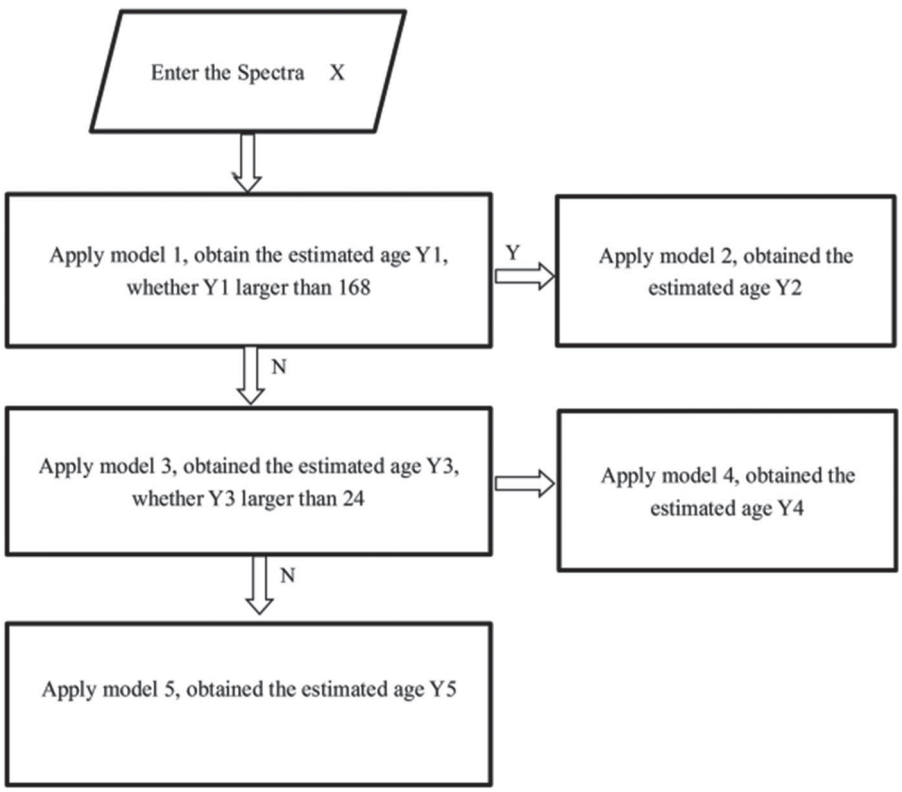
IEEE Photonics Journal
An IEEE Photonics Society Publication

Open Access

Accurate Age Estimation of Bloodstains Based on Visible Reflectance Spectroscopy and Chemometrics Methods

Volume 9, Number 1, February 2017

Huimin Sun
Yongfang Dong
Pingli Zhang
Yaoyong Meng
Wei Wen
Nan Li
Zhiyou Guo



DOI: 10.1109/JPHOT.2017.2651580
1943-0655 © 2017 IEEE

IEEE photonics SOCIETY
IEEE

Accurate Age Estimation of Bloodstains Based on Visible Reflectance Spectroscopy and Chemometrics Methods

Huimin Sun,[1] Yongfang Dong,[1] Pingli Zhang,[1] Yaoyong Meng,[1] Wei Wen,[1] Nan Li,[1] and Zhiyou Guo[2]

[1]MOE Key Laboratory of Laser Life Science and Laboratory of Photonic Chinese Medicine, College of Biophotonics, South China Normal University, Guangdong 510631, China
[2]Institute of Optoelectronic Material and Technology, South China Normal University, Guangdong 510631, China

DOI:10.1109/JPHOT.2017.2651580
1943-0655 © 2017 IEEE. Translations and content mining are permitted for academic research only. Personal use is also permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Manuscript received December 7, 2016; revised January 4, 2017; accepted January 7, 2017. Date of publication January 11, 2017; date of current version February 15, 2017. This work was supported by the Science and technology project of Guangdong Province, China, under Grant 2014B010121001, Grant 2014B010119004, Grant 2013B010204065; the Special Project for Key Science and Technology of Zhongshan City, Guangdong Province, China, under Grant 2014A2FC204; and the Science and Technology Program Project in Huadu District of Guangzhou City, China, under Grant HD15PT003. Corresponding author: Yaoyong Meng (e-mail:yaoyongmeng@aliyun.com).

Abstract: Visible reflectance spectroscopy technique combined with three chemometrics methods were applied to predict the age of bloodstains for forensic purposes. The performance of models established by principal component regression (PCR), partial least-squares regression (PLSR), and least-squares-support vector machines (LS-SVM) methods were compared. Three models, built on three different age periods from 2 to 24 h, one to seven days, and seven to 45 days, were implemented to improve the predictive accuracy. The performance of three LS-SVM models are much better than that of PCR models and just a little better than that of PLSR models. Considering the effect of the specificity of bloodstains on model, LS-SVM model, built on the age period from 2 hours to 45 days, achieved correlation coefficient of the prediction set (Rp) = 0.9900, root mean square error of prediction (RMSEP) = 42.7920 hours, residual predictive deviation (RPD) = 7.6709. Models built on three age periods could significantly improve the predictive capability. The results demonstrate that the visible reflectance spectroscopy combined with LS-SVM would be a reliable tool to accurately estimate the age of bloodstains for forensic practical application.

Index Terms: Age estimation, chemometrics methods, visible reflectance spectroscopy, bloodstains, forensic.

1. Introduction

Bloodstains, which are found most frequently at crime scenes, provide enormous forensic value in many forensic investigations, such as pattern reconstruction and DNA analysis [1]. Accurately estimating age of bloodstains can be used to determine the time when a crime was committed, which is particularly significant in case bloodstains are the only evidence available [2]. The estimated age of a bloodstain would enable investigator to verify witnesses' statements, to set terms of reference for an alibi or to judge whether a bloodstain is crime-related. In 1930, Schwarzacher attempted to

explore the relationship between the age of bloodstains and the solubility of bloodstains in water [3]. Recently, various spectroscopy techniques, including oxygen electrode [4], RNA degradation [5], [6], electron paramagnetic resonance spectroscopy [7], [8], high performance liquid chromatography [9], [10], near infrared (NIR) spectroscopy [11], atomic force microscopy (AFM) [12], hyperspectral imaging [2], [13], Raman spectroscopy [14], [15], and visible reflectance spectroscopy [16]–[18], have been proposed for estimating the age of bloodstains. However, most of methods need sophisticated, expensive equipment and have disadvantage of destructing specimens, which limit their implementation in forensic practice.

The color of a fresh blood stain changes from red to dark brown since the haemoglobin (HbO_2) would convert to met-haemoglobin (met-Hb) and then to hemichrome (HC) [14]. Enlightened by the color change, visible reflectance spectroscopy technique has been implemented to determine the age of bloodstains [14], [16]. Moreover, with the distinguished advantages of non-destructive, non-contact, simple, low cost, visible reflectance spectroscopy technique attracted highly attention in the forensic investigation community. Recently, Bremember *et al.* have proposed diffuse reflectance spectroscopy for estimating the age of bloodstains from 0 to 60 days [14]. Based on the reference absorption spectra of three hemoglobin derivatives, the average fractions of HbO_2, met-Hb and HC in the training experimental spectra were estimated by using a linear least squares. The age of bloodstains in the test set was estimated on the basis of three above average fractions. The estimated age of a bloodstain with an actual age of 35 days varied between 25 and 55 days. Li *et al.* used a microspectropotometer (MSP) to record visible reflectance spectra of bloodstains and established a linear discriminant analysis (LDA) model to predict the age of bloodstains [16]. The predictive results demonstrated that the predictive accuracy decreased to 37.3% when a separate bloodstain was used for the test dataset. The effect of specificity of bloodstains withdrawn from the same individual [15] on model was not taken into account. Obviously, both of these methods did not take the specificity of bloodstains withdrawn from different individual into account and resulted in large predictive errors. It is noted that an effective statistics method could play a crucial role in optimizing predictive ability [19]–[21]. However, statistics methods used in previous studies, such as linear discriminant analysis [13], [18] and linear least squares fitting [16], did not build a robust model. Due to the poor effectiveness of statistics method, the accurate age estimation of bloodstains remains an unsolved situation in forensic practice [22]. Hence, in order to solve this problem, it is urgent to build a powerful multivariate statistics model with the specificity of bloodstains considered.

The factor analysis methods, especially principal component regression (PCR) and partial least square regression (PLSR), as most widely and powerfully used multivariate chemometrics methods, have been successfully applied to the quantitative analysis of spectroscopic data [23]–[29]. The support vector machine (SVM), as a non-linear multivariate data analysis method, is widely used for pattern regression [28] and predicting future data [30]. The least squares support vector machine (LS-SVM), which is a re-formulations of the standard SVM, can reduce the computational complexity and improve the accuracy of the regression [31], [32]. Therefore, a question naturally arises as whether the powerful statistics methods could be employed to improve the accuracy of age estimation of bloodstains further.

In this work, we applied visible reflectance spectroscopy combined with three multivariate statistical regression techniques to accurately estimating the age of bloodstains. Models, established in different age regions, were used to accurately estimate the age of bloodstains and improve the predictive capabilities. In order to obtain the model, the prediction performance of models, built by methods of PLSR, PCR, LS-SVM, were evaluated and compared.

2. Methods and Materials

2.1 Sample Preparation and Reflectance Spectroscopy

Eight blood samples were withdrawn from eight healthy volunteers between 8:00 to 8:20 a.m. Eight bloodstains were prepared by placing a small 15 μL drop on a glass slide. Eight bloodstains were stored in calorstats where temperature is 37 °C, and the relative humidity was about 10%.

The reflectance spectra were collected using a USB-4000 spectrometer (Ocean Optics Inc., USA). The spectrometer, equipped with a fiber optic reflection probe (R400-7-Vis/NIR), has a high

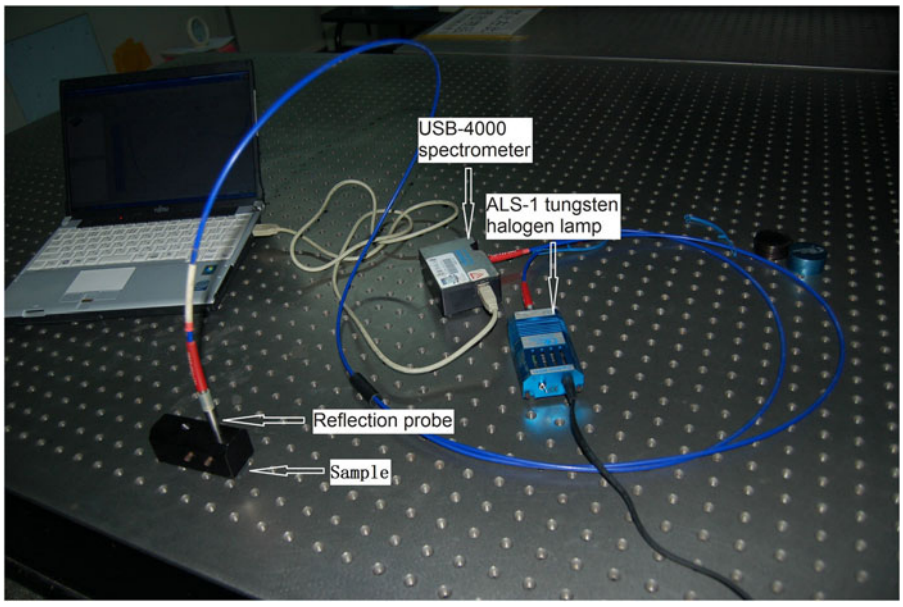


Fig. 1. Experimental setup.

sensitivity range from 200 nm to 1100 nm at a sampling interval of 0.22 nm. Besides, an ALS-1 tungsten halogen lamp (3.6W output) was used as light source. Before collecting the reflection spectra, the white standard was implemented to balance the spectrum. Spectra were saved as percent reflection relative to a reflection standard. The reflection spectra of eight bloodstains were measured at 2h, 4h, 8h, 12h, 24h, 36h, 48h, 3d, 7d, 10d, 15d, 30d, and 45d and 104 measurements in total. The photo of the complete experimental setup is presented in Fig. 1.

2.2 Chemometrics Analysis

2.2.1. Spectral Pre-Processing: The reflectance spectra can be utilized for gaining information about the chemical composition of a specimen. However, several non-chemical factors, including the particle size of the red blood cells [33], the roughness and thickness of the bloodstains [16], were collected at the same time, which can induce systematic variations between spectra. Hence, it is necessary to apply suitable spectral pre-processing techniques to reduce or minimize variations, such as standard normal variate correction (SNV) [16], multiplicative scattering correction (MSC) [34] and Savitzky-golay differentiation [35]. It was difficult to obtain an essential common reference signal for MSC, thus, SNV was selected to remove variations resulting from scattering effects and path length variations [36]. The following equation shows the mathematical transformation required for SNV [16].

$$x_{corr} = \frac{\left|x_{org} - a_0\right|}{a_1} \tag{1}$$

where a_0 is the average value of the sample spectrum to be corrected, a_1 is the standard deviation of the sample spectrum, X_{corr} is the spectrum after SNV transformation, and X_{org} is the original spectrum.

2.2.2. Chemometrics Methods: In this work, three chemometrics methods, including partial least square regression, principal component regression and least square support vector machines, were used to establish models for estimating the age of bloodstains.

2.2.2.1. Partial Least Square Regression: In this work, PLSR was used to established mathematical model for accurately estimating the age of bloodstains. PLSR method undertakes to clarify the latent variables (LVs) which account for the systematic majority of variation in spectral data versus the age of bloodstains. The first several LVs usually account for the most amounts of the

spectra and age variances. Leave-one-out cross validation was implemented to determine the optimum numbers of LVs.

2.2.2.2. Principal Component Regression: Principal component regression (PCR) is a two–step procedure, where the first step is to decompose the spectral data by a principal component analysis; the second step is to fit a multiple linear regression model with a small number of principal components (PCs). The principal components mainly account for most variances in the observed variables instead of the original variables as predictors [20], [37], [38]. Leave-one-out cross validation was used to determine the optimal number of PCs.

2.2.2.3. Least Square Support Vector Machines: Least-square support vector machines (LS-SVM), as a modified classical SVM algorithm, which is capable of dealing with linear and nonlinear multivariate calibration in a relatively fast way [39]. LS-SVM regression model can be expressed as follows [21]:

$$y(x) = \sum_{i=1}^{n} a_i K(x, x_i) + b \tag{2}$$

where $K(x, x_i)$ is the kernel function, x_i is the input vector, a_i is the Lagrange multipliers called support value, and b is the bias term.

In this work, the Radial basis function (RBF) is selected as the kernel function to construct the LS-SVM models, and its formula can be expressed as follows:

$$K(x, x_i) = \exp(-\|x - x_i\|^2 / 2\sigma^2). \tag{3}$$

The implementation of LS-SVM requires the specification of two free parameters, σ and γ, which influence the performance of LS-SVM model and control the risking of over-fitting and the complexity of the boundary [30], [40].

2.3 Model Reliability

A combination of root mean square error (RMSE), correlation coefficient (R) and the value of residual predictive deviation (RPD) are used to evaluate the validity and the prediction capability of the model. Root mean square error of calibration (RMSEC) and root mean square error of prediction (RMSEP) is defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum (y_{pred} - y_{ref})^2}{N}} \tag{4}$$

where y_{pred} is the predicted value, y_{ref} is the laboratory measured value, and N is the number of samples.

Correlation coefficients between the predicted and the measured values are calculated for the calibration set (Rc) and the prediction set (Rp), which are calculated as follows:

$$R = \sqrt{1 - \frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{5}$$

where $\hat{y}_i$, y_i are the predicted and measured value of sample i in the calibration set and the prediction set, $\bar{y}$ is the mean of the reference measurement results of all samples in the calibration set and prediction set and n is the number of observation in calibration and prediction set.

The value of residual predictive deviation (RPD) is used to evaluate the robustness of a model and a relative high RPD value indicate model having greater power for prediction purpose [41]. It was defined as follows:

$$\mathrm{RPD} = \frac{\mathrm{S.D.}}{\mathrm{RMSEP}}. \tag{6}$$

S.D. is the standard deviation for the prediction set. Generally, an RPD value greater than 3 could be considered that the model is very well for prediction purposes [37], [41].

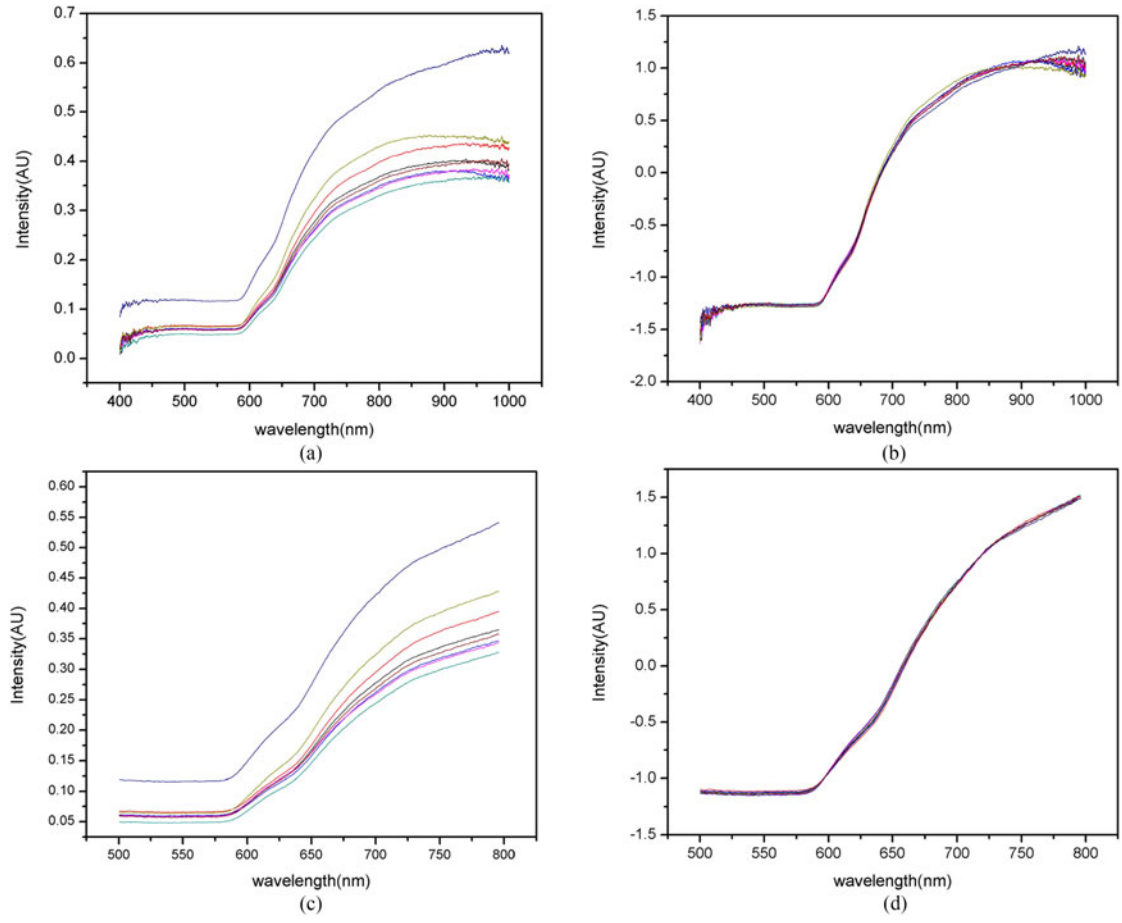


Fig. 2. Spectral of eight bloodstains at the same age before and after pre-processing. (a) Reflectance spectra on the edge region of 400–1000 nm, (b) spectra after applying standard normal variate correction, (c) reflectance spectra on the edge region of 500–780 nm, and (d) spectra after applying standard normal variate correction.

2.4 Software

Custom data analysis software created in MATLAB2012b (Math Work, USA) was used for all the data processing including spectral pre-processing, PLSR, PCR, and LS-SVM analysis.

3. Results and Discussion

Eight bloodstains were randomly divided into two sets in a proportion of 3 to 1. Calibration set, consisting of 78 spectra which were measured from six bloodstains, was used to establish the PCR, PLSR and LS-SVM models. In addition, prediction set, including 26 spectra which were measured from two bloodstains, was used for testing the robustness of the models. Generally, a good model should have lower RMSEC and RMSEP, and higher R and RPD but small differences between RMSEC and RMSEP.

3.1 Spectral Pre-Processing

In this paper, SNV was selected as the method for spectral pre-processing. The effect of spectral pre-processing on reflectance spectra was demonstrated in Fig. 2. The left panes showed the raw spectra measured from eight bloodstains at the age of 8 hours. As can be seen, the spectra with

TABLE 1
PLSR Models With Different Spectral Regions and Pretreatments

Age periods	Spectral regions (nm) and pretreatments	Calibration		Prediction			LVs
		R_c	RMSEC(h)	R_p	RMSEP(h)	RPD	
2h-45d	Full + SNV	0.9847	55.4718	0.9542	95.2676	3.4083	8
2h-45d	500–780 + SNV	0.9848	55.3774	0.9838	57.1007	5.6865	8
2h-45d	400–500 + SNV	0.5660	262.4778	0.3204	301.6134	1.0765	3
2h-45d	500–780	0.7620	206.1949	0.2407	309.0350	1.0507	4
2h-45d	Full	0.9784	65.8375	0.9354	112.5478	2.8852	8
2h-24h	500–780 + SNV	0.9935	0.8868	0.9928	0.9365	9.7768	6
2h-7d	500–780 + SNV	0.9820	9.4044	0.9849	8.6153	5.9413	7
24h-7d	500–780 + SNV	0.9968	4.1280	0.9881	7.7423	7.0384	7
7d-45d	500–780 + SNV	0.9946	36.9566	0.9880	52.6722	6.8236	6

the same age showed relative large differences, which were caused by the baseline shifts and scattering effects. After pre-processing, the spectral variations were greatly removed, as shown in the right panes. Through the comparison of the two panes, it was concluded that SNV method could effectively remove the variations resulted from baseline shifts and scattering effects.

3.2 PLSR Model

In general, the number of LVs more than 10 is correlated with a high risk of over-fitting [42], [43]. Therefore, the number of LVs used in PLSR models was no more than 10.

The performance of PLSR models on three fragments of 400-500, 500-780, 780-1000 nm, as well as full spectra were compared. The results can be seen in the Table 1. The model built on fragment of 500-780 nm presented better predictive ability than those on other regions. The predictive ability of PLSR model built on the edge region of 500-780 nm was obviously improved, while, the predictive ability of models built on the edge regions of 400-500 nm and 780-1000 nm were decreased, which indicated that the spectral data on the two edge regions contain too much noise. Table 1 also showed the performance of the models on fragment of 500-780 nm. It can be observed that the pretreatment of SNV could increase the predictive accuracy of models. The effect of LVs on RMSECV values was visualized in Fig. 3(a). Taking the lowest RMSECV into account, the best LVs number was selected as 8. The performance of the model was shown in line 2 of Table 1 with Rc = 0.9848, Rp = 0.9838, RMSEP = 57.1007 hours and RPD = 5.6865. The model with good robustness was considered acceptable for predicting the age because that the value of the RPD was greater than 3. However, as seen from the Fig. 4 (A1), the relative predicted error was large when the age of bloodstains ranged between 2 hours and 7 days. Thus, models built on different age periods were introduced to solve this problem.

Two models were established on two age periods from 2 hours to 7 days and 7 days to 45 days respectively. As presented in the Fig. 3(e) and Fig. 3(d), the best latent variables of two models were selected as 7, 6 separately. The performance of the model established on the age period from 2 hours to 7 days was shown in line 7 of Table 1 with RPD of 5.9413. The performance of the model from 7 days to 45 days was shown in line 9 of Table 1 with RPD of 6.8236. The values of RPD were increased, meanwhile, the values of RMSEP and RMSEC were decreased. It was

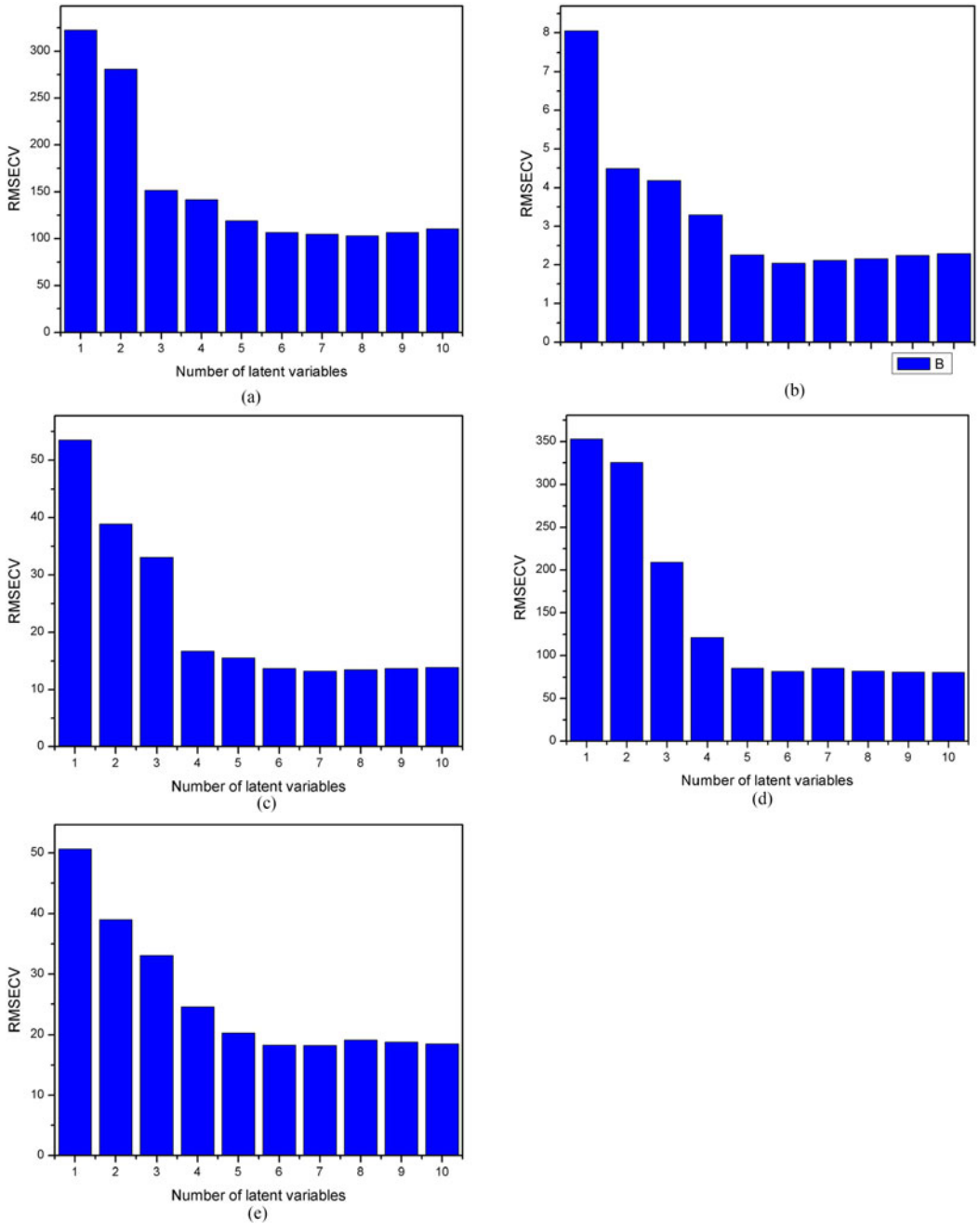


Fig. 3. RMSECV bar plot as a function of number of latent variables. PLSR model built on age periods (a) from 2 hours to 45 days, (b) from 2 hours to 24 hours, (c) from 24 hours to 7 days, (d) from 7 days to 45 days, and (e) from 2 hours to 7 days.

demonstrated that the performance of two models established on two age period were better than that of the model established on the age period from 2 hours to 45 days. Building models on two age periods could improve the predictive ability. Similarly, two models were built on the age periods from 2 hours to 24 hours and 24 hours to 7 days. The best latent variables of two models were selected as 6, 7 separately according to the lowest RMSECV shown in Fig. 3(b) and Fig. 3(c). The performance of the model built on the age period from 2 hours to 24 hours was shown in line 6

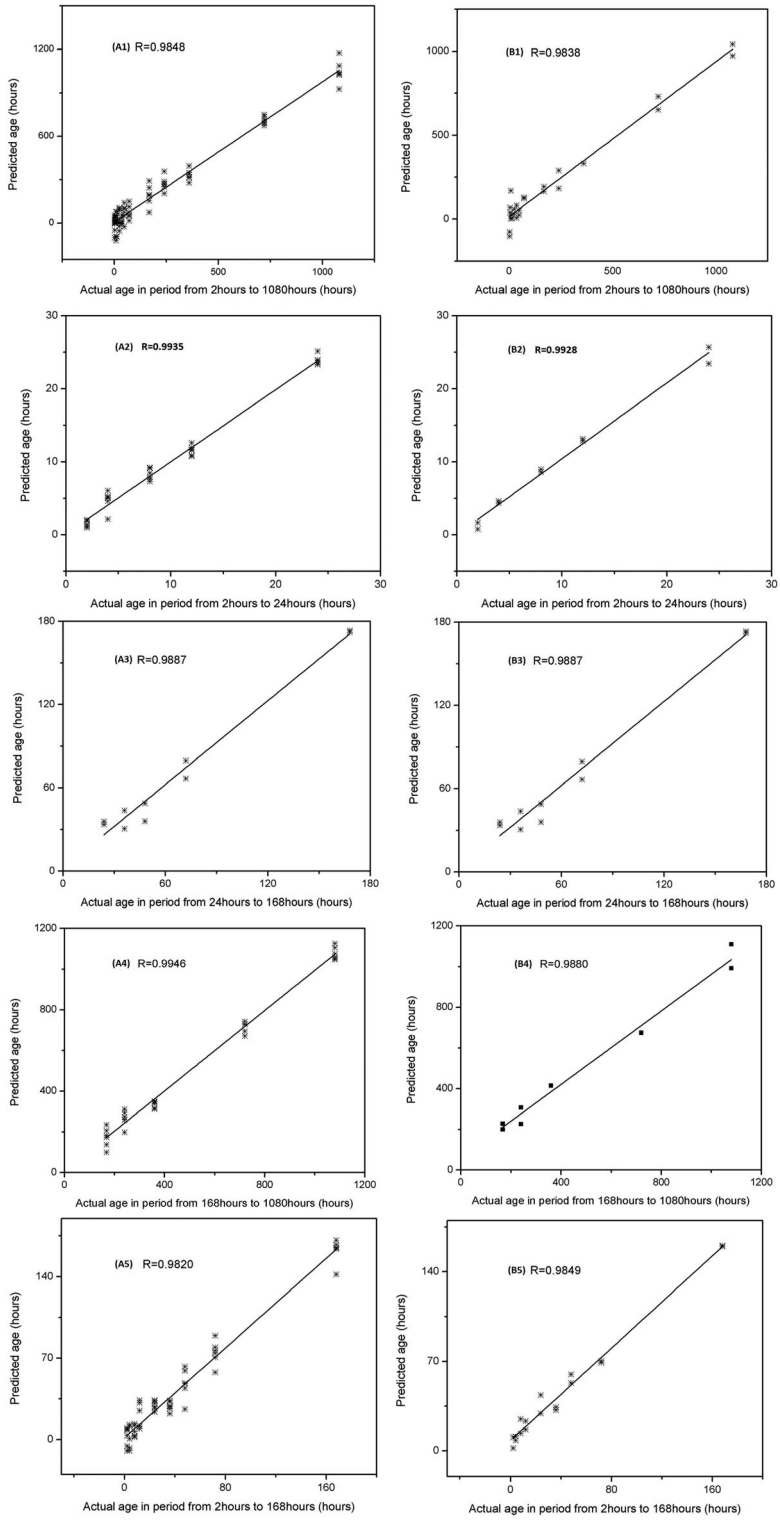


Fig. 4. Prediction results of five models built on different age periods for calibration set (A) and prediction set (B). (A1) and (B1), from 2 hours to 1080 hours; (A2) and (B2), from 2 hours to 24 hours; (A3) and (B3), from 24 hours to 168 hours; (A4) and (B4), from 168 hours to 1080 hours; (A5) and (B5), from 2 hours and 168 hours.

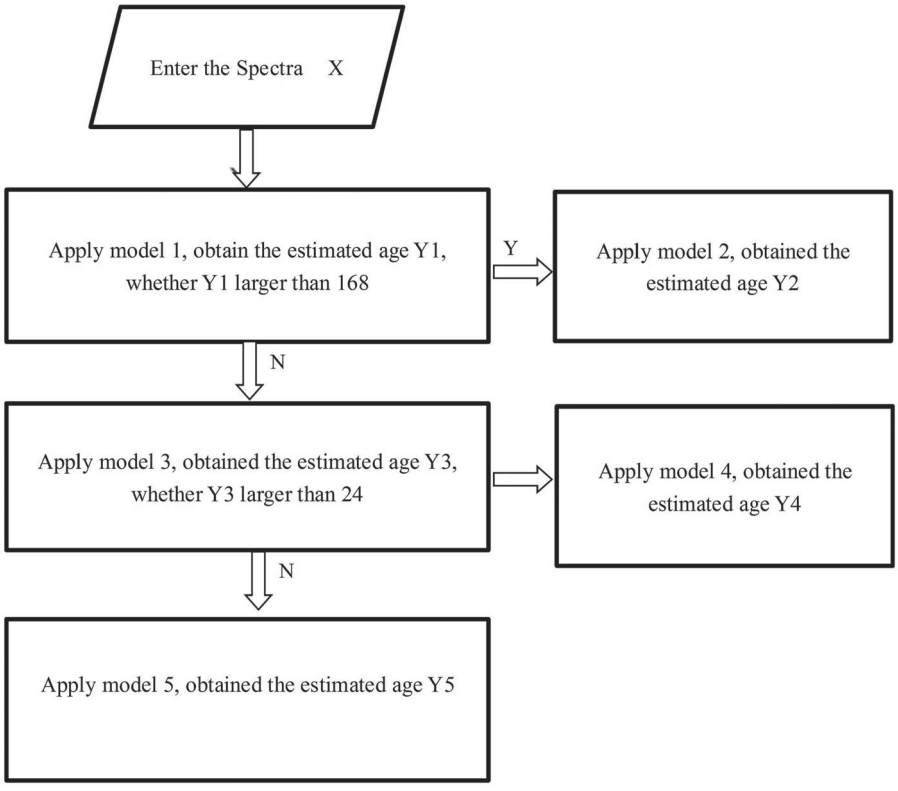


Fig. 5. Procedure used to predict the age of bloodstains.

of Table 1 with RPD = 9.7768 and RMSEP = 0.9365 hours. The performance of the model from 24 hours to 7 days can be seen in line 8 of Table 1 with RPD = 7.0384. The predictive ability of two models was largely improved. Fig. 4 shows the prediction results of five PLSR models built on five different age periods for the calibration data set (A) and prediction data set (B). Finally, the age of bloodstains can be predicted by three models built on the age periods from 2 hours to 24 hours, 24 hours to 7 days and 7 days to 45 days. The predictive ability of three PLSR models with the RPD value of 9.7768, 7.0384, and 6.8236 were satisfactory.

The procedure to predict the age of the bloodstain was shown in Fig. 5. Three spectra were used to predict the age of bloodstains following the procedure. The real ages of three spectra were 8 hours, 48 hours and 720 hours. The predicted result was demonstrated in Table 2. It can be seen that the final estimated age were 9.2271 hours, 48.2462 hours, and 672.7706 hours. The ages of three spectra, estimated by applying the model which was built on the age period from 2 hours to 45 days, were 40.1776 hours, 50.3443 hours and 651.8326 hours. It is obvious that the error between the estimated and the real age is decreased.

3.3 PCR Models

In this work, the spectral data on the edge region of 500-780 nm after pre-processing were used to establish five PCR models on five age period from 2 hours to 45 days, 2 hours to 7 days, 2 hours to 24 hours, 1 day to 7 days and 7 days to 45 days. The performance of five PCR models was presented in Table 3. The RPD values of models, built on the age periods from 2 hours to 24 hours, 1 day to 7 days and 7 days and 45 days, were 4.6217, 3.5942, 3.7757, respectively, and all values are larger than 3; thus, this model could be considered to have good prediction ability, but the RPD values of PCR models are lower than that of PLSR models, and the RMSEP values of models are larger than that of PLSR models. The prediction ability of the PLSR models was much better than

TABLE 2
Predicted Age of Three Spectrums Were Obtained Following the Procedure

	Spectrum 1	Spectrum 2	Spectrum 3
Real age	8 hours	48 hours	720 hours
Model 1	40.1776 hours	50.3443 hours	651.8326 hours
Model 2			672.7706 hours
Model 3	3.2293 hours	59.7096 hours	
Model 4		48.2462 hours	
Model 5	9.2271 hours		

Model built on age periods: Model 1, from 2 hours to 45 days; Model 2, from 7 days to 45 days; Model 3, from 2 hours to 7 days; Model 4, from 24 hours to 168 hours; Model 5, from 2 hours to 24 hours.

TABLE 3
Best Results of PCR Models

age	Spectral regions(nm) and pretreatment	Calibration		Prediction			LVs
		R_c	RMSEC(h)	R_p	RMSEP(h)	RPD	
2h-45d	500-780 + SNV	0.9206	124.3536	0.9227	122.7470	2.6453	5
2-24h	500-780 + SNV	0.9911	1.0403	0.9736	1.7784	4.6217	10
2h-7d	500-780 + SNV	0.9753	10.9968	0.8731	24.2543	2.1104	9
24h-7d	500-780 + SNV	0.9540	15.4987	0.9560	15.1618	3.5942	4
7d-45d	500-780 + SNV	0.9847	59.4070	0.9602	95.1908	3.7757	7

that of PCR models. Therefore, the article did not show the images which were about the actual age versus the estimated age.

3.4 LS-SVM Models

Five LS-SVM models with kernel function of RBF were established on five age periods. The aforementioned LVs were applied as inputs of LS-SVM models for improving the training speed and reducing the training error. The regularization parameter gamma (γ), and the kernel parameter sigma^2 (σ^2) are two very critical parameters. A two-step grid searching method with leave-one-out cross validation was used to obtain the global optimal (γ and σ^2).

The prediction results of five LS-SVM models were shown in Table 4. The actual age values and the predicted values calculated from the optimal model (calibration set and prediction set) were shown in Fig. 6. The performance of the LS-SVM model built on the age periods from 2 hours to 45 days was shown in line 1 of Table 4 with RPD $=$ 7.6709 and Rp $=$ 0.9900. The precision and robustness of this model was very well. However, it was difficult to accurately predict the age of bloodstains at the age periods from 2 hours to 168 hours. Thus, models built on different age periods were implemented to improve the predictive capability. Although the values of the RPD

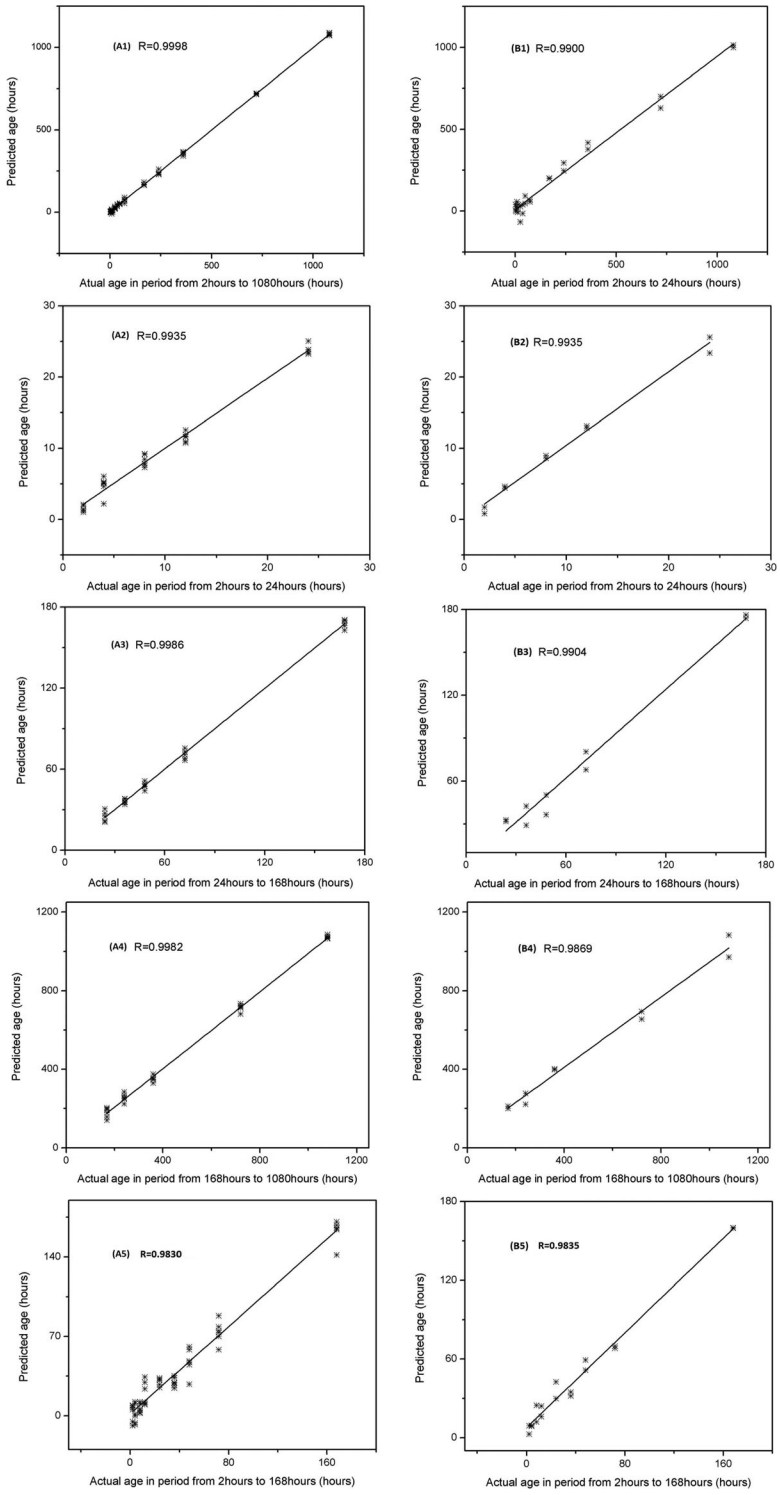


Fig. 6. Prediction results of LS-SVM models built on different age periods for calibration set (A) and prediction set (B). (A1) and (B1), from 2 hours to 1080 hours; (A2) and (B2), from 2 hours to 24 hours; (A3) and (B3), from 24 hours to 168 hours; (A4) and (B4), from 168 hours to 1080 hours; (A5) and (B5), from 2 hours and 168 hours.

TABLE 4
Best Results of LS-SVM Models

age	Input data	Parameters		Calibration		Prediction		RPDs
		γ	σ	R_c	RMSEC(h)	R_p	RMSEP(h)	
2h-45d	8LVs	434.280	0.1307	0.9977	8.2579	0.9900	42.792	7.6709
2h-24h	6LVs	90703	467	0.9935	0.9219	0.9935	0.9219	8.9153
24h-7d	7LVs	1 706 800	200	0.9986	2.7385	0.9904	7.4566	7.3082
2h-7d	7LVs	15 681	43	0.9830	8.8055	0.9835	8.3371	6.1396
7d-45d	6LVs	51.1387	0.7	0.9982	20.6531	0.9869	49.534	7.2559

calculated from the models built on the age periods from 1 days to 7 days and 7 days to 45 days were a little lower than 7.6709, the value of RMSEP were largely decreased. The models could be considered to have excellent precision and robustness. The results showed that LS-SVM models outperformed PLSR and PCR models.

3.5 Comparison of the Multivariate Data Analysis Methods

Both calibration and prediction sets were used to evaluate the efficiencies of three multivariate techniques (PCR, PLSR and LS-SVM). The methods of PLSR, PCR, and LS-SVM were compared and, with respect to prediction accuracy, they could be arranged in the following order: LS-SVM $>$ PLSR $>$ PCR. In PLSR, the data decomposition is done by using spectral information and age data, while PCR employs only spectral. It is noted that the relationship of the fractions of three hemoglobin derivatives and the age of bloodstains is non-linearity [14]; meanwhile, the performance of LS-SVM models was a little better than that of PLSR models, which indicated that LS-SVM with RBF kernel function could make good use of the linear and latent nonlinear information of spectral data and the latent nonlinear information.

4. Conclusion

Three chemometrics methods combined with visible reflectance spectroscopy were successfully applied for accurately determining the age of bloodstains. The predictive accuracy and robustness of PLS and LS-SVM models for age determination between 2 hours and 45 days old are significantly better than all previous studies in the literature. Models, established on three age periods by PLSR and LS-SVM methods, can greatly improve the predictive capability. The performance of LS-SVM models was a little bit better than that of PLSR models. The result shows that visible reflectance spectroscopy combined with LS-SVM method is a reliable tool to estimate the age of bloodstains accurately and non-destructively for forensic purposes.

References

[1] S. H. James, P. E. Kish, and T. P. Sutton, *Principles of Bloodstain Pattern Analysis: Theory and Practice*. Boca Raton, FL, USA: CRC, 2005.
[2] G. Edelman, T. G. van Leeuwen, and M. C. G. Aalders, "Hyperspectral imaging for the age estimation of blood stains at the crime scene," *Forensic Sci Int.*, vol. 223, no. 1, pp. 72–77, 2012.

[3] J. Agudelo, C. Huynh, and J. Halámek, "Forensic determination of blood sample age using a bioaffinity-based assay," *Analyst*, vol. 140, no. 5, pp. 1411–1415, 2015.
[4] T. Matsuoka, T. Taguchi, and J. Okuda, "Estimation of bloodstain age by rapid determinations of oxyhemoglobin by use of oxygen electrode and total hemoglobin," *Biol. Pharmaceutical Bull.*, vol. 18, no. 8, pp. 1031–1035, 1995.
[5] S. E. Anderson, G. R. Hobbs, and C. P. Bishop, "Multivariate analysis for estimating the age of a bloodstain," *J. Forensic Sci.*, vol. 56, no. 1, pp. 186–193, 2011.
[6] M. Bauer, S. Polzin, and D. Patzelt, "Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: A possible indicator of the age of bloodstains?" *Forensic. Sci. Int.*, vol. 138, no. 1, pp. 94–103, 2003.
[7] Y. Fujita, K. Tsuchiya, and S. Abe, "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.*, vol. 152, no. 1, pp. 39–43, 2005.
[8] H. Sakurai, K. Tsuchiya, and Y. Fujita, "Dating of human blood by electron spin resonance spectroscopy," *Naturwissenschaften*, vol. 76, no. 1, pp. 24–25, 1989.
[9] J. Andrasko, "The estimation of age of bloodstains by HPLC analysis," *J. Forensic Sci.*, vol. 42, no. 4, pp. 601–607, 1997.
[10] H. Inoue, F. Takabe, and M. Iwasa, "Identification of fetal hemoglobin and simultaneous estimation of bloodstain age by high-performance liquid chromatography," *Int. J. Legal Med.*, vol. 104, no. 3, pp. 127–131, 1991.
[11] G. Edelman, V. Manti, and S. M. van Ruth, "Identification and age estimation of blood stains on colored backgrounds by near infrared spectroscopy," *Forensic Sci. Int.*, vol. 220, no. 1, pp. 239–244, 2012.
[12] S. Strasser, A. Zink, and G. Kada, "Age determination of blood spots in forensic medicine by force spectroscopy," *Forensic Sci. Int.*, vol. 170, no. 1, pp. 8–14, 2007.
[13] B. Li, P. Beveridge, and W. T. O'Hare, "The age estimation of blood stains up to 30days old using visible wavelength hyperspectral image analysis and linear discriminant analysis," *Sci. Justice*, vol. 53, no. 3, pp. 270–277, 2013.
[14] W. R. Premasiri, J. C. Lee, and L. D. Ziegler, "Surface-enhanced Raman scattering of whole human blood, blood plasma, and red blood cells: Cellular processes and bioanalytical sensing," *J. Phys. Chem. B.*, vol. 116, no. 31, pp. 9376–9386, 2012.
[15] S. Boyd, M. F. Bertino, and S. J. Seashols, "Raman spectroscopy of blood samples for forensic applications," *Forensic Sci. Int.*, vol. 208, no. 1, pp. 124–128, 2011.
[16] R. H. Bremmer, A. Nadort, and T. G. Van Leeuwen, "Age estimation of blood stains by hemoglobin derivative determination using reflectance spectroscopy," *Forensic Sci. Int.*, vol. 206, no. 1, pp. 166–171, 2011.
[17] E. K. Hanson and J. Ballantyne, "A blue spectral shift of the hemoglobin soret band correlates with the age (time since deposition) of dried bloodstains," *PLoS One*, vol. 5, no. 9, 2010, Art. no. e12830.
[18] B. Li, P. Beveridge, and W. T. O'Hare, "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.*, vol. 212, no. 1, pp. 198–204, 2011.
[19] A. M. Mouazen, B. Kuang, and J. De Baerdemaeker, "Comparison among principal component, partial least squares and back propagation neural network analyses for accuracy of measurement of selected soil properties with visible and near infrared spectroscopy," *Geoderma*, vol. 158, no. 1, pp. 23–31, 2010.
[20] B. Özbalci, İ. H. Boyaci, and A. Topcu, "Rapid analysis of sugars in honey by processing Raman spectrum using chemometric methods and artificial neural networks," *Food Chem.*, vol. 136, no. 3-4, pp. 1444–1452, 2013.
[21] F. Liu, F. Zhang, and Z. Jin, "Determination of acetolactate synthase activity and protein content of oilseed rape (Brassica napus L.) leaves using visible/near-infrared spectroscopy," *Anal. Chim. Acta*, vol. 629, no. 1, pp. 56–65, 2008.
[22] R. H. Bremmer, K. G. de Bruin, and M. J. C. van Gemert, "Forensic quest for age determination of bloodstains," *Forensic Sci. Int.*, vol. 216, no. 1, pp. 1–11, 2012.
[23] B. Hemmateenejad, M. Akhond, and F. Samari, "A comparative study between PCR and PLS in simultaneous spectrophotometric determination of diphenylamine, aniline, and phenol: Effect of wavelength selection," *Spectrochim. Acta A*, vol. 67, no. 3, pp. 958–965, 2007.
[24] P. Druilhet and A. Mom, "Shrinkage structure in biased regression," *J. Multivariate Anal.*, vol. 99, no. 2, pp. 232–244, 2008.
[25] B. Hemmateenejad, R. Ghavami, and R. Miri, "Net analyte signal-based simultaneous determination of antazoline and naphazoline using wavelength region selection by experimental design-neural networks," *Talanta*, vol. 68, no. 4, pp. 1222–1229, 2006.
[26] M. Daszykowski, Y. Vander Heyden, and B. Walczak, "Robust partial least squares model for prediction of green tea antioxidant capacity from chromatograms," *J. Chromatogr. A*, vol. 1176, no. 1, pp. 12–18, 2007.
[27] R. Ergon, "Reduced PCR/PLSR models by subspace projections," *Chemometr. Intell. Lab. Syst.*, vol. 81, no. 1, pp. 68–73, 2006.
[28] F. Chauchard, R. Cogdill, and S. Roussel, "Application of LS-SVM to non-linear phenomena in NIR spectroscopy: Development of a robust and portable sensor for acidity prediction in grapes," *Chemometr. Intell. Lab. Syst.*, vol. 71, no. 2, pp. 141–150, 2004.
[29] M. Blanco, J. Coello, and H. Iturriaga, "NIR calibration in non-linear systems: Different PLS approaches and artificial neural networks," *Chemometr. Intell. Lab. Syst.*, vol. 50, no. 1, pp. 75–82, 2000.
[30] Y. Su, B. Xiang, and J. Xu, "Feasibility research on rapid detection of dimethoate in water by near-infrared spectroscopy," *Anal. Methods*, vol. 4, no. 6, pp. 1742–1746, 2012.
[31] R. H. Bremmer, D. M. De Bruin, and M. De Joode, "Biphasic oxidation of oxy-hemoglobin in bloodstains," *PloS One*, vol. 6, no. 7, 2011, Art. no. e21845.
[32] R. H. Bremmer, G. Edelman, and T. D. Vegter, "Remote spectroscopic identification of bloodstains," *J. Forensic Sci.*, vol. 56, no. 6, pp. 1471–1475, 2011.
[33] Å. Rinnan, F. van den Berg, and S. B. Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC-Trends Anal. Chem.*, vol. 28, no. 10, pp. 1201–1222, 2009.

[34] G. J. Edelman, E. Gaston, and T. G. Van Leeuwen, "Hyperspectral imaging for non-contact analysis of forensic traces," *Forensic Sci. Int.*, vol. 223, no. 1, pp. 28–39, 2012.
[35] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures," *Anal. Chem.*, vol. 36, no. 8, pp. 1627–1639, 1964.
[36] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, "Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra," *Appl. Spectrosc.*, vol. 43, no. 5, pp. 772–777, 1989.
[37] D. Cozzolino, W. U. Cynkar, and N. Shah, "Multivariate data analysis applied to spectroscopy: Potential application to juice and fruit quality," *Food Res. Int.*, vol. 44, no. 7, pp. 1888–1896, 2011.
[38] L. Pasti, D. Jouan-Rimbaud, and D. L. Massart, "Application of Fourier transform to multivariate calibration of near-infrared data," *Anal. Chim. Acta*, vol. 364, no. 1, pp. 253–263, 1998.
[39] A. Borin, M. F. Ferrao, and C. Mello, "Least-squares support vector machines and near infrared spectroscopy for quantification of common adulterants in powdered milk," *Anal. Chim. Acta*, vol. 579, no. 1, pp. 25–32, 2006.
[40] O. Devos, C. Ruckebusch, and A. Durand, "Support vector machines (SVM) in near infrared (NIR) spectroscopy: Focus on parameters optimization and model interpretation," *Chemometr. Intell. Lab. Syst.*, vol. 96, no. 1, pp. 27–33, 2009.
[41] D. Cozzolino, M. J. Kwiatkowski, and M. Parker, "Prediction of phenolic compounds in red wine fermentations by visible and near infrared spectroscopy," *Anal. Chim. Acta*, vol. 513, no. 1, pp. 73–80, 2004.
[42] Z. Guo, Q. Chen, and L. Chen, "Optimization of informative spectral variables for the quantification of EGCG in green tea using Fourier transform near-infrared (FT-NIR) spectroscopy and multivariate calibration," *Appl. Spectrosc.*, vol. 65, no. 9, pp. 1062–1067, 2011.
[43] X. Zou, J. Zhao, and Y. Li, "Selection of the efficient wavelength regions in FT-NIR spectroscopy for determination of SSC of 'Fuji'apple based on BiPLS and FiPLS models," *Vib. Spectrosc.*, vol. 44, no. 2, pp. 220–227, 2007.