, indicating that these two calibrated PLS-DA models were robust, reliable and well-fitted for classification purposes. The expanding test was used to attest the

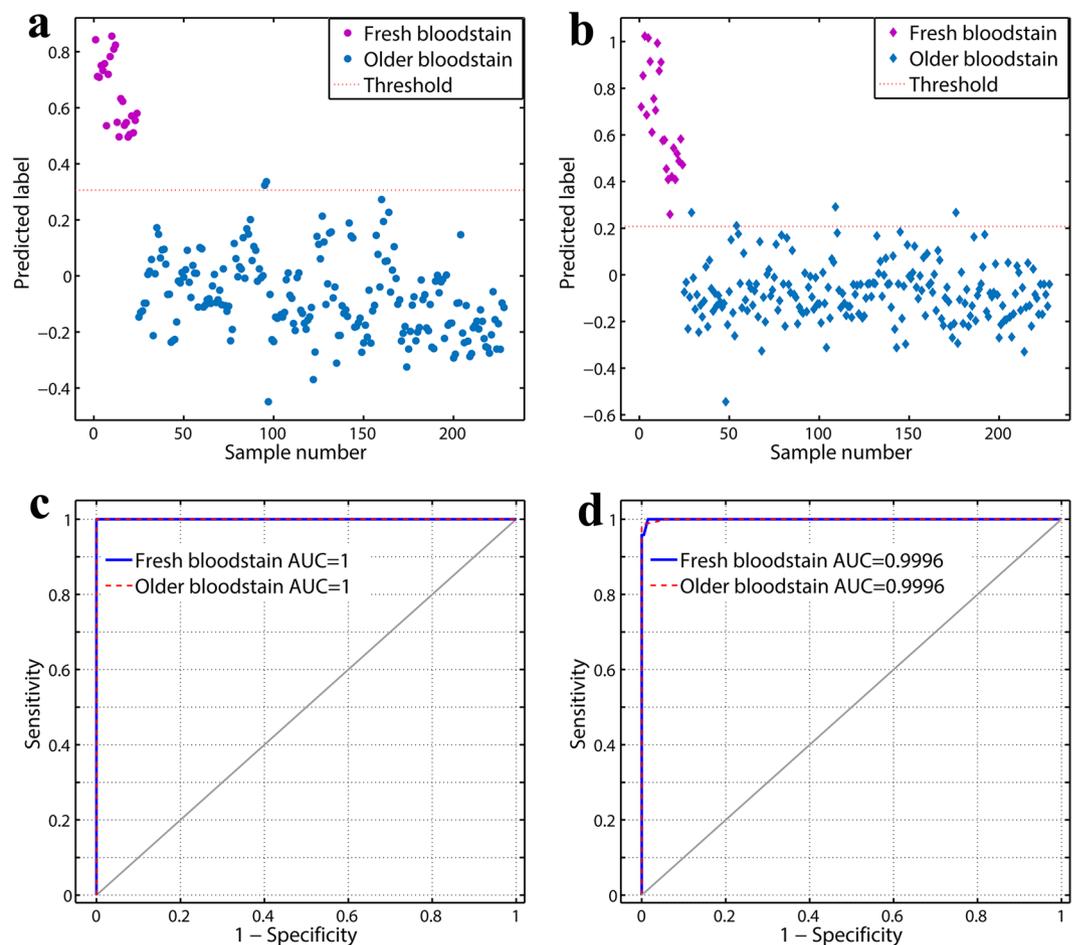


Figure 4. (a) Prediction scores of the indoor training dataset using the indoor PLS-DA model. (b) Prediction scores of the outdoor training dataset using the outdoor PLS-DA model. The red dotted line represents the default classification threshold. ROC curves with AUC for fresh and older bloodstain classes in the (c) indoor and (d) outdoor PLS-DA classification models. Random choice is denoted by the grey diagonal line.

	Accuracy	Fresh bloodstain (age \leq 1 d)		Older bloodstain (age $>$ 1 d)	
		Sensitivity	Specificity	Sensitivity	Specificity
Indoor model					
Cross-validation	0.99	0.99	1	1	0.99
External validation	0.99	1	0.99	0.99	1
Expanding test	0.92	0.25	1	1	0.25
Outdoor model					
Cross-validation	0.96	0.96	0.96	0.96	0.96
External validation	0.99	1	0.99	0.99	1
Expanding test	0.85	0.92	0.84	0.84	0.92

Table 4. PLS-DA classification parameters obtained in the cross-validation, external validation and expanding tests.

ability of each PLS-DA model to discriminate between fresh and older bloodstains in different environments. With regard to discriminating fresh bloodstains, the outdoor classification model showed better performance with a sensitivity value of 0.92 (only one fresh bloodstain from the indoor environment was misclassified) compared with the indoor classification model for classifying outdoor fresh bloodstains (the sensitivity value was only 0.25). The characteristics of the PLS-DA classification models, as well as the aforementioned PLSR models, were defined by each LV's loading variable, which contained numerous peaks throughout the spectral "fingerprint region" (1800-900 cm^{-1}). In other words, spectral information related to all chemical components of a bloodstain during aging contributed to the constructions of the models. This was in accordance with the degradation of bloodstains, which incorporates oxidation of haemoglobin, RNA degradation, and degradation of serum proteins

of blood plasma and other blood components². Therefore, this is one feature of our approach that can easily probe the spectroscopic statistical differentiation of the chemical components of bloodstain samples without knowing the specific components.

In conclusion, ATR-FTIR spectroscopy is rapid, easy to use, and non-destructive—properties that are favourable in forensic practice. Its application in bloodstain identification and species determination has been reported previously^{21,45}. Nevertheless, to the best of our knowledge, this is the first study demonstrating that ATR-FTIR spectroscopy can be a valuable tool for estimating bloodstain age in mimicked indoor and outdoor crime scenes. Chemometric analysis proved to be powerful for extracting and analysing the universal biospectral information of bloodstains with aging and establishing prediction models for age estimation.

Notably, our approach was more useful for longer-term (7–85 d) estimation of the age of bloodstains regardless of whether they were in an indoor or outdoor environment. The rough performance of our PLSR models in predicting the age of bloodstains in the 0.25- to 7-d time period was partly compensated for by two PLS-DA classification models, which could easily discriminate fresh (age ≤ 1 d) bloodstains from older (age > 1 d) bloodstains in both indoor and outdoor environments. This discrimination was a key finding of our study, and it is highly desirable because it can be applied to forensic practices to help reconstruct a more realistic timeline of events. Nevertheless, prior to applying our approach in real forensic practice, more work needs to be done. Expanding the number of donors, determining the effect of common substrates and contaminations, and developing a robust chemometric framework are important tasks for the future studies.

References

- James, S. H., Kish, P. E. & Sutton, T. P. *Principles of bloodstain pattern analysis: theory and practice*. (CRC Press, 2005).
- Bremmer, R. H., De Bruin, K. G., Van Gemert, M. J. C., Van Leeuwen, T. G. & Aalders, M. C. G. Forensic quest for age determination of bloodstains. *Forensic Sci. Int.* **216**, 1–11, <https://doi.org/10.1016/j.forsciint.2011.07.027> (2012).
- Matsuoka, T., Taguchi, T. & Okuda, J. Estimation of bloodstain age by rapid determinations of oxyhemoglobin by use of oxygen electrode and total hemoglobin. *Biol Pharm Bull* **18**, 1031–1035 (1995).
- Fujita, Y. *et al.* Estimation of the age of human bloodstains by electron paramagnetic resonance spectroscopy: Long-term controlled experiment on the effects of environmental factors. *Forensic Sci. Int.* **152**, 39–43, <https://doi.org/10.1016/j.forsciint.2005.02.029> (2005).
- Inoue, H., Takabe, F., Iwasa, M., Maeno, Y. & Seko, Y. A new marker for estimation of bloodstain age by high performance liquid chromatography. *Forensic Sci. Int.* **57**, 17–27 (1992).
- Anderson, S., Howard, B., Hobbs, G. R. & Bishop, C. P. A method for determining the age of a bloodstain. *Forensic Sci. Int.* **148**, 37–45 (2005).
- Bauer, M., Polzin, S. & Patzelt, D. Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Sci. Int.* **138**, 94–103 (2003).
- Agudelo, J., Huynh, C. & Halánek, J. Forensic determination of blood sample age using a bioaffinity-based assay. *The Analyst* **140**, 1411–1415, <https://doi.org/10.1039/C4AN02269F> (2015).
- Agudelo, J. *et al.* Ages at a Crime Scene: Simultaneous Estimation of the Time since Deposition and Age of Its Originator. *Anal. Chem.* **88**, 6479–6484, <https://doi.org/10.1021/acs.analchem.6b01169> (2016).
- Guo, K., Achilefu, S. & Berezin, M. Y. Dating bloodstains with fluorescence lifetime measurements. *Chemistry* **18**, 1303–1305 (2012).
- Shine, S. M., Suhling, K., Beavil, A., Daniel, B. & Frascione, N. The applicability of fluorescence lifetime to determine the time since the deposition of biological stains. *Anal. Methods* **00**, 1–7, <https://doi.org/10.1039/C6AY03099H> (2017).
- Strasser, S. *et al.* Age determination of blood spots in forensic medicine by force spectroscopy. *Forensic Sci. Int.* **170**, 8–14, <https://doi.org/10.1016/j.forsciint.2006.08.023> (2007).
- Shin, J. *et al.* Smart Forensic Phone: Colorimetric analysis of a bloodstain for age estimation using a smartphone. *Sens. Actuators, B* **243**, 221–225, <https://doi.org/10.1016/j.snb.2016.11.142> (2017).
- Muro, C. K., Doty, K. C., Bueno, J., Halámková, L. & Lednev, I. K. Vibrational spectroscopy: recent developments to revolutionize forensic science. *Anal. Chem.* **87**, 306–327 (2015).
- Balabin, R. M. & Smirnov, S. V. Variable selection in near-infrared spectroscopy: benchmarking of feature selection methods on biodiesel data. *Anal. Chim. Acta* **692**, 63–72 (2011).
- Byrne, H. J., Knief, P., Keating, M. E. & Bonnier, F. Spectral pre and post processing for infrared and Raman spectroscopy of biological tissues and cells. *Chem. Soc. Rev.* **45**, 1865–1878 (2016).
- Gajjar, K. *et al.* Fourier-transform infrared spectroscopy coupled with a classification machine for the analysis of blood plasma or serum: a novel diagnostic approach for ovarian cancer. *Analyst* **138**, 3917–3926 (2013).
- Kelly, J. G. *et al.* Biospectroscopy to metabolically profile biomolecular structure: a multistage approach linking computational analysis with biomarkers. *J. Proteome Res.* **10**, 1437–1448 (2011).
- Khoshmanesh, A. *et al.* Detection and Quantification of Early-Stage Malaria Parasites in Laboratory Infected Erythrocytes by Attenuated Total Reflectance Infrared Spectroscopy and Multivariate Analysis. *Anal. Chem.* **86**, 4379–4386 (2014).
- McLaughlin, G., Doty, K. C. & Lednev, I. K. Raman spectroscopy of blood for species identification. *Anal. Chem.* **86**, 11628–11633, <https://doi.org/10.1021/ac5026368> (2014).
- Mistek, E. & Lednev, I. K. Identification of species' blood by attenuated total reflection (ATR) Fourier transform infrared (FT-IR) spectroscopy. *Anal. Bioanal. Chem.* **407**, 7435–7442 (2015).
- McLaughlin, G. & Lednev, I. K. Spectroscopic Discrimination of Bone Samples from Various Species. *Am. J. Anal. Chem.* **3**, 161–167 (2012).
- Cascant, M. M. *et al.* Burned bones forensic investigations employing near infrared spectroscopy. *Vib. Spectrosc.* **90**, 21–30 (2017).
- Grobério, T. S., Zacca, J. J., Botelho, É. D., Talhavi, M. & Braga, J. W. Discrimination and quantification of cocaine and adulterants in seized drug samples by infrared spectroscopy and PLSR. *Forensic Sci. Int.* **257**, 297–306 (2015).
- Zhang, J. *et al.* Characterization of postmortem biochemical changes in rabbit plasma using ATR-FTIR combined with chemometrics: A preliminary study. *Spectrochim. Acta, Part A* **173**, 733–739, <https://doi.org/10.1016/j.saa.2016.10.041> (2017).
- Wang, Q. *et al.* UV-Vis and ATR-FTIR spectroscopic investigations of postmortem interval based on the changes in rabbit plasma. *PLoS one* **12**, e0182161 (2017).
- Engel, J. *et al.* Breaking with trends in pre-processing? *TrAC, Trends Anal. Chem.* **50**, 96–106 (2013).
- Wold, S., Sjöström, M. & Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **58**, 109–130 (2008).
- Ballabio, D. & Consonni, V. Classification tools in chemistry. Part 1: linear models. PLS-DA. *Anal. Methods* **5**, 3790–3798 (2013).
- Faber, N. M. & Rajkó, R. How to avoid over-fitting in multivariate calibration—the conventional validation approach and an alternative. *Anal. Chim. Acta* **595**, 98–106 (2007).
- Cozzolino, D., Cynkar, W. U., Shah, N. & Smith, P. Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality. *Food Res. Int.* **44**, 1888–1896 (2011).

32. Cozzolino, D. *et al.* Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy. *Anal. Chim. Acta* **513**, 73–80 (2004).
33. Trevisan, J., Angelov, P. P., Carmichael, P. L., Scott, A. D. & Martin, F. L. Extracting biological information with computational analysis of Fourier-transform infrared (FTIR) biospectroscopy datasets: current practices to future perspectives. *Analyst* **137**, 3202–3215 (2012).
34. Baker, M. J. *et al.* Developing and understanding biofluid vibrational spectroscopy: a critical review. *Chem. Soc. Rev.* **45**, 1803–1818, <https://doi.org/10.1039/c5cs00585j> (2016).
35. And, S. C. & Singh, B. R. A Distinct Utility of the Amide III Infrared Band for Secondary Structure Estimation of Aqueous Protein Solutions Using Partial Least Squares Methods. *Biochemistry* **43**, 2541–2549 (2004).
36. Caine, S., Heraud, P., Tobin, M. J., McNaughton, D. & Bernard, C. C. The application of Fourier transform infrared microspectroscopy for the study of diseased central nervous system tissue. *NeuroImage* **59**, 3624–3640, <https://doi.org/10.1016/j.neuroimage.2011.11.033> (2012).
37. Perez-Guaita, D. *et al.* Protein determination in serum and whole blood by attenuated total reflectance infrared spectroscopy. *Anal. Bioanal. Chem.* **404**, 649 (2012).
38. Bremmer, R. H., Nadort, A., van Leeuwen, T. G., van Gemert, M. J. C. & Aalders, M. C. G. Age estimation of blood stains by hemoglobin derivative determination using reflectance spectroscopy. *Forensic Sci. Int.* **206**, 166–171, <https://doi.org/10.1016/j.forsciint.2010.07.034> (2011).
39. Edelman, G., van Leeuwen, T. G. & Aalders, M. C. G. Hyperspectral imaging for the age estimation of blood stains at the crime scene. *Forensic Sci. Int.* **223**, 72–77, <https://doi.org/10.1016/j.forsciint.2012.08.003> (2012).
40. Bremmer, R. H. *et al.* Biphasic Oxidation of Oxy-Hemoglobin in Bloodstains. *Plos One* **6**, e21845 (2011).
41. Doty, K. C., McLaughlin, G. & Lednev, I. K. A Raman spectroscopic clock for bloodstain age determination: the first week after deposition. *Anal. Bioanal. Chem.* **408**, 3993–4001, <https://doi.org/10.1007/s00216-016-9486-z> (2016).
42. Sun, H. *et al.* Accurate Age Estimation of Bloodstains Based on Visible Reflectance Spectroscopy and Chemometrics Methods. *IEEE Photonics J.* **9**, 1–14, <https://doi.org/10.1109/JPHOT.2017.2651580> (2017).
43. Tsuruga, M., Matsuoka, A., Hachimori, A., Sugawara, Y. & Shikama, K. The molecular mechanism of autoxidation for human oxyhemoglobin. Tilting of the distal histidine causes nonequivalent oxidation in the beta chain. *J. Biol. Chem.* **273**, 8607–8615 (1998).
44. Zweig, M. H. & Campbell, G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical chemistry* **39**, 561–577 (1993).
45. Elkins, K. M. Rapid Presumptive “Fingerprinting” of Body Fluids and Materials by ATR FT-IR Spectroscopy. *J. Forensic Sci.* **56**, 1580–1587 (2011).
46. Li, B., Beveridge, P., O’Hare, W. T. & Islam, M. The estimation of the age of a blood stain using reflectance spectroscopy with a microspectrophotometer, spectral pre-processing and linear discriminant analysis. *Forensic Sci. Int.* **212**, 198–204, <https://doi.org/10.1016/j.forsciint.2011.05.031> (2011).
47. Edelman, G., Manti, V., Van Ruth, S. M., Van Leeuwen, T. & Aalders, M. Identification and age estimation of blood stains on colored backgrounds by near infrared spectroscopy. *Forensic Sci. Int.* **220**, 239–244, <https://doi.org/10.1016/j.forsciint.2012.03.009> (2012).
48. Li, B., Beveridge, P., O’Hare, W. T. & Islam, M. The age estimation of blood stains up to 30 days old using visible wavelength hyperspectral image analysis and linear discriminant analysis. *Sci. Justice* **53**, 270–277, <https://doi.org/10.1016/j.scijus.2013.04.004> (2013).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos 81730056, 81722027, 81601645, 81471819, 81273339 and 81671869).

Author Contributions

Z.Y.W., P.H., and H.C.L. designed this research project. H.C.L. and Y.M.Z. performed the experiments. H.C.L., Q.W., and B.L. analyzed the results. H.C.L. wrote this manuscript and prepared Figures 1–4 and Tables 1–4. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-017-13725-1>.

Competing Interests: The authors declare that they have no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017