Research on Rapid Identification Technology of Microfiber Evidence in Criminal Cases Based on Raman Spectroscopy and Deep Learning

Hui Sun
Investigation Department, Shaanxi Police College, Xi'an 710021, Shaanxi, China
Corresponding author: Hui Sun (e-mail: qiqibaobei0330@126.com)

ABSTRACT
Rapid and accurate identification of trace microfiber evidence in criminal investigations is essential for forensic analysis and relies on precise characterization of textile materials. As an important electromagnetic wave–matter interaction technique, Raman spectroscopy provides non-destructive optical sensing capabilities for revealing the molecular composition and structural characteristics of fiber samples. This study integrates Raman spectroscopy with a deep learning framework to establish an intelligent identification system for microfiber evidence. A total of 8,500 high-quality Raman spectra covering 15 common and blended fiber types were collected, and a preprocessing workflow including baseline correction, smoothing, denoising, and normalization was developed. A one-dimensional convolutional neural network was then constructed for spectral feature extraction and classification. Experimental results demonstrate that the proposed model achieves an overall recognition accuracy of 98.7% on an independent test set. The results confirm that the combined approach effectively captures the chemical signatures of complex textile fibers and enables reliable identification of trace evidence, providing a robust technical solution for forensic analysis and electromagnetic spectroscopy-based intelligent material recognition.

INDEX TERMS
raman spectroscopy, deep learning, fiber evidence, electromagnetic spectroscopy, optical sensing

I. INTRODUCTION
FORENSIC identification factors are critical in criminal science because they help reconstruct the facts of a case by establishing precise links between suspects and textile fiber clues recovered from crime scenes. By analyzing the unique microscopic morph ology, dye composition, and spinning structures of these textile fragments, investigators can scientifically validate the connection between specific materials and criminal evidence. Among them, fiber, as a common trace evidence, can easily remain in the crime scene, on the victim's clothes, and in places where the suspect himself comes into contact or transfer during the crime process. Correct determination of these fibers can also provide vital scientific evidence from which to infer weapons, crime scene reconstructions, and causal relationships between suspects and crime scenes. Therefore, obtaining fast, accurate, and non-destructive fiber identification technology is the frontier and research topic of criminology. This study focuses on the key conditions for rapid detection of trace fibers in criminal investigation practice. Committed to systematically learning the combination of Raman spectroscopy analysis technology and deep learning algorithms to develop effective and reliable intelligent identification technology. The given system will be aimed at addressing the signal preprocessing issues of the Raman spectroscopy in fiber analysis and constructing a deep-learning model that meets the features of spectral data. Finally, studies on large-scale experimental data are required to confirm its accuracy, efficacy and timeliness in identifying multiple categories of fibers. The results of this research promote the transformation of fiber evidence review from a traditional expert-based model to an automated paradigm that integrates high-precision textile morphology and spinning data. This development enhances the accuracy of evidence identification by leveraging a digital textile database to provide more powerful capabilities for forensic applications.

II. RELATED WORK
The non-destructive nature of Raman spectroscopy and its very specific characteristics have led to a series of theoretically groundbreaking methodological advances in the fields of analytical chemistry and materials science. Ibtehaz

et al. [1] proposed RamanNet to process one-dimensional sequence data of Raman spectra. Adar et al. [2] studied how the latter uses the intensity, ratio and displacement of individual peak positions to provide evidence of protein secondary structure composition and clarify and integrate the interpretation rules of protein Raman spectra. Wu et al. [3] outlined a self-supervised learning framework for Raman spectroscopy denoising, which is a general framework designed to train denoising models without clean labeled data. Li et al. [4] proposed a new paradigm for matching Raman spectra based on contrastive representation learning. Zhao [5] applied attenuated total reflection infrared spectroscopy (ATR) to measure the infrared spectrum, conducted a comprehensive analysis of the infrared spectrum of common fibers, and used each absorption peak in the infrared spectrum to identify fiber samples. Wang et al. [6] reviewed the relevant research on thermal damage to human hair and textile fibers at home and abroad, in order to provide a reference for related research and accident investigation. Zhong and Yang [7] used scanning electron microscopy to analyze the microscopic morphology of fibers, distinguished the fibers in appearance, and used infrared spectroscopy to analyze the functional group structure of fibers to determine the fiber type. Zhao and Liu [8] discussed three cases of using EDS component analysis to identify electrical melt marks, and put forward the precautions in the sampling of physical evidence at fire scene, melt mark identification and the application of identification opinions. Hu et al. [9] reviewed the current status of the application of pyrolysis gas chromatography in the examination and identification of trace physical evidence such as combustion residues, biological materials, rubber, paint, ink toner, fibers, organic matter in soil and wood chips. Song et al. [10] collected 600 Raman spectral data and expanded them to 6000, established 12 classification prediction models and compared and discussed their identification effects.

Existing research has made significant progress in the advancement of Raman spectroscopy algorithms and the accumulation of methods for the identification of scientific fibers. Based on this, this paper proposes a rapid identification technology system based on Raman spectroscopy and deep learning, specifically designed for criminal trace fiber evidence, to address the increasingly complex challenges of evidence examination and improve the scientific and standardized level of identification.

III. METHODS

A. CONSTRUCTION OF FIBER SAMPLE LIBRARY AND ACQUISITION OF RAMAN SPECTROSCOPY DATA

The sample collection focused on common fiber types in criminal cases, and 15 categories were finally identified, including natural fibers (cotton, wool, silk, linen), synthetic fibers (polyester, nylon 6, nylon 66, acrylic, spandex, viscose, acetate) and blended fibers commonly seen in actual cases (such as cotton/polyester, wool/acrylic, polyester/spandex, nylon/spandex). All samples were cross-validated using Fourier transform infrared spectroscopy and microscopy to verify the correctness of the class labels. Sample preparation was carried out in accordance with the precautions for handling trace evidence to ensure no contamination.

Spectrum collection was performed in a confocal micro-Raman spectrometer with a laser wavelength of 785 nm, which effectively reduced fluorescence background interference. The spectrum collection range is adjusted to 200–1800 cm^{-1}, covering the main fingerprint area of fiber chemical bond vibration.

Measurements of no fewer than 10 points were selected for each fiber sample to include possible microscopic inhomogeneities (including orientation differences) in the fibers themselves. Optimize laser power and integration time in a systematic manner to provide a compromise between maintaining signal strength and preventing thermal damage to the sample. Finally, we obtained to our satisfaction 28,500 raw spectra of 495 independent samples. After strict quality assurance (i.e., signal strength and signal-to-noise ratio), 8,500 high-quality Raman spectra were ultimately retained for future research by eliminating spectra containing weak signals or obvious anomalies [11]. Of the 28,500 original spectra, nearly 70% were discarded due to low signal-to-noise ratio, strong fluorescence interference, or spectral anomalies. This was mainly due to the strong fluorescence background of some fiber samples and insufficient optimization of laser power and integration time parameters in the initial stage of the experiment. Subsequent experiments significantly improved the spectral quality by optimizing acquisition parameters and increasing the number of measurement points. The statistics of Raman spectrum data collection are shown in Table 1:

TABLE 1. Statistics on Raman spectral data acquisition for 15 types of fiber samples

Fiber type number	Fiber type description	Number of independent samples	Total number of spectra collected	Spectral data through quality control
F01	cotton	35	2100	601
F02	wool	32	1920	549
F03	silk	30	1800	515
F04	flax	28	1680	481
F05	Polyester	40	2400	687
F06	Nylon 6	38	2280	653
F07	Nylon 66	36	2160	618
F08	acrylic fiber	33	1980	567
F09	spandex	29	1740	498
F10	viscose fiber	31	1860	532
F11	cellulose acetate	27	1620	464
F12	Cotton/polyester blend (65/35)	34	2040	584
F13	Wool/Acrylic blend (70/30)	30	1800	515
F14	Polyester/spandex blend (85/15)	37	2220	635
F15	Nylon/spandex blend (80/20)	35	2100	601
Total	15 categories	495	28,500	8500

The sample library covers key fiber types used in criminal investigations, with sufficient independent samples and measurement replicates for each type, providing statistical robustness for model training. The inclusion of blended fibers enhances the model's ability to handle complex real-world evidence. Rigorous quality control ensures that subsequent analyses are based on high signal-to-noise ratio data.

B. SPECTRAL PREPROCESSING AND DATA AUGMENTATION WORKFLOW

Raw Raman spectra typically contain instrument noise, fluorescence background, and intensity fluctuations, necessitat-

ing preprocessing to highlight characteristic signals relevant to chemical structure. The preprocessing procedure in this study comprises three sequential steps:

Baseline Correction

An asymmetric weighted penalized least squares method is employed. This method adaptively estimates and subtracts complex fluorescence baselines while preserving the original shape and intensity information of the Raman peaks.

Smoothing and Denoising

A Savitzky-Golay filter (window size 9, polynomial order 3) is used for smoothing. This filter effectively suppresses random noise through local polynomial fitting and has minimal distortion of the signal peak shape.

Vector Normalization

Normalize the spectrum after each baseline correction and smoothing to make its Euclidean norm (modulus) 1. This eliminates the absolute intensity changes caused by laser power fluctuations, sampling position differences, etc., and makes the model focus on the relative differences in spectral shape [12].

To improve the generalization ability and robustness of deep learning models and prevent overfitting on small datasets, data augmentation strategies were implemented. Augmentation was performed on the normalized spectra, and the main methods included spectral shifting, adding Gaussian white noise, and random intensity scaling (scaling factors ranging from 0.9 to 1.1). Each training set spectrum was used to generate two augmented samples by randomly combining these operations, thereby tripling the amount of training data. Data augmentation is applied only to the training set to increase the diversity of training samples. The test and validation sets use the original, unaugmented spectral data to ensure the authenticity and generalization ability of the evaluation results.

C. DEEP LEARNING MODEL ARCHITECTURE AND TRAINING CONFIGURATION

The core classification task is implemented through a specially designed one-dimensional deep convolutional neural network. The one-dimensional data input to the Raman spectroscopy (length 1600, i.e. range 200–1800 cm^{-1}) is preprocessed data. It has a network topology consisting of four consecutive convolution blocks with 1D convolutional layers (kernel size 5), batch normalization, ReLU activation function, and max pooling (stride 2 to reduce dimensionality and add translation invariance). The hierarchical network is trained to generate features in hierarchical groups, local maxima at the base layer, and global spectra at the top layer. A global average pooling layer follows the convolution, thereby converting the feature maps into small feature vectors. Finally, the probability distribution of 15 categories is obtained through the connection of the fully connected layer and the softmax activation function [13,14]. The goal of model training is to reduce the cross-entropy loss of classification tasks. The optimization process used is Adam with an initial learning rate of 0.001 and a learning rate decay strategy implemented. In order to avoid overfitting, Dropout technology (dropout rate 0.5) is used before the fully connected layer, and the L2 attenuation term of the loss function is introduced. The data were randomly divided into training set (70%), validation set (20%), and test set (10%). The training set is used to update model parameters, the validation set is used to monitor the training process and fine-tune hyperparameters and pass early stopping decisions, and the test set is used to measure the generalization performance of the model. Each experiment was run with a fixed random seed value to guarantee reproducible results [15].

IV. RESULTS AND DISCUSSION

A. COMPREHENSIVE EVALUATION AND COMPARATIVE ANALYSIS OF MODEL RECOGNITION PERFORMANCE

Its performance was fully assessed after model training in completely independent test set (with 850 spectra). The suggested deep machine learning model demonstrated great classification outcomes. Its confusion matrix indicates that the test samples of the statistically significant majority were correctly as well with the majority of the values on the main diagonal being much greater than the off-diagonal. The total recognition rate of the model made 98.7%. Further examination of the precision, recall, F1 score of individual categories revealed that all these score values were over 97%, and the performance difference between categories was slight. The macro-mean F1 score was 98.6%, which means that the model had a very stable and correct recognition ability of all fibers. To highlight the advantages of deep learning methods, we compared them with three classic machine learning models: Support Vector Machine (using radial basis function kernels), Random Forest (100 decision trees), and K-Nearest Neighbors (K=5). The comparison models used the same training/validation/test set partitioning. For these traditional models, features need to be manually extracted from the preprocessed spectrum. In this experiment, we extracted the wavelength, intensity, peak area ratio of the main peak, and the scores of the top 20 principal components after dimensionality reduction via principal component analysis as feature inputs. We introduced an ensemble learning model based on extreme gradient boosting (XGBoost), which has proven to significantly outperform traditional partial least squares regression in pulp fiber performance prediction tasks. XGBoost effectively captures the nonlinear relationship between spectral features and fiber category by iteratively training weak classifiers and optimizing the loss function. Results of different fiber recognition models on the test set are shown in Figures 1 and 2:

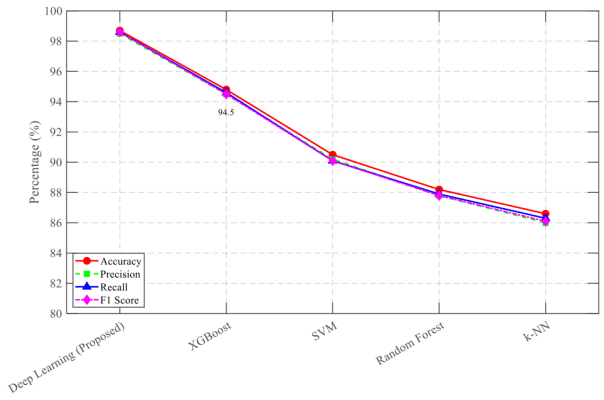

FIGURE 1. Recognition performance

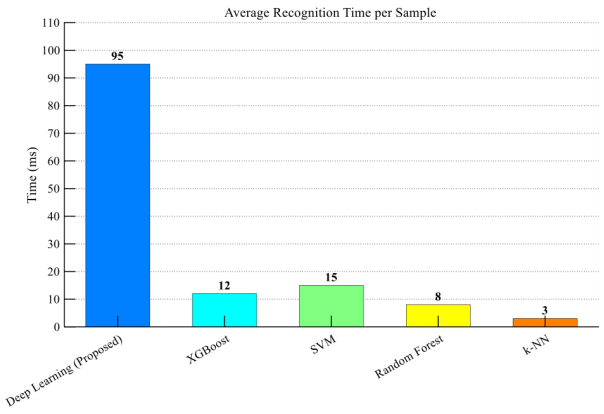

FIGURE 2. Recognition Time

Figure 2 results analysis:

The comparison results clearly show that the deep learning model has an absolute advantage in recognition accuracy, exceeding the second-best performing 1D-CNN model by 2.5 percentage points. This is mainly due to the fact that CNNs can automatically learn relevant features from raw data, avoiding information loss and subjective bias that may be introduced by manual feature engineering. Although traditional models have faster inference speed, their lower accuracy makes it difficult to meet the high reliability requirements of criminal identification.

B. IN-DEPTH ANALYSIS OF MISJUDGMENT CASES AND DISCUSSION OF MODEL DECISION-MAKING MECHANISM

Although the overall performance of the model is very good, in-depth analysis of a few false positive cases can help understand its limitations and directions for future improvements. Classification errors mainly include the following errors: First, fibers have highly similar chemical structures; Nylon 6 and nylon 66, their macromolecular chains are arranged slightly differently, resulting in only small differences in their Raman spectra; second, the mixed fibers are filled with similar proportions of components, where the model may be biased towards the loudest components. To understand the process by which the model makes decisions, a gradient-weighted class activation derived derivative method is used to generate saliency maps to help visualize the wavenumber portions of the input spectrum that contribute most to the classification choices made by the model. The study found that when correctly classifying the model, it paid great attention to the characteristic peaks related to specific chemical bond vibrations (such as the C=O stretching vibration peak of the ester carbonyl group near 1720 cm^{-1} in polyester; the breathing vibration peak of the benzene ring). However, if the classification is incorrect, its focus may be in an important discriminating region or be disturbed by non-characteristic variation. This goes a long way in confirming that the model is not a black box type, but in fact does learn the spectral physics related to the chemical composition of the fiber. Visualization not only improves model interpretability but also provides chemists with a mechanism to ensure that model decisions are sound.

V. CONCLUSION

This study successfully constructed and validated a technical system for the rapid identification of trace fiber evidence by deeply integrating Raman spectroscopy with deep learning algorithms to analyze the chemical signatures of diverse textile polymers and spinning additives. This integration ensures that microscopic variations in textile weaving structures and dye compositions are accurately captured, providing a robust forensic tool for distinguishing between complex industrial materials. Through systematic data acquisition, the study obtained a large-scale, high-quality Raman spectral dataset covering 15 common fiber types. An effective spectral preprocessing workflow was designed and optimized, significantly improving the quality of the original spectral data. The constructed deep convolutional neural network model automatically extracts discriminative features from spectral data to achieve high-precision identification of multiple fiber categories, including complex blends of natural and synthetic textile materials. This technical approach effectively maps the subtle chemical and structural variations produced during the textile spinning and weaving processes to ensure the accurate classification of microscopic forensic evidence.

Future work will focus on the following aspects: First, expanding the fiber type database, especially more types of blended fibers and fibers that have undergone special treatments (such as dyeing and flame retardancy); second, exploring the model's ability to identify severely degraded or contaminated fibers and studying corresponding enhancement strategies; and third, attempting to integrate the model with miniaturized Raman probes to develop a prototype system that can be used for preliminary on-site screening, ultimately making a greater contribution to achieving higher efficiency and precision in criminal science and technology.

AUTHOR CONTRIBUTIONS

Hui Sun designed, collected and analyzed the data, and drafted the manuscript. Hui Sun conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Hui Sun participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This work was financially supported by the Shaanxi Police College 2023 Research and Innovation Team - Trace Material Evidence Inspection and Analysis Research Center Project.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] N. Ibtehaz, M. E. H. Chowdhury, A. Khandakar, S. Kiranyaz, M. S. Rahman, and S. M. Zughaier, "RamanNet: a generalized neural network architecture for Raman spectrum analysis," Neural Computing and Applications, vol. 35, no. 25, pp. 18719-18735, 2023, doi: 10.1007/s00521-023-08700-z.

[2] F. Adar, N. Zheng, Z. Wang, F. Zhao, H. Li, and H. He, "Interpretation of Raman spectrum of proteins," Spectroscopy, vol. 37, no. 2, pp. 9-13, 2022, doi: 10.56530/spectroscopy.lo2270l5.

[3] S. Wu, Y. Zhang, C. He, Z. Luo, Z. Chen, and J. Ye, "Self-Supervised Learning for Generic Raman Spectrum Denoising," Analytical Chemistry, vol. 96, no. 44, pp. 17476-17485, 2024, doi: 10.1021/acs.analchem.4c01550.

[4] B. Li, M. N. Schmidt, and T. S. Alstrøm, "Raman spectrum matching with contrastive representation learning," Analyst, vol. 147, no. 10, pp. 2238-2246, 2022, doi: 10.1039/D2AN00403H.

[5] S. Zhao, "Fourier transform infrared spectroscopy analysis and application of trace fiber evidence," Journal of Jilin Chemical Engineering Institute, vol. 42, no. 3, pp. 86-90, 2025.

[6] P. Wang, Z. Zang, J. Di, and J. Jin, "Research progress in thermal damage testing of fiber evidence," Fire Science and Technology, vol. 43, no. 6, pp. 892-895, 2024.

[7] Y. Zhong and L. Yang, "Application examples of scanning electron microscopy and infrared spectroscopy in fiber evidence analysis," Shandong Chemical Industry, vol. 53, no. 6, pp. 152-155, 2024.

[8] X. Zhao F; Liu, "Application of EDS component analysis in the identification of electrical melt marks," Shandong Chemical Industry, vol. 53, no. 8, pp. 167-169, 2024.

[9] Y. Hu, X. Sun, and J. Jin, "Research progress on the application of pyrolysis gas chromatography in the examination and identification of trace evidence," Guangdong Chemical Industry, vol. 50, no. 7, pp. 200-202, 2023.

[10] L. Song, J. Meng, F. Zhang, S. Nan, and Y. Tao, "Research progress on trace evidence detection based on Raman spectroscopy," Guangdong Chemical Industry, vol. 52, no. 21, pp. 53-5569, 2025.

[11] A. A. Timoshenko, "Procedural Conditions for the Admissibility of an Expert's Opinion as Evidence in a Criminal Case," Russian Journal of Legal Studies (Moscow), vol. 11, no. 1, pp. 81-85, 2024, doi: 10.17816/RJLS629892.

[12] S. Suhadi, S. Muchtar, and A. M. Muin, "Implementation of borrow-used evidence criminal case in Makassar City," The International Journal of Politics and Sociology Research, vol. 11, no. 1, pp. 119-129, 2023, doi: 10.35335/ijopsor.v11i1.109.

[13] M. H. Ahmad, A. S. Baharuddin, H. Hashim, R. Razak, N. S. Saharudin, and S. N. Omar, "Forensic evidence as a mean of proof in prima developing facie case in takhbib criminal offence," UUM Journal of Legal Studies (UUMJLS), vol. 13, no. 01, pp. 221-248, 2022, [Online]. Available: https://ejournal.uum.edu.my/index.php/uumjls/article/view/15618.

[14] C. M. Morăreanu-Dragnea, "ADMINISTRATION OF EVIDENCE IN THE TRIAL PHASE OF A CRIMINAL CASE," Studii Juridice şi Administrative, vol. 32, no. 1, pp. 107-118, 2025, [Online]. Available: https://www.ceeol.com/search/article-detail?id=1392043.

[15] S. Ghignone, M. Prencipe, P. Manzotti, M. Bruno, F. Boero, A. Borghini, et al., "The Raman spectrum of florencite-(REE)[REEAl3 (PO4) 2 (OH) 6]: An integrated experimental and computational approach," Journal of Raman Spectroscopy, vol. 55, no. 3, pp. 394-405, 2024, doi: 10.1002/jrs.6640.