Analytical Methods

ROYAL SOCIETY OF CHEMISTRY

PAPER

View Article Online
View Journal | View Issue

Cite this: *Anal. Methods*, 2025, **17**, 6179

Comparative analysis of hyperspectral and near-infrared spectroscopy for bloodstain deposition time estimation†

Ying Liang,[a] Qifu Yang,[a] Jiaquan Wu,*[a] Kun Ma,[a] Xinyu Zhang,[a] Huihui Ren,[a] Hanyu Zhu,[a] Xingshuai Peng,[a] Jiateng Wang[a] and Jianqiang Zhang*[b]

Background: Bloodstains are a prevalent and critical type of forensic evidence at crime scenes. Accurate determination of bloodstain age is essential for crime resolution, and non-destructive spectral methods are instrumental in this process. While extensive research has established the practicality of hyperspectral imaging (HSI) in specific forensic contexts, limited studies have explored near-infrared (NIR) spectroscopy. Owing to its superior penetration capabilities and high sensitivity, NIR holds promise in addressing certain limitations of HSI. This study aims to assess the applicability of NIR spectroscopy for bloodstain age estimation in forensic contexts and to compare its efficacy with HSI. Results: Bloodstains were aged on various substrates over a 60 day period, with periodic analyses conducted using both spectral methods. Chemometric analysis of the spectral data was performed following SNV preprocessing and application of different regression algorithms. First, linear regression analysis was utilized to determine the effect of material on bloodstain deposition. Under the premise of distinguishing materials, partial least squares (PLS) regression was employed to extract eight latent variables from HSI and NIR spectral data for regression prediction. However, the prediction performance was suboptimal. To address this, polynomial features were introduced into the PLS regression algorithm to capture the nonlinear relationships in the spectral data, and the improved model significantly enhanced the prediction performance. Furthermore, PLS polynomial regression was applied to predict homologous data, and the results also demonstrated favorable performance. Finally, to optimize the prediction accuracy of multimodal data, a multilayer perceptron (MLP) was introduced for regression prediction through multimodal data fusion, further improving the overall performance of the model. Finally, predictive performance was evaluated across models, emphasizing their specific strengths. For homologous data fusion, comparable root mean square errors of prediction (RMSEP) were achieved for HSI and NIR spectra, at 8.35 and 8.15 days, respectively. Similar RMSEP values were observed in multimodal data fusion, and the accuracy of both low-level and intermediate-level fusion methods was evaluated. Significance: HSI and NIR spectroscopy each provide unique advantages in bloodstain detection. Data fusion of these methods helps mitigate external influences, enhancing the approach's general applicability. This integrated method facilitates rapid estimation of bloodstain age at crime scenes, aiding in crime timeline determination and presenting valuable potential for forensic applications.

Received 28th November 2024
Accepted 19th February 2025

DOI: 10.1039/d4ay02167c

rsc.li/methods

1 Introduction

The rapid identification of human bloodstains is critically important in criminal investigations, as bloodstains represent one of the most objective and stable types of evidence available at a crime scene.[1] Blood evidence offers crucial forensic information and is considered indispensable in forensic identification.[2] In particular, determining the bloodstain deposition time is valuable in establishing the timing of a crime and assessing whether a bloodstain may be linked to individuals involved in the incident. However, bloodstains are highly susceptible to environmental influences, complicating their analysis. Hemoglobin (HbO_2) in blood quickly reacts with atmospheric oxygen and undergoes auto-oxidation to methemoglobin (metHb) within 48 h. Once formed, metHb progressively denatures into hemichrome (HC) over hours or days, depending on humidity and temperature. This complex transformation makes it challenging for forensic analysts to precisely determine the time of an incident, underscoring the need for

[a]College of Science, Kunming University of Science and Technology, Kunming 650500, China

[b]Yunnan Police College, 650000, Kunming, China. E-mail: 1765301431@qq.com

† Electronic supplementary information (ESI) available. See DOI: https://doi.org/10.1039/d4ay02167c

rapid, accurate, and non-destructive methods to estimate bloodstain age.

In forensic practice, conventional blood detection methods include direct observation, reagent-based approaches, DNA analysis,[3] the tetramethylbenzidine (TMB) test,[4] and the luminol test (LT).[5] Direct observation relies heavily on subjective judgment and experience, while reagent-based techniques often damage blood components, thereby compromising subsequent DNA profiling and reducing the available DNA quantity for analysis.[6–9] Many chemical reagents used in these tests produce visible color changes,[10] fluorescence,[11] or luminescence[5] when in contact with blood. However, DNA analysis methods are time-intensive and costly, and none of these approaches provide precise timing based solely on the bloodstain itself.[12] To overcome these limitations, researchers have explored non-destructive techniques for identifying and analyzing bloodstains, including ultraviolet-visible (UV-vis) spectroscopy,[13] Raman spectroscopy,[14] Fourier-transform infrared (FTIR) spectroscopy,[15] hyperspectral imaging (HSI),[16] and near-infrared (NIR) spectroscopy.[17] These methods offer several advantages, such as being non-destructive, rapid, convenient, and contamination-free, while also demonstrating high sensitivity and adaptability to various environmental conditions. Following plasma water evaporation, blood degradation predominantly involves the breakdown of iron-containing proteins and hemoglobin in red blood cells, which gradually convert into methemoglobin, heme, and hemochromes.[18,19] This transformation is distinctly captured in HSI spectra. NIR spectroscopy, on the other hand, is effective for observing bloodstain dehydration and methemoglobin formation during later stages of aging.[20] Given the complementary information derived from these different spectroscopic methods, combining HSI and NIR spectroscopy may yield a more accurate estimation of blood stain deposition time (TSD) by focusing on distinct aspects of blood degradation. However, no studies to date have attempted to integrate these two approaches.

In recent years, single spectroscopic techniques have been applied to predict blood TSD. For example, in 2012, Gerda Edelman *et al.*[21] evaluated the utility of NIR technology for bloodstain identification and age estimation on dark backgrounds. Using partial least squares (PLS) regression, they successfully estimated the age of month-old bloodstains, achieving a root mean square error of 8.9%. In a subsequent study, E. Gerda *et al.*[22] combined HSI with cluster analysis to determine the absolute age of bloodstains. In 2013, L. Bo *et al.*[23] utilized visible HSI technology along with Linear Discriminant Analysis (LDA) to age bloodstains up to 30 days. However, many current approaches adopt a strictly analytical perspective rather than addressing forensic applications. Models for bloodstain deposition time estimation aim to provide a general chronological framework that also considers the specific environmental conditions of the crime scene. Such conditions impart distinct characteristics to each case, substantially influencing the rate and nature of blood degradation and, consequently, the timing conclusions derived from these models.

Crime scenes often present complex and varied environments, where bloodstain samples may originate from individuals of differing biological profiles and deposit on a range of substrates. Moreover, environmental contaminants frequently affect bloodstains, complicating both their extraction and analysis. To address these challenges, this study integrates HSI and NIR spectroscopic techniques to deepen the understanding of the transformation processes in bloodstains as they age. This study assessed both the individual capabilities of HSI and NIR methods and their combined potential to monitor blood degradation and estimate bloodstain age. For model development, optimal prediction models were incrementally refined under various conditions. Initially, a basic PLS model was developed for each substrate to create spectral prediction frameworks. Polynomial extensions were incorporated to capture more complex nonlinear relationships to enhance prediction accuracy.

2 Materials and methods

2.1 Sample

All experiments were conducted in accordance with relevant Chinese laws and the guidelines of the Chinese Forensic Medicine Association. Ethical approval was obtained from the Medical Ethics Committee of Kunming University of Science and Technology, and informed consent was acquired from all volunteers. In the experiment, blood samples were prepared on three different substrates: cement board, gypsum board, and iron board. Further details are provided in ESI Fig. S1.†

2.2 Spectral acquisition equipment

An HSI system was established in the laboratory to capture bloodstain images, The HSI system is provided by Shenzhen Zhongda Ruihe Technology Co. This setup consisted of a sample stage, a hyperspectral camera, and four halogen-based broadband white light sources. The system used a liquid crystal tunable filter (LCTF) connected to a complementary metal-oxide-semiconductor (CMOS) camera with a resolution of 2048 × 2046 pixels. The LCTF featured a 35 nm aperture. To eliminate external light interference during data acquisition, the entire system was enclosed. Before acquisition, the light source was calibrated using a light intensity calibration function. The HSI system covered a spectral range of 400–1000 nm with a spectral sampling interval of 5 nm and an exposure time of 40 ms.

The MicroNIR handheld NIR spectrometer is produced by VIAVI Solutions in Beijing, China, the NIR system employed a 128-pixel detector array to capture spectral data. It was equipped with two tungsten filament lamps as radiation sources and a linear variable filter (LVF) connected directly to an Indium Gallium Arsenide (InGaAs) linear array detector. This system operated within a wavelength range of 908–1676 nm, offering a spectral resolution of 4 nm with a data sampling interval of 6 nm. The integration time was fixed at 9.6 μs. To enhance measurement accuracy, a 99% diffuse reflectance whiteboard was used as a reference, and the entire acquisition was conducted under dark conditions to minimize light interference.

2.3 Spectral data acquisition

The steps from sample preparation to spectral data acquisition are illustrated in ESI Fig. S2.† Data were collected across 102 time intervals, with both HSI and NIR spectra recorded at each interval. For each collection, three spectra were captured at each sample point, and the mean of the three measurements was used as the final spectral data for that time point. NIR spectral data were exported using MicroNIR 1.5.7 software, while HSI data were processed and exported using ENVI software. HSI spectral data acquisition involved selecting a region of interest (ROI) in each sample, with details on ROI selection provided in the ESI (Fig S3†).

As summarized in Table 1, spectral data were collected from five sample points, where sample points 1, 2, and 3 corresponded to male blood samples, and points 4 and 5 to female blood samples. Data were acquired over 102 intervals spanning a 60 day period, yielding a total of 510 HSI and NIR spectra for each substrate, or 1530 HSI and 1530 NIR spectra in total. Subsequent preprocessing and analysis of spectral data were conducted using MATLAB® 2022a software.

3 Data processing methods

Fig. 1 illustrates the process of spectral data processing in this study, which is primarily divided into three parts. First, the raw data undergoes preprocessing. Then, homologous data from HIS and NIR are fused separately. Finally, multimodal data fusion is performed.

3.1 Data structure

The collected HSI and NIR data were processed using MATLAB® 2022a. The HSI and NIR datasets were organized as independent matrices, formatted as row vectors, where each row represented a sample spectrum, and variables were stored as column vectors. The HSI dataset comprised 121 wavelengths with 4590 sample spectra, while the NIR dataset included 125 wavelengths with 4590 sample spectra. After averaging the spectral data across three measurements, the HSI matrix dimensions were 1530 × 121, and the NIR matrix dimensions were 1530 × 125. MATLAB® 2022a was employed for data preprocessing, dimensionality reduction, and multivariate regression analysis.[24] For regression analysis, the data were partitioned into training and test sets in a 7 : 3 ratio, and *K*-fold cross-validation was applied to the training set for validation.

3.2 Preprocessing methods

Three substrates—cement board, gypsum board (permeable materials), and iron plate (non-permeable material)—were used in the experiment. All samples were collected using HSI and NIR spectroscopy in reflectance mode to assess changes in bloodstain deposition on different surfaces. Principal Component Analysis (PCA)[25] was applied to the pre-processed data to evaluate the dependence of sample distribution on variables such as gender, substrate material, and time.

To eliminate variables unrelated to bloodstain deposition time (*e.g.*, gender, substrate material), raw HSI and NIR data were preprocessed using methods such as Multiplicative Scatter Correction (MSC),[26] Savitzky–Golay (SG) smoothing, First Derivative (1D), Standard Normal Variate (SNV),[27] and a combined First Derivative and Savitzky–Golay (1DSG) approach. These preprocessing methods were used to focus the analysis on the actual chemical changes within the bloodstains, thereby enhancing the accuracy of deposition time predictions. Additionally, these techniques reduced variations due to differences in physical properties among samples, improved spectral comparability across different samples, and better highlighted the chemical characteristics of blood.

3.3 Linear regression analysis

Linear regression analysis was performed on the HSI and NIR spectral data, incorporating time variables, material feature matrices, and material types. The time variable represented different intervals following bloodstain deposition, while the material feature matrix contained spectral feature values for each substrate, categorized by permeable and non-permeable materials. To simplify material feature dimensions, PCA was first applied to the material feature matrix, with the first principal component (PC1) selected as the representative material

Table 1 Spectral data sampling intervals and number of spectra for bloodstain samples on different substrates

Substrate	Time interval/days	Samples number
Cement board Gypsum board Iron board	0.02, 0.04, 0.08, 0.13, 0.17, 0.21, 0.25, 0.33, 0.42, 0.50, 0.88, 0.92, 1.00, 1.08, 1.17, 1.25, 1.33, 1.96, 2.00, 2.08, 2.17, 2.25, 2.33, 3.00, 3.17, 4.00, 4.17, 5.00, 5.17, 6.00, 6.17, 8.00, 8.17, 9.00, 9.17, 9.96, 10.13, 11.00, 11.17, 12.00, 12.17, 13.00, 13.17, 14.00, 14.17, 15.00, 15.17, 16.00, 16.17, 17.00, 17.17, 18.00, 18.17, 19.00, 19.17, 20.00, 20.17, 21.00, 21.17, 22.00, 22.17, 23.00, 23.17, 24.00, 24.17, 25.00, 25.17, 26.00, 26.13, 27.00, 27.17, 28.00, 28.17, 29.17, 30.00, 30.17, 31.00, 31.17, 32.00, 33.00, 34.00, 35.00, 36.00, 37.00, 38.00, 39.00, 40.00, 41.00, 42.00, 43.00, 44.00, 45.00, 48.00, 49.00, 50.00, 51.00, 54.00, 55.00, 56.00, 57.00, 58.00, 60.00	1530

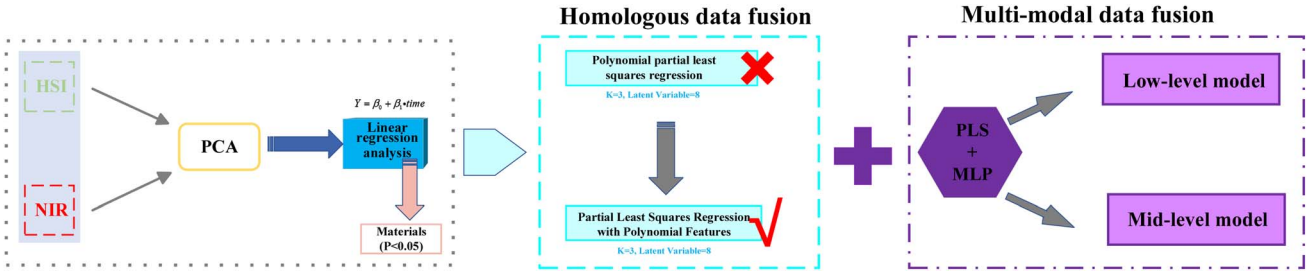

Fig. 1 Data processing flowchart.

feature. To investigate the relationship between time and material features, a simple linear regression model was established as follows:

$$Y = \beta_0 + \beta_1 \cdot \text{time}$$

where Y represents the material feature (PC1) extracted through PCA, β_0 is the intercept, and β_1 is the regression coefficient reflecting the relationship between time and the material feature. Regression coefficients were estimated using the least squares method, and predicted values and residuals were subsequently calculated.

Residuals from the regression model were grouped by material type, with permeable materials coded as 1 and impermeable materials coded as 0. An independent samples t-test was performed to compare the residual distributions between permeable and impermeable materials using P-values. $P < 0.05$ indicated a significant effect of material type on bloodstain deposition time, whereas $P > 0.05$ indicated a negligible effect.

3.4 Regression analysis and homologous data fusion

3.4.1 Regression analysis by material type. Following the approach in Section 3.3, analysis of the HSI and NIR spectral data revealed that material type influenced bloodstain deposition time. Based on this finding, polynomial partial least squares (PLS) regression was applied to the SNV-preprocessed HSI and NIR spectral data, aiming to extract key features from the high-dimensional spectral data and capture nonlinear

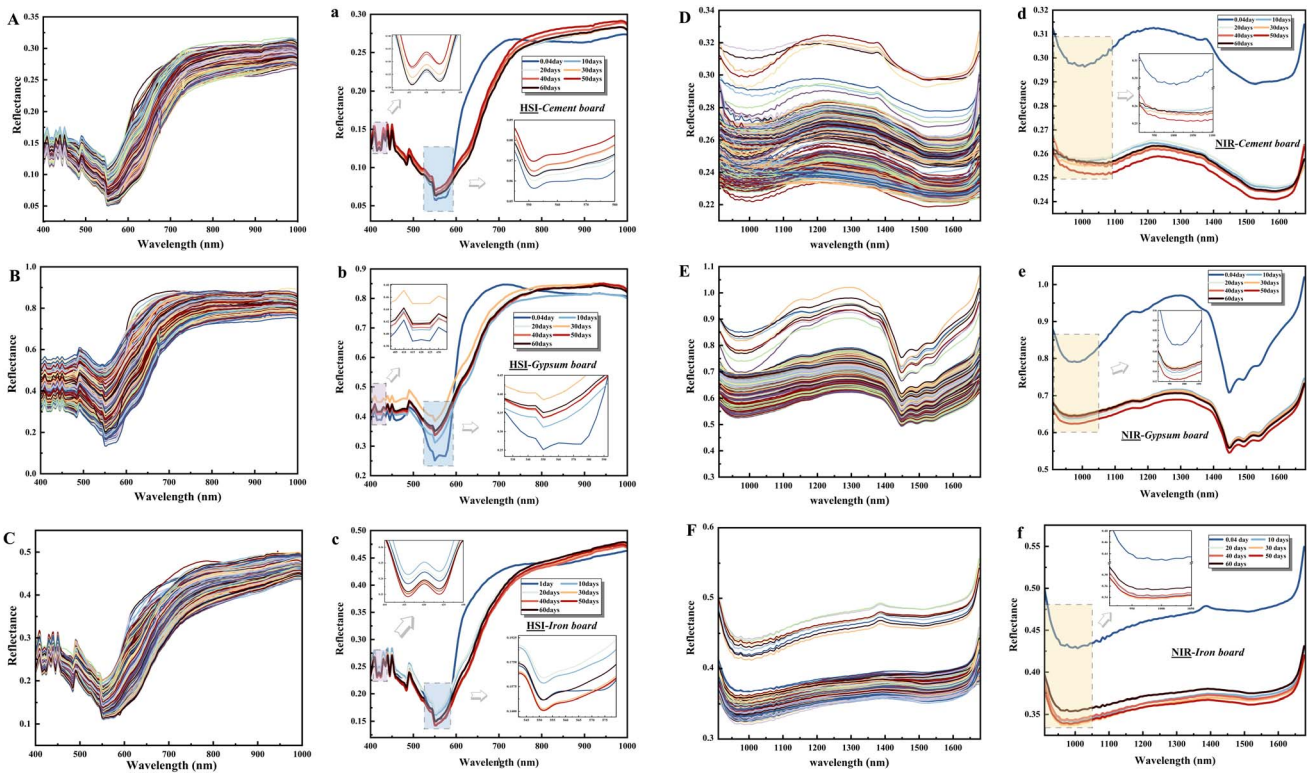

Fig. 2 Raw spectra and time-interval spectra: (A)–(C) represent the raw spectral data of HSI obtained from bloodstains on cement board, gypsum board and iron board, respectively, while (a)–(c) are the average spectra calculated for every 10 day interval (D)–(F) show the raw spectra data of NIR obtained from bloodstains on cement board, gypsum board and iron board, respectively, with (d)–(f) representing the average spectra calculated for every 10 day interval.

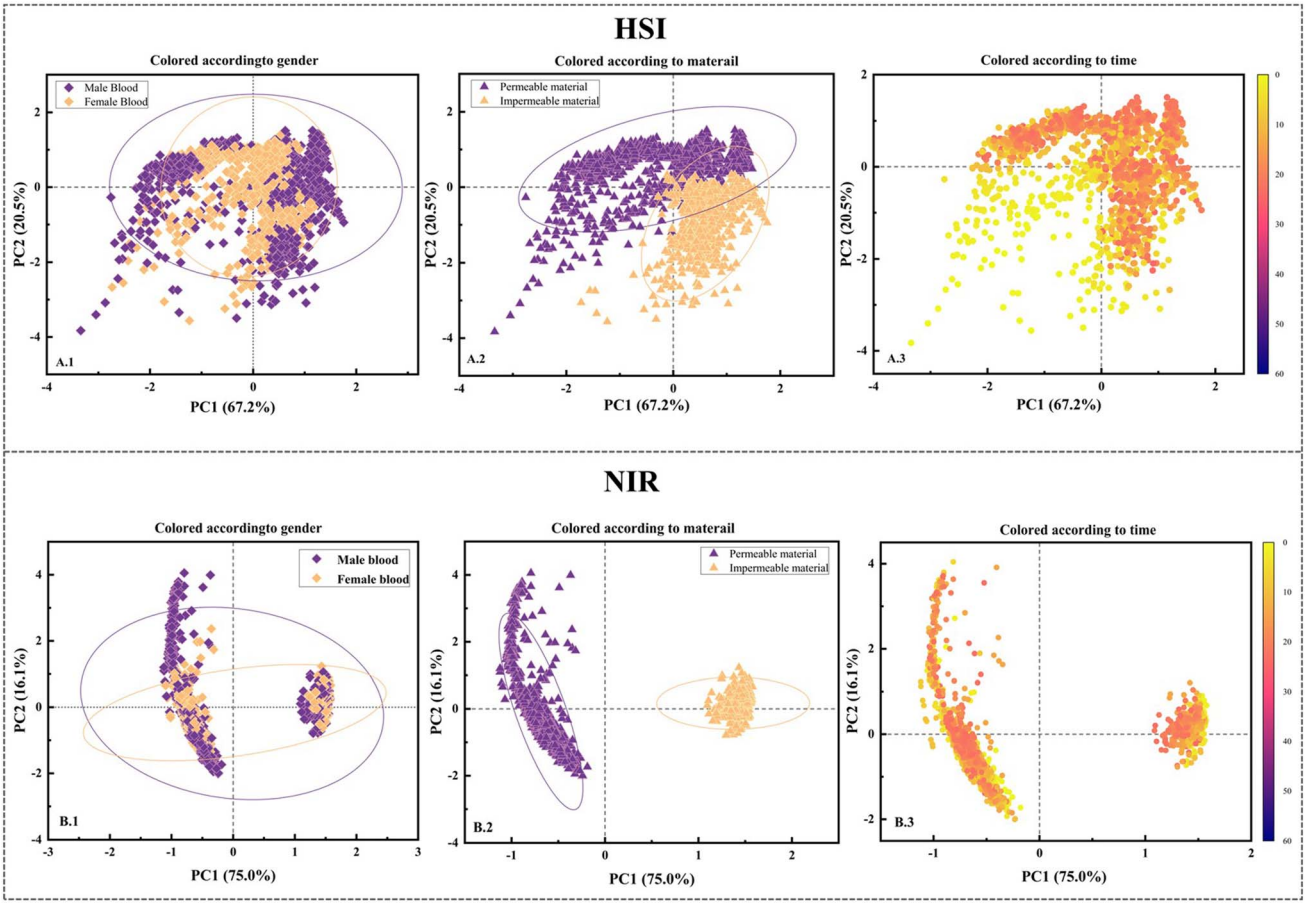

Fig. 3 PCA score plots after preprocessing. (A.1–A.3) show the score plots for HSI: (A.1) is the score plot colored by gender, (A.2) is the score plot colored by material, and (A.3) is the score plot colored by time. (B.1–B.3) shows the score plots for NIR: (B.1) is the score plot colored by gender, (B.2) is the score plot colored by material, and (B.3) is the score plot colored by time.

relationships between independent and dependent variables. Polynomial feature expansion was first applied to the original spectral matrix $\boldsymbol{X}$ for each material, generating a polynomial expansion matrix $\boldsymbol{X}'$, *i.e.*, $\boldsymbol{X}' = [\boldsymbol{X}, \boldsymbol{X}^2]$, where $\boldsymbol{X}^2$ represents the squared term for each spectral feature. This feature expansion was designed to capture potential quadratic nonlinear patterns in the independent variables, thereby improving the model's capability to represent complex spectral changes. To ensure data stability and enhance model robustness, the expanded feature matrix was standardized before applying PLS regression to model the data for each material type.[28,29] By projecting both the independent variables $\boldsymbol{X}'$ and the dependent variable $\boldsymbol{Y}$ (*i.e.*, bloodstain deposition time), the analysis extracted the most relevant latent variables. *K*-Fold cross-validation was used to select the optimal number of latent variables, and the root mean square error of cross-validation (RMSECV) was calculated

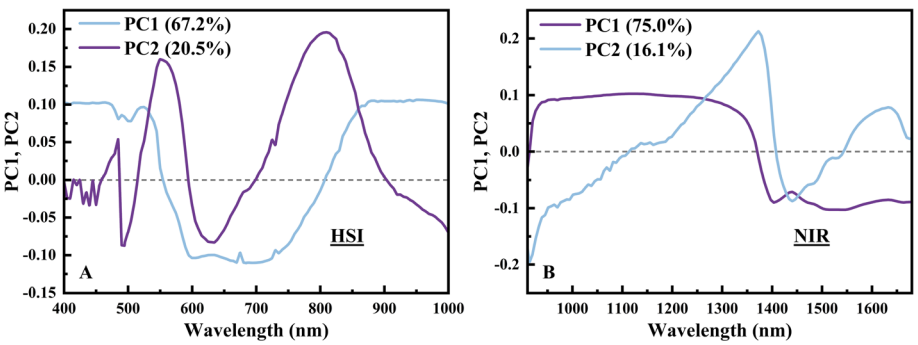

Fig. 4 PCA loading plots after SNV preprocessing. (A) shows the loadings of PC1 and PC2 for the HSI spectrum PCA, and (B) shows the loadings of PC1 and PC2 for the NIR spectrum PCA.

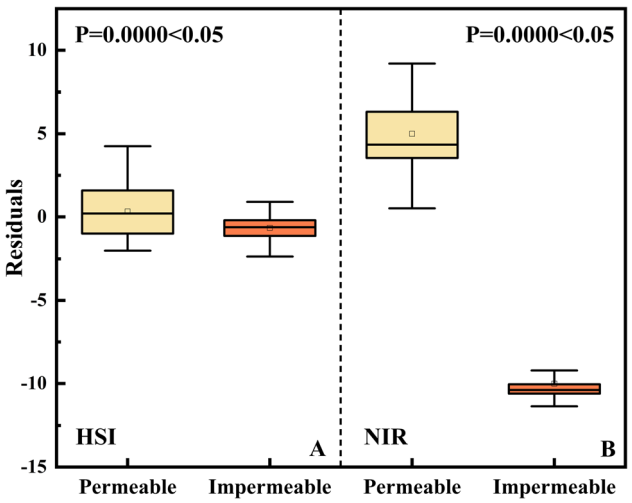

Fig. 5 Box plot of residuals divided by material type, where (A) is HSI spectral data and (B) is NIR spectral data.

to assess model accuracy. This process was repeated across all bloodstain samples, and the average RMSECV was calculated to obtain the final RMSECV value for the model.

According to the method described in Section 3.3, analysis of the HSI and NIR spectral data indicated that material type influences bloodstain deposition time. To address this, PLS regression was applied to the SNV-preprocessed HSI and NIR spectral data, aiming to extract essential features from high-dimensional spectral data while capturing the nonlinear relationships between independent and dependent variables. First, polynomial feature expansion was applied to the original spectral matrix $\boldsymbol{X}$ for each material, generating a polynomial expansion matrix where $\boldsymbol{X}^2$ represents the squared term for each spectral feature. This expansion captures potential quadratic nonlinear patterns in the independent variables, enhancing the model's capacity to represent complex spectral changes. The expanded feature matrix was standardized to ensure data stability and model robustness, followed by PLS regression modeling for each material. By projecting both independent variables and the dependent variable $\boldsymbol{Y}$ (*i.e.*, bloodstain deposition time), the analysis extracted the most relevant latent variables. *K*-Fold cross-validation was employed to determine the optimal number of latent variables, with root mean square error of cross-validation (RMSECV) calculated as the model evaluation metric. This process was repeated for all bloodstain samples, and the average RMSECV served as the final model evaluation criterion.

3.4.2 Homologous data fusion. The HSI and NIR spectral data, preprocessed by SNV, were integrated into a single feature matrix to minimize the material's effect on bloodstain deposition time prediction. First, polynomial feature expansion was applied to the preprocessed HSI and NIR spectral data to capture quadratic nonlinear features. The expanded feature matrix was then standardized, and PLS regression was used to construct the predictive model, achieving dimensionality reduction by extracting latent variables strongly correlated with the target variable. *K*-Fold cross-validation was conducted to select the optimal number of latent variables, with RMSECV calculated at each fold. The average RMSECV across all samples served as the model's performance indicator. This homologous data fusion approach effectively extracted key spectral features related to deposition time while reducing material-induced variability, thereby enhancing the model's generalization ability across different material substrates.

3.5 Multi-modal data fusion

Data fusion integrates variables by matrix juxtaposition,[30] enabling the integration of HSI and NIR spectral data derived from different materials and equipment. This process aims to mitigate material-induced effects on bloodstain deposition time, enhance the model's generalizability, and reduce material variability by combining both spectral datasets within a unified sample set.

For spectral measurements, each time point corresponds to six samples, with three spectral measurements taken per sample. The average of these measurements is then calculated to yield a representative spectrum for each material at each time point. This approach ensures that each material is associated with a unique spectrum per time point, consolidating multiple measurements into a single result. In this study, the spectral data for the first five samples were selected, and the averaged spectral data from HSI and NIR were integrated for each material. During the fusion process, the HSI and NIR spectral data were concatenated by columns, forming a fusion matrix that contains spectral information from both techniques. The rows of the fusion matrix correspond to the material samples, while the columns represent the combined HSI and NIR wavelength variables, yielding a total of $n + m = 246$. This data fusion method enables comprehensive integration of spectral information for each material across time points, providing a robust data foundation for further analysis and modeling.

Table 2 Summary of figs of merit for the polynomial PLS regression models

Data	Material	Preprocessing	Latent variable	RMSEC	RMSEP	R^2	RMSECV	Bias
HSI	Permeation	SNV	8	5.95	6.56	0.80	5.08	0.92
	Impermeable		8	5.97	6.24	0.85	5.23	1.35
NIR	Permeation		8	5.00	6.25	0.84	5.45	1.07
	Impermeable		8	5.98	6.11	0.88	5.21	0.96
HSI	No distinction		8	7.93	8.35	0.76	7.17	2.98
NIR	No distinction		8	7.60	8.15	0.79	7.98	2.23

Table 3 Prediction results of low-level and intermediate-level data fusion models

Data fusion	Data	Preprocessing	Latent variable	RMSEC	RMSEP	R^2	RMSECV	Bias
Low-level	HSI + NIR	SNV	8	2.97	3.12	0.87	3.24	0.24
Mid-level	HSI + NIR		8	2.21	2.59	0.95	2.61	0.20

Before data fusion, SNV preprocessing was performed separately on HSI and NIR data. During modeling, data standardization was applied to eliminate dimensional differences across wavelengths. PLS regression was then employed to reduce the dimensionality of the fused features and extract latent variables. Through PLS, both the fused spectral features (independent variables) and time labels (dependent variables) were dimensionally reduced, extracting the latent variables most relevant to regression predictions. Subsequently, a multilayer perceptron (MLP) neural network model was constructed with a two-hidden-layer structure, comprising 10 and 5 hidden units in the respective layers. This model, optimized for capturing nonlinear data relationships, employed the Levenberg–Marquardt (trainlm) algorithm for parameter optimization, set to a maximum of 1000 iterations, with a regularization parameter of 0.1 and a learning rate of 0.1 to ensure thorough training. Five-fold cross-validation was used to validate the model, calculating error metrics such as root mean square error (RMSE) and prediction bias (Bias) at each fold. The average values of these error metrics were then used to assess the model's overall performance.

In the final stage, intermediate data fusion was applied by combining significant latent variables from independent regression models built with each spectral technique on the same samples.[31] To ensure sample comparability, both spectral datasets used identical samples, adhering to the same strategy as in low-level data fusion. Initially, separate PLS regression models were developed for each spectral dataset, extracting the respective significant latent variables. These significant latent variables were then combined, and a PLS model was applied to the resulting latent variable matrix for further regression analysis. The remaining modeling steps and algorithm parameters remained consistent with those applied in low-level data fusion.

4 Results and discussion

4.1 Comparative analysis of HSI and NIR spectroscopy

Exposure of human blood to air initiates a series of chemical changes. The original and averaged HSI and NIR spectra obtained from bloodstains on various surfaces are illustrated in Fig. 2. The original HSI spectra (Fig. 2A–C) include individual spectra of bloodstain samples across all analyzed objects, while the corresponding averaged spectra (Fig. 2a–c) display signals recorded at specific time points, with colors indicating different times. In the HSI spectra, time-dependent changes were particularly prominent in the Soret band around 415 nm, visible in the purple region. As samples aged, a blue shift occurred, with wavelengths moving from 415 nm to shorter values, primarily due to changes in the spin state of iron ions within hemoglobin.[32] In the blue region, spanning 500–650 nm, spectral distortion was evident. Fresh bloodstains, characterized by high oxyhemoglobin content, exhibited distinct Q bands[33]—specifically, the α and β bands—peaking at 542 nm and 576 nm, respectively. Over time, these peaks progressively weakened, broadened, and eventually merged into a single band, signifying the oxidation of hemoglobin from its oxygenated form to methemoglobin. This spectral transformation reflects the blood's chemical evolution as deposition time progresses. The presence of the Q and Soret bands indicates $\pi \rightarrow \pi^*$ electronic transitions originating from the delocalized electron configuration on the porphyrin tetrapyrrole ring.[34] These bands exhibit high sensitivity to structural changes, especially during ligand exchange processes involving nitrogen atoms in the pyrrole ring, leading to band distortions. Additionally, the transition from oxyhemoglobin to methemoglobin was observed around 630 nm.[35] The averaged spectra indicated an initial increase in bloodstain reflectance, followed by a decrease and eventual

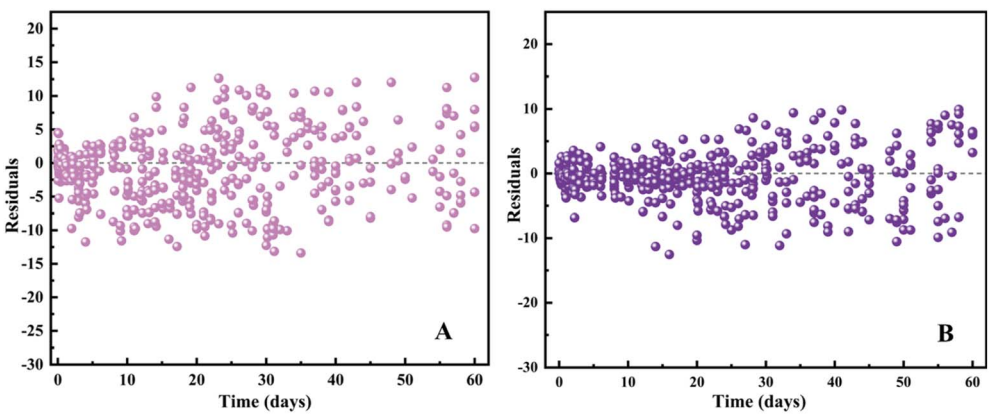

Fig. 6 Residual distribution plot of the test set for the low-level model and the mid-level model. Where, (A) shows the residual distribution of the test set for the low-level model, and (B) shows the residual distribution of the test set for the mid-level model.

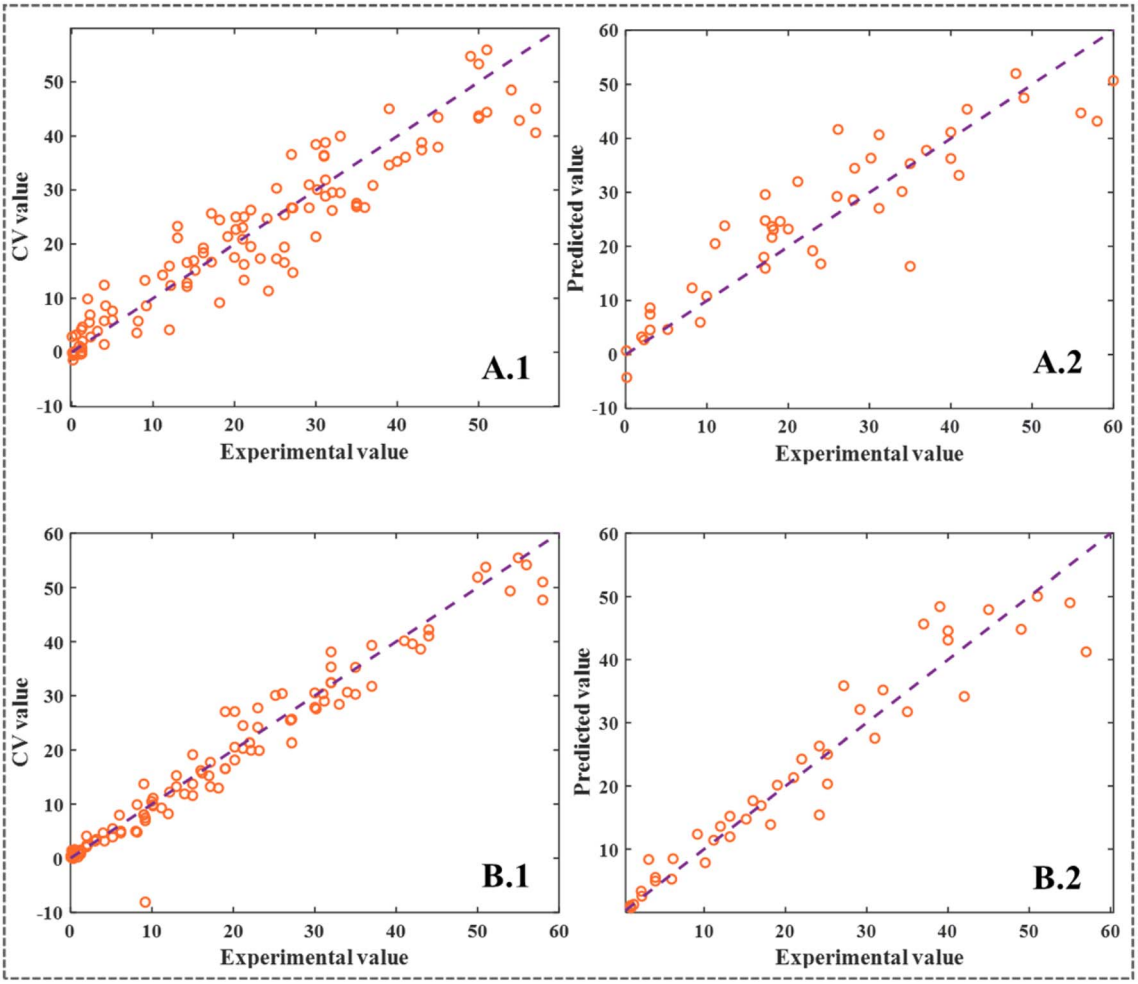

Fig. 7 Comparison of experimental values and predicted values under different models. Where, the (A.1) and (A.2) shows the experimental *versus* predicted values for the training and testing sets of the low-level fusion model, respectively, while (B.1) and (B.2) shows the experimental *versus* predicted values for the training and testing sets of the mid-level fusion model.

stabilization over time. This trend provides further insight into the degradation and oxidation processes impacting bloodstain spectral characteristics over various deposition times.

Fig. 2D–F depict the original NIR spectra of bloodstain samples on cement boards, gypsum boards, and iron plates, respectively, while Fig. 2a–c present the color-coded average signals by deposition time. In the 900–1680 nm range, reflectance intensity generally declined with increasing time since deposition (TSD), primarily due to hemoglobin degradation and the drying process. Early-stage spectra showed a significant impact from water, with rapid evaporation leading to notable decreases in reflectance, particularly within bands associated with the overtone stretching vibrations of the O–H bond in water (1000–1400 nm). This absorption was especially pronounced, as illustrated in Fig. 2d–f. The light orange bands in these figs highlight the 900–1100 nm region, where absorption is linked to the N–H bond vibrations in hemoglobin, marked by an isosbestic point near 1050 nm. Notable degradation was also observed in the early stages within the 1250–1540 nm range, which corresponds to the first overtone stretching vibration of O–H; this degradation diminished over time. Reduced spectral variations in this region, as shown in Fig. 2d–f, partly corresponded to distortions arising from scattering effects and variations in electromagnetic wave path lengths. These changes are closely related to shifts in sample geometry, thickness, and physical properties during measurement.[36] Spectral alterations are also associated with proteins, particularly C–H bond stretching vibrations in protein structures.[20]

In summary, HSI was limited by its weak penetration capacity within this wavelength range, rendering it less effective for detecting changes in water content and degradation products over time, thus confining its utility to early-stage bloodstain detection. Conversely, while hemoglobin absorption peaks were less pronounced within the 900–1700 nm range, this wavelength region captured specific absorption characteristics for water and hemoglobin degradation products,[37] facilitating analyses of water content and the degradation process in bloodstains. Additionally, the enhanced penetration and anti-interference properties of NIR light provided more

comprehensive physical and chemical information, making it highly suitable for monitoring bloodstain changes over extended periods.

4.2 Pattern recognition and PCA dimensionality reduction

Fig. 3 presents the PCA score plots for the preprocessed HSI and NIR spectral data, shown in panels A and B, respectively, with classifications color-coded by gender (A.1/B.1), material (A.2/B.2), and time (A.3/B.3). Analysis of the HSI score plot revealed no discernible gender-based grouping, indicating that gender does not influence bloodstain deposition time. However, distinct groupings by material type (permeable and impermeable) and time were observed. Similarly, the NIR score plot (B1–B3) showed clear groupings by material and a pronounced gradient over time, with no significant separation by gender. While this study primarily focused on the relationship between bloodstain deposition and time, material influence was also evident: permeable materials, due to higher absorbency, accelerated blood drying and amplified spectral changes. In contrast, blood on impermeable surfaces remained in a liquid state for longer, exerting less effect on spectral characteristics. To minimize interference from material effects on deposition time predictions, subsequent analyses implemented separate modeling and data fusion techniques to reduce potential material-induced bias.

Fig. 4A illustrates the PCA loadings for HSI spectral data within the 400–1000 nm range following SNV preprocessing. Here, PC1 demonstrated a high positive contribution in the Soret band (415 nm), while both PC1 and PC2 showed significant contributions within the Q band, reflecting the high reflectance of ferric hemoglobin in this range—features associated with bloodstain aging. Additionally, PC1 displayed pronounced fluctuations between 600–900 nm, where oxyhemoglobin exhibits high reflectance with increasing concentration, suggesting that this spectral range is indicative of deposition time. PC2 also exhibited notable fluctuations within the 500–700 nm and 800–900 nm regions, with positive loadings indicating greater contributions in these bands.

Fig. 4B shows the PCA loadings for NIR data in the 900–1680 nm range after SNV preprocessing. Here, PC1 predominantly reflected water content in bloodstains, remaining stable from 1000–1200 nm but showing marked changes beyond 1300 nm, especially within the 1500–1600 nm range, where amide bond vibrations become prominent. These bands are strongly associated with shifts in water content. Additionally, combination bands from O–H stretching and bending vibrations appeared around 850 nm. PC2 primarily captured features of the protein band region, with higher PC2 loadings corresponding to stronger absorption within this region. Significant fluctuations were noted between 900–1000 nm, 1100–1400 nm, and 1500–1600 nm, indicative of organic components such as proteins, lipids, and water. The range from 1300–1500 nm, in particular, corresponded to combination bands of C–H and O–H stretching vibrations, likely reflecting compositional changes in the blood and contributing to bloodstain deposition time identification.

4.3 Homogeneous data fusion regression model

A simple linear regression model was employed to evaluate the effect of material type (permeable *vs.* impermeable) on predicting bloodstain deposition time, with significance assessed using a *t*-test ($P < 0.05$ denoting a significant effect). As illustrated in Fig. 5, the *P*-values for both HSI and NIR spectra were below 0.05, indicating a significant influence of material type on bloodstain deposition time. Analysis of residual box plots revealed that on permeable materials, NIR residual fluctuations were markedly larger than those for HSI. This outcome suggests that the greater penetrative capacity of NIR made it more susceptible to material-type variations on permeable surfaces, resulting in a wider residual range. Conversely, on non-permeable materials, both NIR and HSI exhibited reduced residual fluctuations, with NIR displaying a narrower residual range, indicating a more stable response on impermeable surfaces. These results highlight that combining HSI and NIR modalities could enhance detection capability across diverse material types, overcoming the limitations of each individual technique on specific substrates and improving the stability and accuracy of detection results.

The PLS algorithm, which extracts latent variables from high-dimensional data, is particularly advantageous in cases of multicollinearity between independent and dependent variables, making it commonly applied in spectral data analysis and modeling. Here, a simple PLS regression was implemented to predict bloodstain deposition time (up to 60 days) using HSI and NIR data. The results, detailed in Table S1,† indicated a significant difference between RMSECV and RMSEC, suggesting underfitting in the test set and reflecting the model's difficulty in capturing complex nonlinear data features, which led to substantial prediction errors. To address this, a polynomial extension was introduced, enhancing the model's capacity to account for nonlinear variation, significantly improving fitting performance for extended prediction intervals. In *K*-fold cross-validation, the optimal number of latent variables was determined to be 8 when $K = 3$, achieving the lowest RMSECV. As presented in Table 2, R^2 values exceeded 0.8 for both HSI and NIR, with minimal difference between RMSEC and RMSECV. R^2 values for impermeable materials were higher at 0.85 and 0.88, corroborating that impermeable surfaces retain more blood moisture, resulting in subtler spectral variations.

To improve model generalizability and stability across varying material types, the study controlled for material influence by combining data from different substrates (by directly concatenating raw data) and applying $K = 3$ cross-validation to determine latent variables. Post-fusion prediction performance, as shown in Table 2, was lower than pre-fusion values. The R^2 value for HSI spectral data decreased to 0.76, significantly lower than pre-fusion values of 0.80 and 0.85. RMSEP values for HSI and NIR were 8.35 and 8.15 days, respectively, both higher than results from material-specific models. Furthermore, RMSECV values for HSI and NIR were 7.17 and 7.98 days, with Bias values of 2.98 and 2.23, indicating model stability with no signs of overfitting or underfitting. However, the increased heterogeneity from merging data across materials intensified sample

complexity and nonlinear relationships. Although polynomial regression captured some nonlinear features, fully addressing the added complexity proved challenging, resulting in elevated RMSEP and prediction bias.

4.4 Multimodal data fusion regression model

HSI and NIR data demonstrated notable complementarity in predicting bloodstain deposition time, with substrate material exerting a significant influence on deposition characteristics. To improve model generalizability across different substrates, multimodal data fusion was applied to spectral data from various materials and sources. The resulting multimodal data fusion regression model, constructed using the PLS algorithm in combination with a multilayer perceptron (MLP), followed the fusion methods detailed in Section 3.5. Results from data fusion are presented in Table 3. For low-level data fusion, five-fold cross-validation ($K = 5$) yielded 8 latent variables, producing an R^2 value of 0.87, a root mean square error of prediction (RMSEP) of 3.12 days, and a difference of 0.73 between RMSECV and RMSEC, indicating reliable model fit across both training and test sets. In intermediate-level data fusion, the model again employed 8 latent variables, achieving an improved R^2 of 0.95, a reduced RMSEP of 2.59 days, and a smaller RMSECV–RMSEC difference of 0.4. Prediction biases were measured at 0.24 for low-level fusion and 0.20 for intermediate-level fusion, confirming that data fusion enhances prediction accuracy, with intermediate-level fusion exhibiting superior performance.

Fig. 6 displays residual plots for the test sets of both low-level and intermediate-level data fusion. Residuals were distributed evenly around the zero line, with low-level fusion residuals ranging from −15 to 10 and intermediate-level fusion residuals from −10 to 10. In the first 20 days, residuals clustered more closely, especially in the intermediate-level fusion model. The residual distribution of intermediate-level fusion also exhibited greater symmetry than that of low-level fusion. Fig. 7 compares experimental and predicted values for the low-level and intermediate-level models, with 10% of data from both training and test sets selected for plotting. In the training set (Fig. 7A.1), the low-level fusion model showed a strong overall fit, with most data points aligned closely with the reference line ($y = x$). However, in the high-value range (>40), deviations emerged. In the test set (Fig. 7A.2), model performance weakened in this high-value range, as evidenced by sparse data point distribution, suggesting limitations in capturing complex multimodal features and reduced model stability. In the training set for intermediate-level fusion (Fig. 7B.1), results showed a dense distribution of data points around the reference line, particularly in the high-value range (>40), with minimal deviation, indicating robust extraction of nonlinear features. In the test set (Fig. 7B.2), the model exhibited high accuracy in the low-to-medium range (0–40), with only minor deviations in the high-value range (>50). Overall, predictions aligned more closely with experimental values, affirming the model's generalization capability and robustness over the low-level fusion approach, thus underscoring the effectiveness of intermediate-level data fusion.

5 Conclusion

This study leverages two complementary spectroscopic techniques—HSI and NIR spectroscopy—to determine bloodstain deposition time (TSD) in forensic science. HSI is capable of effectively identifying early-stage bloodstains through the detection of hemoglobin structural changes, whereas NIR offers a more comprehensive analysis by monitoring moisture loss and utilizing its deep penetration properties. The integration of these two techniques facilitates more precise TSD assessment in complex environments. Both HSI and NIR spectra demonstrate considerable robustness, and the use of SNV preprocessing effectively removes spectral noise. Polynomial PLS regression and MLP models were employed to address nonlinear characteristics, thereby enhancing predictive performance. In cross-material data fusion results, the predictive accuracy of the model was higher for NIR spectra (RMSEP = 8.15 days) compared to HSI spectra (RMSEP = 8.35 days). Multimodal data fusion achieved RMSEP values of 3.12 days for low-level fusion and 2.59 days for intermediate-level fusion, yielding robust results for TSD prediction up to 60 days. Although this study successfully minimized the influence of material types on TSD, practical applications may still face challenges from variables such as contaminants (*e.g.*, coffee and lipstick) and a wider diversity of materials. While HSI and NIR spectroscopy exhibit complementary capabilities and adaptability, more complex models may be required to provide comprehensive predictions in these scenarios. Additionally, as bloodstains age, spectral features become increasingly difficult to detect, leading to reduced prediction accuracy over time, an issue warranting further investigation.

This study highlights the potential for longer, more precise TSD determination through the synergistic application of different spectroscopic techniques and advanced chemometric methods. The multidimensional analytical approach outlined here enhances forensic science's capacity for temporal inference in criminal investigations and sets the groundwork for future research aimed at bridging laboratory findings with practical forensic applications.

Data availability

Data available on request from the authors.

Author contributions

Ying Liang was responsible for writing – original draft and writing – review & editing, Jianqiang Zhang, Jiaquan Wu for funding acquisition, Qifu Yang for validation, Huihui Ren, Kun Ma, Hanyu Zhu, Xingshuai Peng, Jiateng Wang and Xinyu Zhang for conceptualization.

Conflicts of interest

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.

Acknowledgements

This work was financially supported by Physical Evidence Spectral Technology Innovation Team of Yunnan Police College in Yunnan Province (202105AE160007), Key Laboratory of Spectral Technology Physical Evidence of Education of Yunnan Province, Basic Research Project of Yunnan Police College (21A028) and Yunnan Provincial Key Laboratory of Forensic Science (2020zz02, 2020zz07).

References

1 R. H. Bremmer, K. G. Bruin, M. J. C. Gemert, T. G. Leeuwen and M. C. G. Aalders, *Forensic Sci. Int.*, 2012, **216**, 1–11.
2 B. P. Epstein, *J. Forensic Sci.*, 2006, **51**, 462.
3 S. S. Tobe, N. Watson and N. N. Daéid, *J. Forensic Sci.*, 2007, **52**, 102–109.
4 E. D. Vittori, F. Barni, S. W. Lewis, G. Antonini, C. Rapone and A. Berti, *Forensic Chem.*, 2016, **2**, 63–74.
5 F. Barni, S. W. Lewis, A. Berti, G. M. Miskelly and G. Lago, *Talanta*, 2007, **72**, 896–913.
6 G. Patel and A. Hopwood, *J. Leg. Med.*, 2013, **127**, 723–729.
7 J. P. Almeid, N. Glesse and C. Bonorino, *Forensic Sci. Int.*, 2011, **206**, 58–61.
8 N. S. Alenazy, A. M. Refaat and S. R. Babu, *Egypt. J. Forensic Sci.*, 2015, **5**, 103–108.
9 V. Stewart, P. Deacon, N. Zahra, M. L. Uchimoto and K. J. Farrugia, *Sci. Justice*, 2018, **58**, 386–396.
10 S. Bell, *Forensic science: an introduction to scientific and investigative techniques*, CRC press, 2019, vol. 29, p. 3.
11 B. Budowle, J. L. Leggitt, D. A. Defenbaugh, K. M. Keys and S. F. Malkiewicz, *J. Forensic Sci.*, 2000, **45**, 1090–1092.
12 G. J. Edelman, T. G. van Leeuwen and M. C. G. Aalders, *Proc. SPIE*, 2013, **8**, DOI: **10.1117/12.2021509**.
13 S. Gariglio, C. Malegori, A. Menżyk, G. Zadora, M. Vincenti, M. Casale and P. Oliveri, *Talanta*, 2024, **278**, 126444.
14 R. Zhang, P. Wang, J. Chen, Y. Tian and J. Gao, *Spectrochim. Acta, Part A*, 2023, **290**, 122–284.
15 H. Lin, Y. Zhang, Q. Wang, B. Li, S. Fan and Z. Wang, *J. Leg. Med.*, 2018, **132**, 667–674.
16 G. S. Cooney, H. Köhler, C. Chalopin and C. Babian, *Forensic Sci., Med., Pathol.*, 2024, **20**, 490–499.
17 K. Wang, X. Bian, M. Zheng, L. Peng, L. Ligang and T. Xiaoyao, *Spectrochim. Acta, Part A*, 2021, **263**, 120–138.
18 W. Yangzhe, H. Yi, C. Jiye C, M. Shuyuan, W. Xiaoping, C. Yong and P. Yunlong, *Micron*, 2009, **40**, 359–364.
19 S. Arany and S. Ohtani, *Forensic Sci. Int.*, 2011, **212**, 1–3.
20 S. Huimin, D. Yongfang, Z. Pingli, M. Yaoyong, W. Wei, L. Nan and G. Zhiyou, *IEEE Photonics J.*, 2017, **9**, 1–14.
21 E. Gerda, M. Vicky, M. Saskia M van Ruth, L. Ton van and A. Maurice, *Forensic Sci. Int.*, 2012, **220**, 239–244.
22 E. Gerda, L. Ton G van and A. C. G. Maurice, *Forensic Sci. Int.*, 2012, **223**(1–3), 72–77.
23 L. Bo, B. Peter, O. William and I. Meez, *Sci. Justice*, 2013, **53**(3), 270–277.
24 P. Oliveri, C. Malegori and M. Casale, *Chem. Anal. Food*, 2020, 33–76.
25 S. Wold, K. Esbensen and P. Geladi, *Chemom. Intell. Lab. Syst.*, 1987, 37–52.
26 J. S. Cleveland and A. D. Weidemann, *Limnol. Oceanogr.*, 1993, **38**, 1321–1327.
27 R. J. Barnes, M. S. Dhanoa and S. J. Lister, *Appl. Spectrosc.*, 1989, **43**, 772–777.
28 S. Wold, A. Ruhe, H. Wold and W. J. Dunn, *SIAM J. Sci. Comput.*, 1984, **5**, 735–743.
29 Wold, M. Sjöström and C. J. L. Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.
30 C. Federico, *Sci. World J.*, 2013, DOI: **10.1155/2013/704504**.
31 D. L. Hall and J. Llinas, *Proc. IEEE*, 1997, 6–23.
32 B. H. Rolf, N. Annemarie, L. Ton G van, G. Martin J C van and A. G. C. Maurice, *Forensic Sci. Int.*, 2011, 166–171.
33 S. S. Kind, D. Patterson and G. W. Owen, *Forensic Sci.*, 1972, 27–54.
34 W. M. Carrie, *Crystallogr. Rev.*, 2014, 224–225.
35 S. Fathima, D. M. Dhammika, B. A. Nicholas and D. H. Andrew, *Ann. Emerg. Med.*, 2010, 184–189.
36 Å. Rinnan, F. van den Berg and S. Engelsen, *TrAC, Trends Anal. Chem.*, 2009, 1201–1222.
37 H. Jansson and J. Swenson, *J. Chem. Phys.*, 2008, **24**, 128.