

## Rapid Identification of the Species of Bloodstain Based on Near Infrared Spectroscopy and Convolutional Neural Network-Support Vector Machine Algorithm

Ying Liang,<sup>1b</sup> Jiaquan Wu,<sup>a</sup> Qi Zeng,<sup>b</sup> Yunxia Zhao,<sup>c</sup> Kun Ma,<sup>a</sup> Xinyu Zhang,<sup>a</sup>  
Qifu Yang,<sup>1b</sup> Jianqiang Zhang<sup>\*,d</sup> and Yueying Qi<sup>\*,d</sup>

<sup>a</sup>College of Science, Kunming University of Science and Technology, 650500, Kunming, China

<sup>b</sup>Yunnan Province Public Security Department, 65000, Kunming, China

<sup>c</sup>Wenshan Public Security Bureau, 663000, Wenshan, China

<sup>d</sup>Yunnan Police College, 650000, Kunming, China

As one of the most important types of evidence at the scene of the crime, the rapid identification of the human bloodstain is of great significance to solve the criminal case. In this paper, the spectral data of different species of bloodstain samples including human, chicken and pig were acquired by using a hand-held near-infrared spectrometer. Then, the training models were established via convolutional neural network-support vector machine algorithm. Meanwhile, the traditional support vector machine, genetic algorithm-back propagation and random forest classification algorithms were also compared. The results showed that the prediction accuracy of convolutional neural network-support vector machine algorithm was the highest and the overall performance of the model was the best. The rapid detection method based on a handheld near-infrared spectrometer and convolutional neural network-support vector machine algorithm could identify the species of bloodstain efficiently, non-destructively, quickly and accurately and it provided a new technical reference for bloodstains detection and identification.

**Keywords:** near-infrared, bloodstains identification, convolutional neural network-support vector machine, non-destructive, rapidly

### Introduction

Blood is one of the most important types of forensic evidence and it can usually be found at crime scenes.<sup>1</sup> In violent crimes, the identification of blood is one of the most critical steps in a crime scene analysis, as it provides information about the dynamics of the crime and the presence of suspects.<sup>2</sup> The blood at the scene of the crime does not exist in isolation and is generally deliberately destroyed by the suspect, which makes the blood at the crime scene difficult to distinguish intuitively, especially the human blood and the animal blood. At present, the identification of blood mainly depends on the experiences of forensic experts<sup>3</sup> and biochemical examination.<sup>4</sup> Forensic experts are subjective and the biochemical examination is expensive. Some spectral detection methods such as hyperspectral imaging and Raman have also been used in

the identification of blood.<sup>5-7</sup> However, the hyperspectral imager is bulky, expensive and the timeliness is not high. In Raman spectroscopy, different vibration peaks may overlap and the scattering intensity is easily affected by factors such as optical system parameters. As a result, it is important to develop a novel spectral detection method which is convenient, cheap and rapid.

Near infrared spectroscopy (NIR) technique is a highly efficient and rapid modern analysis tool. It has the advantages of being rapid, non-destructive, and low cost for sample analysis.<sup>8</sup> It has been widely adopted in various fields, such as agriculture,<sup>9</sup> petrochemical industries,<sup>10,11</sup> medicine,<sup>12</sup> etc. However, there is little research on the detection of blood. The plasma is the liquid portion of blood. The main constituents of plasma are water, proteins and other soluble fractions that contain plenty of O–H, N–H and C–H. Therefore, NIR is suitable for the detection of blood.

As an excellent classification algorithm, convolutional neural network-support vector machine (CNN-SVM) has

\*e-mail: zhangjqxy@163.com; 2363783339@qq.com

Editor handled this article: Eduardo Carasek



been developed rapidly, especially in the field of satellite remote sensing,<sup>13</sup> medicine,<sup>14</sup> image classification,<sup>15</sup> and identification of plant disease.<sup>16</sup> CNN-SVM algorithm is outstanding and has the advantages of convolutional neural network (CNN) in feature extraction and support vector machine (SVM) in the classification. It also maximizes the generalization and accuracy of classification.<sup>17</sup> However, it has not been sufficiently used in the field of spectrum analysis. The dimension of NIR spectral data is huge and usually contains noise. CNN-SVM algorithm can integrate feature extraction and classification recognition, and it is suitable for dealing with the high-dimensional NIR spectral data.

In this study, the identification of bloodstain method based on CNN-SVM algorithm and NIR technology was proposed. The experiment results showed that the CNN-SVM model achieved the best performance compared with the traditional SVM model, random forest (RF), and genetic algorithm-back propagation (GA-BP) model. It could identify human blood, animal blood and non-blood substances accurately.

## Experimental

### Samples

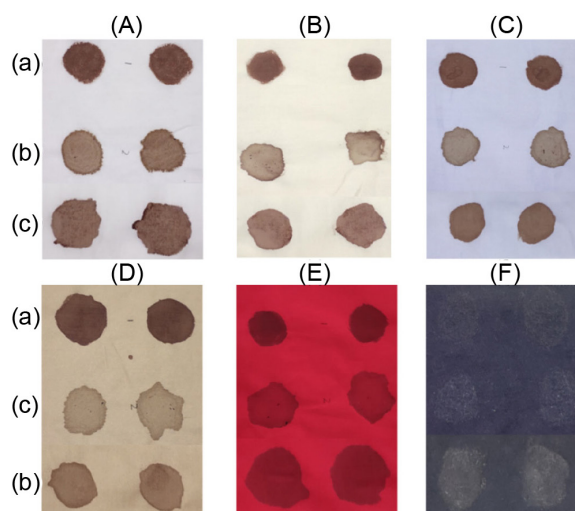
The samples included three types, namely: human blood, chicken blood and pig blood. The total number of samples was 216 and the number of each type was 72. The details of the samples were shown in Table 1. Human blood samples were collected directly from 2-man volunteers. Human blood was collected using a vacuum collection vessel with an anticoagulant. This study involves bloodstains that passed the ethics approvals by Kunming University of Science and Technology medical ethics committee, No. KMUST-MEC-204.

**Table 1.** Samples of different types of bloodstains used in the substrates

Substrate	Human blood	Chicken blood	Pig blood
White fabric	12	12	12
Red fabric	12	12	12
Blue fabric	12	12	12
Beige fabric	12	12	12
Brown fabric	12	12	12
Black fabric	12	12	12
Total	72	72	72

Fresh chicken blood and pig blood come from slaughterhouses. Human blood, chicken blood and pig blood samples were collected in lavender-capped vials with

ethylene diamine tetra acetic acid (EDTA, Kangweishi, Hebei, China) to prevent clotting. Here, 100% cotton fabrics with different colors (white, beige, blue, red, brown and black) were used as substrates. The reason was that most of the clothes were made of cotton fabric in China as they usually occurred at the crime scene. Bloodstains were applied directly on each cotton fabric. All samples were stored in the laboratory. The temperature was 26 °C and the samples were kept in the laboratory for seven days before analysis. Figure 1 shows the blood samples after the drying period and the details of the samples were shown in Figures 1A-1F for the white, beige, blue, cotton, red cotton and black cotton fabric, respectively. Figure 1a shows the human blood, Figure 1b shows the chicken blood and Figure 1c shows pig blood in each fabric.



**Figure 1.** The blood samples after the drying period: (a) human blood, (b) chicken blood and (c) pig blood. (A)-(F) were the white, beige, blue, brown, red cotton and black cotton fabric.

### Spectral acquisition

A hand-held NIR spectrometer (MicroNIR 1700) was used in the experiment. It was provided by VIAVI Solution, Milpitas, CA, United States. The MicroNIR consists of a linear variable filter (LVF) as the dispersing element. The LVF is coupled to a linear detector array (128-pixel uncooled InGaAs photodiode array). The light source is a pair of integrated vacuum tungsten lamps. A 16-bit analog-to-digital converter (ADC) is used for analog conversion. A MicroNIR hand-held NIR spectrometer provided by VIAVI company from America was used to collect the spectral data of the bloodstain samples. The wavelength range and the spectral resolution of the device were 908-1676 and 4 nm, respectively. The data sample interval was set as 6 nm, the integration time was 9.6 ms, the number of scans was 50 and the thermistor was 28.6.

A 99% diffuse reflective white board was placed under the blood sample. Absorbance spectral data of each sample were collected as shown in Figure 1 and the average of the spectral data was set as the final data. The samples were divided into training set and test set with the ratio 7:3 randomly.

## Methods and measures

### CNN-SVM algorithm

Here, a hybrid algorithm that included the proposed CNN and SVM was designed. The proposed CNN was assigned as automatic feature extractor from spectral data and SVM was employed for classifier.<sup>18</sup> In brief, the steps of the method were as follows: (i) the proposed CNN was trained on spectral data, first. (ii) To classify these features, the dataset was split into 70% training set and 30% testing set. Then, the trained net was activated. (iii) SVM classifier was employed to detect type of bloodstain, effectively.

### Measures of classification performance

As a visualization tool, confusion matrix was not only used to evaluate the accuracy for supervised learning, but also for unsupervised learning. The matrix could also display the accuracy of classification results. Figure 2 shows the basic form of the confusion matrix. In Figure 2, TP is true positive, FN is false negative, FP is false positive, and TN true negative.

Confusion matrix		Predicted value	
		Positive	Negative
True value	Positive	TP	FN
	Negative	FP	TN

**Figure 2.** Confusion matrix of the two-category task.

Here are some symbolic meanings in the confusion matrix classification assessment metric: accuracy is the proportion of the total number of predictions that were correct.

The following are metrics for mixed-matrix multi-classification performance: accuracy was the proportion of total observations for which all judgments of the classification model were correct:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

The precision rate was the ratio of the number of positive samples correctly classified to the number of all samples divided by the classifier:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

The definition of sensitivity was shown as follows:

$$\text{Sensitivity} = \text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

Specificity was the correct proportion of actual negatives measured:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (4)$$

The  $F_1$ -score ( $F_1$  score) was the harmonic mean of precision and sensitivity:

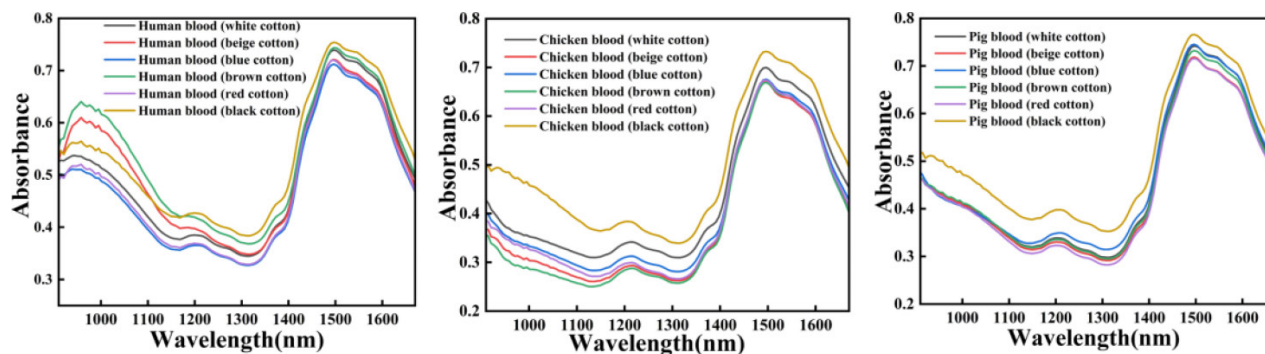
$$F_1\text{-score} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \quad (5)$$

## Results and Discussion

### Data pre-processing

Figure 3 shows the average NIR spectra data of human blood, chicken blood, and pig blood on different cotton fabrics. It showed that the spectral data of the same species of blood on different colors of cotton fabrics were different. It meant that the different colors of cotton fabrics had some influence on blood spectra data. Therefore, different colors of cotton fabrics were selected as substrates in the following research.

The raw spectral data contain both the information of the samples and the noise. The pre-processing operation can not only reduce the influence of noise, but also enhance the experimental ability of the model. Here, Savitzky-Golay + 1<sup>st</sup> derivatives (SG+D1), standard normal variate (SNV), multiplicative signal correction (MSC), and Savitzky-Golay (SG) were chosen to establish the bloodstain identification models by using SVM algorithm. The parameters of SVM algorithm were set as the kernel function was radical basis function (RBF). The core of SVM modelling was polynomial, C and gamma were hyperparameters. Penalty factor was set as 10 and gamma was set as 0.01. Accuracies of training set and test set based on SVM algorithm with different preprocessing methods, the training set and test set include human blood, chicken



**Figure 3.** Average NIR spectra of human blood, chicken blood, and pig blood on different cotton fabrics.

blood and pig blood total data were shown in Table 2. In Table 2, the accuracies of training set and test set were the highest when using the SNV pre-processing operation. Therefore, the SNV algorithm pre-processing operation was used before building the training models for each algorithm in the follow-up study. Figure 4 shows the original spectral data and the pre-processing result after the SNV operation.

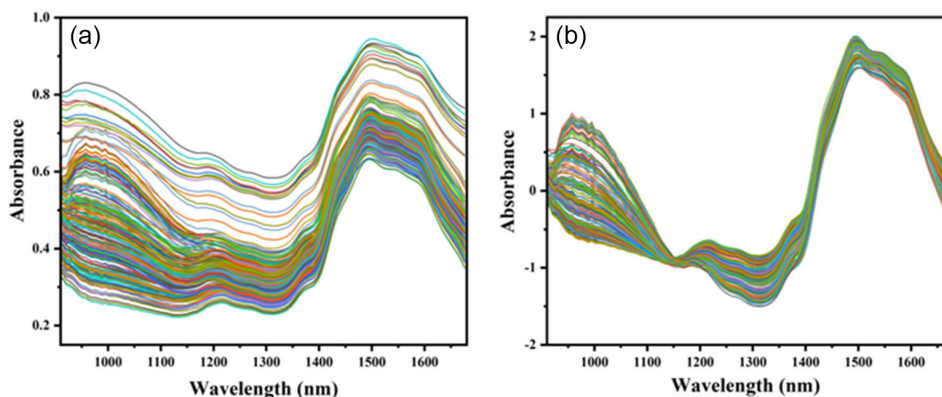
#### Construction of the qualitative model

SVM,<sup>18,19</sup> GA-BP,<sup>20</sup> RF,<sup>21</sup> CNN-SVM<sup>22</sup> algorithms were

**Table 2.** Accuracies of training set and test set based on SVM algorithm with different preprocessing methods, the training set and test set include human blood, chicken blood and pig blood total data

Algorithm	Preprocessing method	Training accuracy / %	Test accuracy / %
SVM	SG	60.19	60.18
	SG+D1	62.50	61.60
	MSC	77.78	77.70
	SNV	84.76	73.84

SVM: support vector machines; SG: Savitzky-Golay; SG+D1: Savitzky-Golay + 1<sup>st</sup> derivatives; MSC: multiplicative signal correction; SNV: standard normal variate.



**Figure 4.** (a) The original near infrared spectroscopy data and (b) after pre-processing near infrared spectroscopy data, including human blood, chicken blood and pig blood total data.

used to establish the qualitative models of different types of bloodstains, including human, chicken, and pig. The precision, accuracy, sensitivity and specificity were used as the model evaluation.

Table 3 shows the performance of the training model for different types of bloodstains using different modeling approaches. The classification precision, sensitivity, specificity and accuracy of CNN-SVM training model were much higher than those of SVM, GA-BP, and RF training models.

The prediction performance of SVM, GA-BP, RF and CNN-SVM algorithms were shown in Table 4 in the form of a confusion matrix. As shown in Table 4, the accuracy, precision, sensitivity, specificity and *F1*-score of CNN-SVM algorithm were much higher than those of SVM, GA-BP, RF algorithms. A higher sensitivity means a higher recognition capability for a classification model and a higher *F*-score means a lower misdiagnosis rate for a classification model. Besides, the average prediction accuracies of SVM, GA-BP, RF and CNN algorithms were 73.84, 84.62, 93.84 and 98.48%, respectively. The prediction accuracy of CNN-SVM algorithm was 24.64, 13.86, 4.64 higher than that of SVM, GA-BP and RF, respectively. Not only the accuracy but also precision, sensitivity and *F*-score contained in each base classifier's confusion matrix, the CNN-SVM

**Table 3.** Comparison model effects of training models for different bloodstains

Object	Algorithm	Type	Training samples	Training correct number	Number of false positives	PC / %	SN / %	SP / %	F1-score	OAC / %
Cotton fabric	SVM	human blood	46	4	42	87.5	91.3	94.3	0.89	84.76
		chicken blood	51	9	42	91.3	82.4	96.0	0.87	
		pig blood	54	10	44	77.2	81.5	86.6	0.79	
	GA-BP	human blood	53	2	51	85.5	96.2	93.9	0.90	92.05
		chicken blood	46	10	36	97.3	78.3	99.0	0.87	
		pig blood	52	0	52	96.3	100	98.0	0.98	
	RF	human blood	55	0	55	100	100	100	1.00	100
		chicken blood	49	0	49	100	100	100	1.00	
		pig blood	47	0	47	100	100	100	1.00	
	CNN-SVM	human blood	50	0	50	100	100	100	1.00	100
		chicken blood	50	0	50	100	100	100	1.00	
		pig blood	50	0	50	100	100	100	1.00	

SVM: support vector machines; GA-BP: genetic algorithm-back propagation; RF: random forest; CNN-SVM: convolutional neural networks-support vector machines; NIR: near-infrared; PC: precision; SN: sensitivity; SP: specificity; OAC: overall accuracy.

**Table 4.** Comparison model effects of test models for different bloodstains

Objects	Algorithms	Type	Test samples	Test correct number	Number of false positives	PC / %	SN / %	SP / %	F1-score	OAC / %
Cotton fabric	SVM	human blood	26	17	9	85.0	65.4	92.3	0.74	73.84
		chicken blood	21	18	3	85.7	85.7	93.2	0.86	
		pig blood	18	13	5	54.2	72.2	76.6	0.62	
	GA-BP	human blood	19	17	2	81.0	89.5	91.3	0.85	84.62
		chicken blood	26	18	8	90.0	69.2	94.9	0.78	
		pig blood	20	20	0	83.3	100	91.1	0.91	
	RF	human blood	17	16	1	94.1	94.1	97.9	0.94	93.84
		chicken blood	23	21	2	95.5	91.3	97.6	0.93	
		pig blood	24	23	1	92.3	96.0	95.0	0.94	
	CNN-SVM	human blood	22	22	0	100	100	100	1.00	98.48
		chicken blood	22	21	1	95.7	100	98.4	0.98	
		pig blood	21	20	1	100	95.5	100	0.98	

SVM: support vector machines; GA-BP: genetic algorithm-back propagation; RF: random forest; CNN-SVM: convolutional neural networks-support vector machines; NIR: near-infrared; PC: precision; SN: sensitivity; SP: specificity; OAC: overall accuracy.

algorithm occupied the highest accuracy for the calibration, and prediction. The reason was that compared with the traditional methods, the CNN-SVM algorithm has lower loss value and better generalization in modeling training.<sup>13</sup> The training of CNN usually requires large sample data and has the overfitting problem, but SVM has good generalization ability and can solve small sample problems. Moreover, the CNN-SVM combined model could automatically extract features using CNN, and better improve the generalization ability of CNN and the classification accuracy by means of combining the SVM.<sup>22</sup> In summary, the CNN-SVM algorithm could improve the robustness and accuracy of the

model. The above results showed the CNN-SVM algorithm had the best performance to building the models for the different types of bloodstains with NIR spectral data. In this paper, the work was completed with the help of MATLAB<sup>23</sup> and Origin<sup>24</sup> software.

## Conclusions

A novel method was proposed to identify different bloodstains by using a hand-held NIR spectrometer together with CNN-SVM algorithm. The results showed that the method was an alternative to discriminate the

bloodstains rapidly, accurately, and non-destructively. Besides, the CNN-SVM algorithm performed much better than traditional SVM, RF and GA-BP algorithms in dealing with the NIR spectral data. The research of blood spectral detection method proposed in this paper has great research space and application prospects.

## Acknowledgments

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 Ministry of Public Security (2020GABJC41, 2019GABJC40), Yunnan Provincial Department of Science and Technology (202001AU070004, 2018FD160), Basic Research Project of Yunnan Police College (21A028) and Yunnan Provincial Key Laboratory of Forensic Science (2020zz02, 2020zz07).

## Author Contributions

Jianqiang Zhang, Yueying Qi, Ying Liang, Qi Zeng, Jiaquan Wu were responsible for resources; Ying Liang and Yunxia Zhao for data curation and investigation; Jianqiang Zhang, Yunxia Zhao and Qifu Yang for writing original draft; Xinyu Zhang and Kun Ma for the conceptualization and writing-review.

## References

- Pereira, J. F. Q.; Silva, C. S.; Vieira, M. J. L.; Pimentel, M. F.; Braz, A.; Honorato, R. S.; *J. Microchem.* **2017**, *133*, 561. [Crossref]
- Fonseca, A. C. S.; Pereira, J. F. Q.; Honorato, R. S.; Bro, R.; Pimentel, M. F.; *Spectrochim. Acta, Part A* **2022**, *267*, 120533. [Crossref]
- Holtkötter, H.; Dias Filho, C. R.; Schwender, K.; Stadler, C.; Vennemann, M.; Pacheco, A. C.; Roca, C.; *J. Leg. Med.* **2017**, *132*, 683. [Crossref]
- Zubakov, D.; Chamier-Ciemińska, J.; Chamier-Ciemińska, I.; Maciejewska, A.; Martínez, P.; Pawłowski, R.; Haas, C.; Kayser, M.; *J. Forensic Sci. Int.: Genetics* **2018**, *36*, 112. [Crossref]
- Malegori, C.; Alladio, E.; Oliveri, P.; Manis, C.; Vincenti, M.; Garofano, P.; Barni, F.; Berti, A.; *Talanta* **2020**, *215*, 120911. [Crossref]
- Silva, C. S.; Pimentel, M. F.; Amigo, J. M.; Honorato, R. S.; Pasquini, C.; *TrAC, Trends Anal. Chem.* **2017**, *95*, 23. [Crossref]
- Virkler, K.; Lednev, I. K.; *J. Anal. Chem.* **2009**, *81*, 7773. [Crossref]
- Watanabe, A.; Furukawa, H.; Miyamoto, S.; Minagawa, H.; *Constr. Build. Mater.* **2019**, *196*, 95. [Crossref]
- Tsuchikawa, S.; Ma, T.; Inagaki, T.; *Anal. Sci.* **2022**, *38*, 635. [Crossref]
- Yu, H.; Wang, X.; Shen, F.; Long, J.; Du, W.; *Fuel* **2022**, *316*, 123101. [Crossref]
- Quan, Z.; Hua, H.; Shiping, Z.; Baohua, Q.; Xin, T.; *J. Near Infrared Spectrosc.* **2023**, *31*, 63. [Crossref]
- Jamrógiewicz, M.; *J. Pharm. Biomed. Anal.* **2012**, *66*, 1. [Crossref]
- Xiankun, S.; Lan, L.; Chengfan, L.; Jingyuan, Y.; Junjuan, Z.; Wen, S.; *J. IEEE Access.* **2019**, *7*, 164507. [Crossref]
- Ozaltin, O.; Yeniay, O.; *Soft Comput.* **2023**, *27*, 4639. [Crossref]
- Khairandish, M. O.; Jain, V.; Chatterjee, M.; Jhanjhi, N. Z.; *IRBM* **2022**, *43*, 290. [Crossref]
- Chaudhari, D. J.; Malathi, K.; *Opt. Mem. Neural Networks* **2023**, *32*, 39. [Crossref]
- Zhao, W.; Mu, T.; Li, D.; *J. Appl. Remote Sens.* **2020**, *14*, 024514. [Crossref]
- Nkengfack, L. C. D.; Tchiotso, D.; Atangana, R.; Louis-Door, V.; Wolf, D.; *J. Biomed Signal Process Control.* **2020**, *62*, 102141. [Crossref]
- Schaback, R.; *Constructive Approximation* **2005**, *21*, 293. [Crossref]
- Mekni, N.; Claudia, C.; Langier, T.; Rosa, M. D.; Perricone, U.; *Int. J. Mol. Sci.* **2021**, *22*, 7714. [Crossref]
- Asadi, S.; Roshan, S.; Kattan, M. W.; *J. Biomed. Inf.* **2021**, *115*, 103690. [Crossref]
- Wu, H.; Huang, Q.; Wang, D.; Gao, L.; *J. Electromyogr. Kinesiol.* **2018**, *42*, 136. [Crossref]
- Matlab*, version R2022a; The MathWorks Inc.; Natick, MA, USA, 2007.
- Origin*, version 2022; OriginLab, Northampton, USA, 2021.

Submitted: November 23, 2023

Published online: February 22, 2024