激光与光电子学进展
Laser & Optoelectronics Progress
ISSN 1006-4125,CN 31-1690/TN

《激光与光电子学进展》网络首发论文

题目：基于拉曼光谱和机器学习方法的一次性口罩分类识别
作者：刘金坤，李春宇，吕航，孔维刚，孙威，张格菲
收稿日期：2020-08-24
网络首发日期：2020-10-19
引用格式：刘金坤，李春宇，吕航，孔维刚，孙威，张格菲．基于拉曼光谱和机器学习方法的一次性口罩分类识别[J/OL]．激光与光电子学进展．
https://kns.cnki.net/kcms/detail/31.1690.tn.20201015.0915.004.html

网络首发时间：2020-10-19 11:21:46
网络首发地址：https://kns.cnki.net/kcms/detail/31.1690.tn.20201015.0915.004.html

基于拉曼光谱和机器学习方法的一次性口罩分类识别

刘金坤[1]，李春宇[1*]，吕航[1]，孔维刚[2]，孙威[1]，张格菲[1]

[1]中国人民公安大学侦查学院，北京 100038

[2]郑州市公安局刑事科学技术研究所，郑州 450000

摘 要 一次性口罩分类识别在法庭科学物证鉴别中具有重要意义，实验借助拉曼光谱分析技术和机器学习方法实现了口罩类别准确快速区分。实验采集来自不同地市、不同厂家生产的 37 种一次性口罩样品拉曼光谱数据，利用 S-G 平滑算法和数据归一化进行数据预处理，通过主成分分析法和拉曼光谱特征峰对照的方法划分口罩类别，进而构建基于 SVM、贝叶斯判别分析、BP 神经网络算法的一次性口罩分类识别模型。结果表明，SVM 模型训练集的准确率为 93.3%，测试集准确率为 100%，贝叶斯判别分析的训练集和测试集准确率均为 100%，BP 神经网络的训练集准确率为 93.9%，验证集准确率为 60%，测试集准确率为 60%，根据分类要求的严谨性和准确性，最终确定贝叶斯判别分析模型为口罩分类识别的最佳模型。实验仪器的选择和模型设计思路可为法庭科学技术物证检验提供借鉴。

关键词 一次性口罩；拉曼光谱；机器学习；分类识别

中图分类号 0657　　**文献标志码** A

Classification and recognition of disposable masks based on Raman spectroscopy and machine learning

Liu Jinkun[1], Li Chunyu[1*], Lü Hang[1], Kong Weigang[2], Sun Wei[1], Zhang Gefei[1]

1. School of Investigation，People's Public Security University of China, Beijing 100038, China;

2. Institute of Criminal Science and Technology, Zhengzhou Public Security Bureau 450000, China

Abstract: Classification and recognition of disposable masks is of great significance in forensic science. Raman spectroscopy and machine learning are used to distinguish the types of disposable masks accurately and quickly. The Raman spectrum data of 37 disposable masks from different cities and factories were collected. The S-G smoothing algorithm and data normalization were used to preprocess the data. The masks were classified by principal component analysis (PCA) and Raman spectrum characteristic peak comparison. SVM, Bayesian discriminant analysis and BP

基金项目：国家重点研发计划（2017YFC0822004），国家重点研发计划（2019YFF0303405），公安部技术研究计划（2019JSYJC21），中央高校基本科研业务费项目（2019JKF427）

E-mail：18860392253@qq.com；***E-mail：**lichunyu@ppsuc.edu.cn

neural network models were constructed. The results show that the accuracy of SVM model training set is 93.3%, the accuracy of test set is 100%, the accuracy of training set and test set of Bayesian discriminant analysis is 100%, the accuracy of BP neural network training set is 93.9%, the accuracy of verification set is 60%, and the accuracy of test set is 60%. According to the preciseness and accuracy of classification requirements, Bayesian discriminant analysis model is finally determined as mask classification. The selection of experimental instruments and the design of the model can provide reference for forensic science and technology evidence examination.

Key words disposable mask; Raman spectroscopy; machine learning; classification and recognition

OCIS codes 300.6170; 300.6340; 300.6450

1 引言

一次性防护型口罩作为重要的物证经常出现在蓄谋盗窃、杀人、爆炸等犯罪现场，特别是在新冠肺炎疫情暴发期间，口罩的使用量大幅度提升，增加了口罩在犯罪现场出现的可能性，口罩物证研究的重要性日益凸显。法庭科学中，侦查人员可以对故意佩戴口罩遮挡脸部实施作案行为的犯罪嫌疑人进行视频追踪，也可以对遗留在犯罪现场的口罩物证进行人体脱落细胞、携带灰尘、材质等多方面检验。其中，口罩材质的分类识别可以确定嫌疑人购买口罩的品牌或生产口罩的厂家，从而追根溯源，确定犯罪嫌疑人可能出现的位置，缩小侦查范围，提高办案效率。一次性防护型口罩通常有三层，内外两层为非织造布(化学合成纤维)，中间一层为熔喷布，口罩带为绒毛橡筋，鼻梁条为可塑性材料。《日常防护型口罩技术规范GBT32610-2016》从口罩的耐摩擦色牢度、甲醛含量、PH 值、可分解致癌芳香胺染料、环氧乙烷残留量、吸气阻力等方面提出质量检测标准。《丙纶纺粘/熔喷/纺粘复合无纺布标准》从单位面积质量、断裂强力和断裂伸长率、抗渗水性、透气性方面提出口罩质量检测方案。但这些方法的实施需要一定的时间周期，所耗费人力、财力较大，对于法庭科学领域口罩物证来源的确定未能显现有效作用，因此需要探索快速高效的口罩材质分类识别方法才能解决公安机关面临的实际问题。

非织造布作为一次性防护型口罩的主要原料，其化学纤维材质主要有丙纶、涤纶、腈纶和氨纶组成。对纤维材质的鉴定与识别已有多种方法，FZ/T01057《纺织纤维鉴别实验方法》将化学溶解、显微镜观察、燃烧等方法作为纤维鉴定标准。该标准的鉴定结果准确，但损耗检材，化学物质会产生污染，不符合环境友好型社会发展理念。另外，红外光谱法是纤维检验的常用方法，张海煊[1]等用红外光谱仪测得锦纶、氨纶的光谱数据，通过建立偏最小二乘法模型实现纤维快速定量分析。黎海洋[2]等用傅里叶变换衰减全反射红外光谱技术对聚烯烃

弹性纤维、聚酯类弹性纤维、二烯类弹性纤维和氨纶进行检验，通过分析其红外特征峰对纤维分类。近年来，特种光谱成像技术应用于纤维物证的检验也在慢慢兴起，魏子涵[3]等用 FTIR 光谱测量出不同种类的织物纤维的标准谱图，实现了不同纤维的快速无损鉴别。金肖客[4]等用高光谱成像系统鉴别不同种类纺织品，证明了高光谱纤维检验的可行性。

然而，科技的进步需要实验仪器、学科领域、数据分析方法等多方面不断创新，拉曼光谱法作为重要的物质分子结构分析方法可与红外光谱法相互弥补，已应用在化学新材料鉴别、宝石鉴别、气体鉴别、微量物证鉴别等多个领域[5]，因其无损检验，获取谱图速度快，数据稳定性好，也成为化学纤维检验的最佳方法之一。同时，将机器学习建模方法应用在纤维光谱数据分类识别上可增加实验结果预测的准确性和方便快捷性。因此，本文收集到 37 种不同品牌和种类的一次性防护型口罩（部分样品见表 1），通过采集其拉曼光谱数据，利用主成分分析法降维，并建立 SVM、BP 神经网络、贝叶斯判别分析多种模型，以期找到最佳模型，达到对未知类别口罩精确分类和模式识别的目的。

表 1 一次性防护型口罩样品表

Table 1 sample table of disposable protective masks

Label	Brand	Manufactor
1	XuanChen mask	Rizhao nuohuan protective equipment Co., Ltd
2	HuiAn mask	Zhangjiagang meibaiqi Trade Co., Ltd
3	ShengWang mask	Xiantao Siqi protective equipment Co., Ltd
4	JiaBoNeng mask	Guangdong Dongwan Yian labor protection products Co., Ltd
…	…	…
33	Hui An mask	Suzhou Lotte protective equipment Co., Ltd
34	Zhi Chang mask	Nanchang of Jiangxi Province
35	San Bang mask	Foshan Nanhai Weijian sanbang protective equipment Technology Co., Ltd
36	Tai Bang mask	Yunnan Baiyao Group Co., Ltd
37	3M mask	3M China Ltd

2 实验

2.1 实验仪器

显微共聚焦拉曼光谱仪(inVia Raman Microscop)，英国雷尼绍(Renishaw)公司生产，配备 532 nm、633 nm、785 nm 波段激光器，显微镜倍率 5X、20X、50X、100X，光谱扫描范围 100～2000 nm，光谱分辨率 1 cm^{-1}，最低波数 10 cm^{-1}。

2.2 光谱采集

取口罩外层非织造布，裁剪成 0.5 cm × 0.5 cm 小方块，镊子夹取，放置在载玻片上，用双面粘胶带固定，酒精擦拭样品表面，除去污渍、灰尘等，避免杂物干扰。随机选取20#样品做激光器优选实验，设置最大功率 10%，曝光时间 10 s，调整显微镜聚焦倍率，发现在 100X放大焦距下可以获得单根清晰的口罩纤维（见图 1），光谱实验结果（见图 2），由图2 可知，拉曼位移 100～2000 cm^{-1} 范围内 532 nm 激光器获得光谱噪声大，光谱峰值不明显，没有特征峰。633 nm 激光器光谱信号弱，且在拉曼位移 100～800 cm^{-1} 范围内没有信号出现。785 nm激光器得到的谱图特征峰峰值明显，强度大，噪音少，可作为实验最佳激光器。

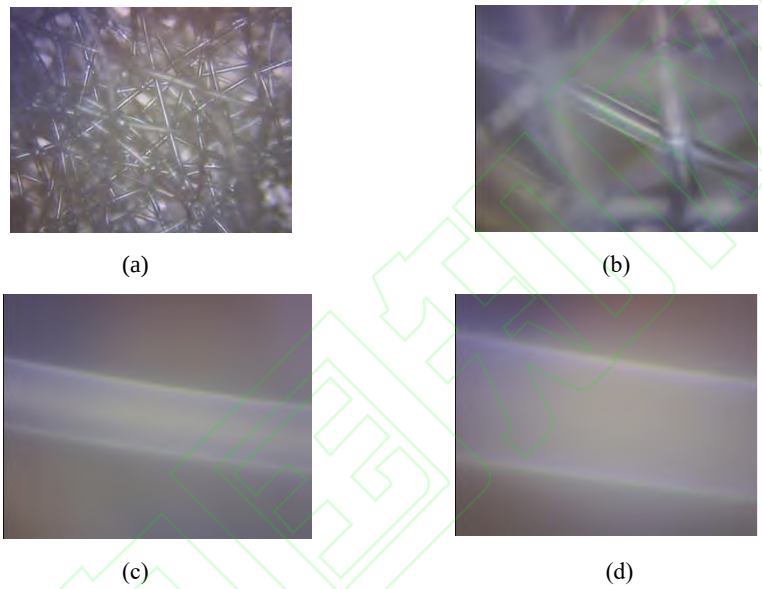


图 1 不同焦距下的口罩外观形态

(a):5x; (b)20x; (c)50x; (d)100x

Fig. 1 The appearance of mask under different focal length

(a):5x; (b)20x; (c)50x; (d)100x

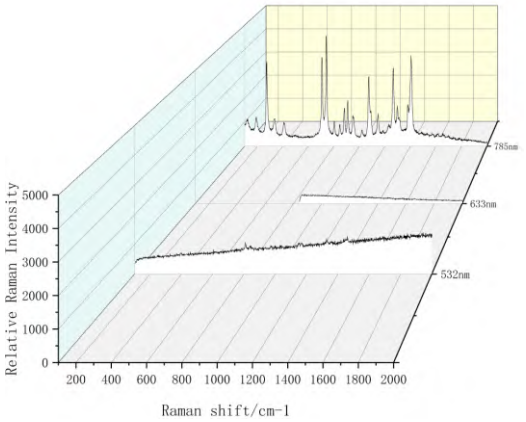
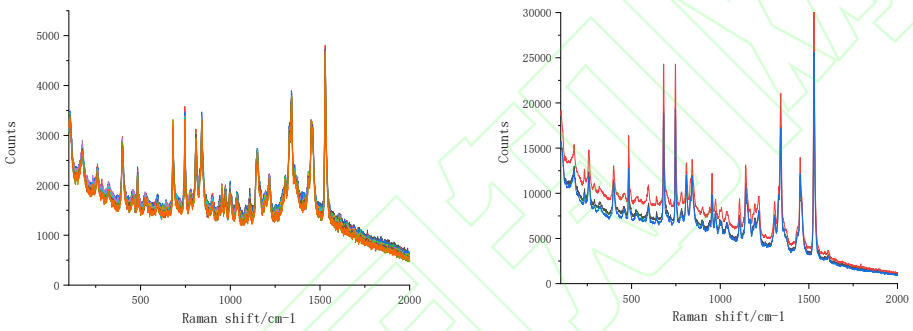


图 2 不同波长激光器的口罩测量结果

Fig. 2 Results of mask measurement for different wavelength lasers

为验证拉曼光谱仪的稳定性和光谱重现性，随机选取 10#样品测量 10 次（见图 3a），由图 3a 可知，10 次实验光谱除了产生较小的噪声差异，特征峰峰位全部重合。为验证不同采样点对实验结果的影响，随机选取 18#样品，在 3 个不同的采样点上验证口罩材料的均匀性（见图 3b）。由图 3b 可知，不同采样点间除噪点、相对拉曼强度稍有差别，特征峰都相同。以上实验结果表明：785 nm 激光器下测得拉曼光谱稳定性好，样品材质均匀性，在不同位点实验效果无明显差异，符合样品实验条件。其他样品测量时，为保证光谱数据的准确性，每个样本测量三次，取平均值，对所有样品实验后采集到 37 组拉曼图谱。



(a) (b)

图 3 口罩样品重现性实验和均匀性实验结果

Fig. 3 Repeatability test and uniformity test results of mask sample

2.3 光谱预处理

实验过程中，除光谱特征峰外，容易产生噪声，主要是受仪器性能稳定性、宇宙射线、外界光照及环境温度等因素影响。光谱预处理是消除其影响的有效方法，主要包括平滑、MSC、SNV、导数。其中，平滑可以除去光谱白噪声、荧光强度不均，提高信噪比，可分为窗口平滑和 S-G 平滑，因 S-G 平滑是窗口平滑的改进并被大量文献所引用，本文中对口罩光谱数据 S-G 平滑的主要公式为：

$$\begin{cases} x_{k,sm} = \dfrac{1}{H}\sum_{i=-m}^{m} x_{k+i}h_i \\ H = \sum_{i=-m}^{m} h_i \end{cases} \tag{2-1}$$

(2-1) 式中 x_k 为输入口罩光谱数据，$x_{k,sm}$ 为输出数据，k 为光谱序号，sm 为平滑的英文缩写，m 表示滤波带宽，h_i 为平滑系数。

实验发现，不同样品在相同激光器功率下得到的拉曼光谱相对强度不同，数量级之间相差较大，为实现更好的机器学习效果，将拉曼相对强度归一化在[0,1]之间，主要公式为：

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{x_{\max} - x_{\min}} \qquad (2\text{-}2)$$

(2-2) 式中，$x_{\min}$ 和 $x_{\max}$ 分别是原始口罩光谱数据的最小值和最大值，y 是映射的范围参数。当 $y_{\min} = 0, y_{\max} = 1$ 时，可实现 (2-2) 式[0,1]区间的归一化。设置滤波带宽 m =5，得到预处理后的结果（见图 4），由图 4 可知，在设定拉曼位移下每组光谱数据的特征峰明显，谱图稳定清晰，可进行下一步数据分析。

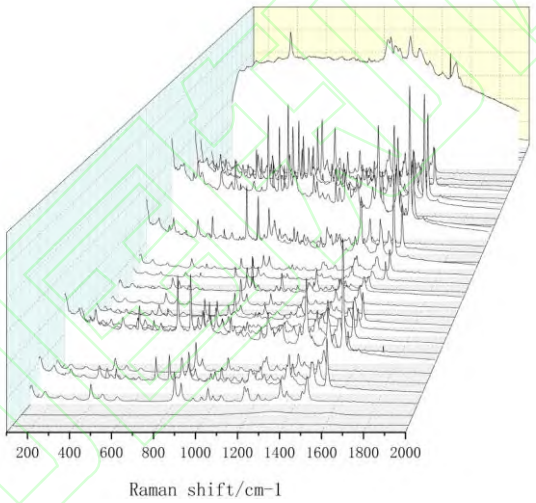


图 4 37 种口罩样品拉曼光谱处理结果

Fig. 4 Results of Raman spectra of 37 kinds of masks

2.4 模型原理

光谱数据维度较高，以本实验为例，拉曼光谱扫描范围从 100～2000 cm^{-1}，将会产生 1900维数据，高维数据中的噪点、奇异点等信息会影响模型的运行时间和预测精度，如果能降低维度，同时保证重要信息的不丢失并提取光谱特征信息，将会是非常好的办法。在多元统计分析中，特征提取的方法主要有小波变换和主成分分析，通过查阅相关参考文献[7-8]，主成分分析用于光谱数据降维效果很好，其主要步骤为：

(1) 计算口罩光谱数据标准化矩阵 X

$$X=\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ x_{p1} & x_{p2} & \cdots & x_{pq} \end{bmatrix} \quad (2-3)$$

x_{pq}：第 p 条口罩样本数据的第 q 维光谱数据

(2)计算相关系数矩阵 P

$$P=\begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1q} \\ r_{21} & r_{22} & \cdots & r_{2q} \\ \cdots & \cdots & \cdots & \cdots \\ r_{p1} & r_{p2} & \cdots & r_{pq} \end{bmatrix} \quad (2-4)$$

r_{pq}：x_i、x_j 的相关系数，$r_{ij}=r_{ji}$，r_{ij} 的推导公式如下：

$$r_{ij}=\sum_{k=1}^{n}\frac{(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{n}x_{kj}-x_i\sum_{k=1}^{n}(x_{kj}-\bar{x}_j)}} \quad (2-5)$$

(3)根据 $|\lambda\alpha-P|=0$ 求出矩阵 P 的特征值 λ_i 和特征向量 β_i 并由大到小排序，最大的特征值和特征向量就是第一主成分的方差和方向，定义每一个特征值在总方差中所占比值为主成分贡献率 PC_i，前 i 个贡献率的综合为累计贡献率 ACC_{Pc_i}，表达式如下：

$$PC_i=\frac{\lambda_i}{\sum_{k=1}^{n}\lambda_k}(i=1,2,...,n),\quad ACC_{Pc_i}=\frac{\sum_{k=1}^{i}\lambda_k}{\sum_{k=1}^{n}\lambda_k}(i=1,2,...,n) \quad (2-6)$$

(4)由 $(2-6)$ 可知特征向量组成的矩阵 $M=[\beta_1,\beta_2,...,\beta_i]$，故主成分矩阵 Y 的计算公式如下：

$$Y=XM \quad (2-7)$$

支持向量机模型(SVM)是一种降低数据结构风险的分类模型，需寻求一个超平面，距离超平面最近的点组成支持向量，通过不断优化支持向量与超平面之间的距离使其尽可能最远，从而实现不同样本分类。影响模型的主要参数有误差惩罚参数 C，可控制错误分类样本数，权衡错分样本比例和算法复杂度。对于非线性问题，还需要选择不同种类的核函数将数据投影到合适的特征空间，核函数通常分为多项式核函数、高斯核函数和径向基核函数。其

中，径向基核函数使用广泛，其重要参数γ可判定特征空间中向量间的距离，对分类准确率的影响很大[9-11]。

贝叶斯判别分析模型是依据贝叶斯定理和特征条件假设对样本进行分类的算法模型，具有降低错分概率风险的优点，模型的实现需要建立关于数据集和对应标签的联合概率分布（先验概率），基于先验概率分布的认识，得到后验分布，当有新的样本数据时，利用后验分布判断其所属类别[12-13]。

BP 神经网络是将每一个样品数据作为输入神经元，隐藏层数据向量与模型自动生成的权重矩阵做内积并加上可以降低误差的偏置项，通过激活函数输出分类概率，根据概率确定口罩的所属类别，如果输入未知测试样品数据，训练好的模型就可以识别出样品对应的标签。因为权重矩阵是随机生成的，预测结果不准确，会产生损失，因此需要计算正向传播的代价函数，通过反向传播的梯度下降算法，优化权重值，降低模型产生的损失，达到准确分类的目的[14-17]。影响模型分类结果的参数有学习率（梯度下降的步长）、隐藏层神经元的个数、激活函数、损失函数以及模型迭代次数。

3 结果与讨论

实验首先通过主成分分析法对光谱数据降维，并结合拉曼光谱特征峰确定口罩样品的类别，从而将无监督问题转化为有监督问题，接着搭建多种机器学习模型[6]，通过调参、改变实验条件，多次迭代等方法训练模型并对测试集预测，观察模型分类识别准确率和运行时间，优选出最佳口罩分类识别方案，具体流程(见图 5).

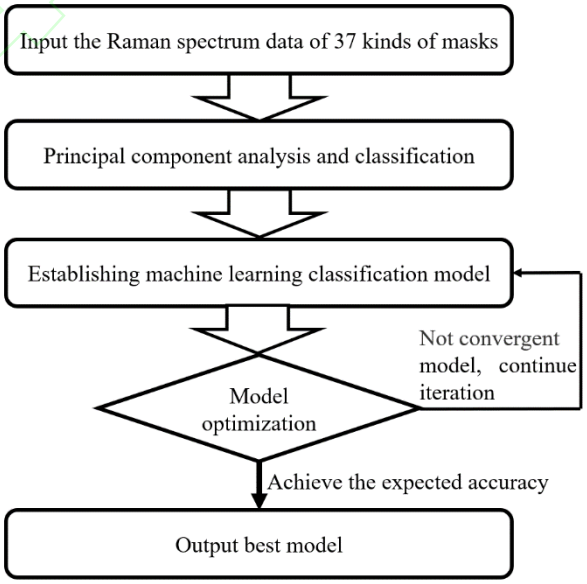


图 5 口罩拉曼光谱数据分类识别模型设计

Fig. 5 Design of mask Raman spectrum data classification and recognition model

实验发现，当提取前 31 个主成分时，累计方差百分比已达 100%（见表 2），其中成分 1 的贡献率为 49.02%，成分 2 的贡献率为 36.13%，对于总体数据可解释性较强，故提取前两个主成分代表的函数建立二维可视化图（见图 6），由图 6 可知，37 个口罩样品大致可分为 6 类，为验证其分类结果的可靠性，需对比拉曼光谱的特征峰进一步分析（见图 7）。

表 2 PCA 特征方差贡献率

Table 2 Contribution rate of PCA characteristic variance

PC.	Eigenvalue	Variance (%)	Cumulative Variance (%)
1	18.80335	49.02%	49.02%
2	13.85901	36.13%	85.15%
3	2.36197	6.16%	91.31%
4	1.76937	4.61%	95.92%
5	1.0663	2.78%	98.70%
6	0.13423	0.35%	99.05%
7	0.09534	0.25%	99.30%
8	0.06265	0.16%	99.46%
9	0.03321	0.09%	99.55%
10	0.02942	0.08%	99.63%
11	0.024	0.06%	99.69%
12	0.02108	0.05%	99.74%
13	0.01752	0.05%	99.79%
14	0.01311	0.03%	99.82%
15	0.01148	0.03%	99.85%
16	0.01001	0.03%	99.88%
17	0.008	0.02%	99.90%
18	0.00729	0.02%	99.92%
19	0.00525	0.01%	99.93%
20	0.00471	0.01%	99.95%
21	0.00387	0.01%	99.96%
22	0.00341	0.01%	99.97%
23	0.0025	0.01%	99.97%
24	0.00202	0.01%	99.98%
25	0.00168	0.00%	99.98%
26	0.00142	0.00%	99.99%
27	0.00116	0.00%	99.99%
28	8.98E-04	0.00%	99.99%
29	8.26E-04	0.00%	99.99%

30	6.32E-04	0.00%	99.99%
31	6.00E-04	0.00%	100.00%

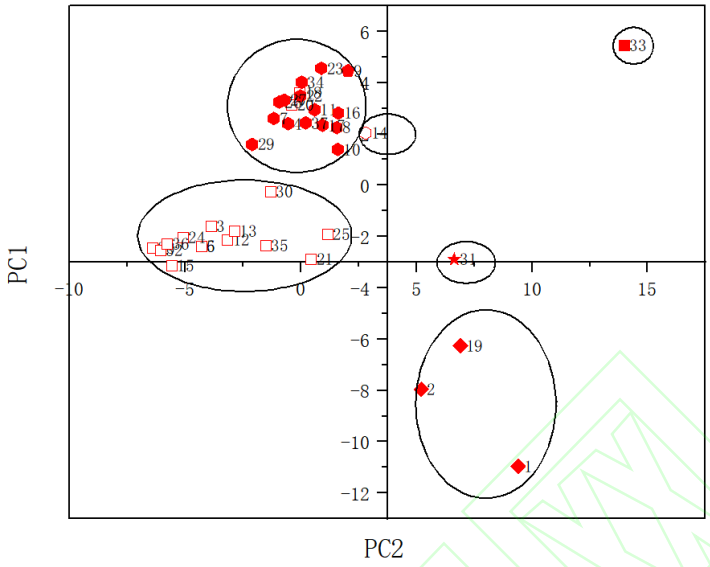
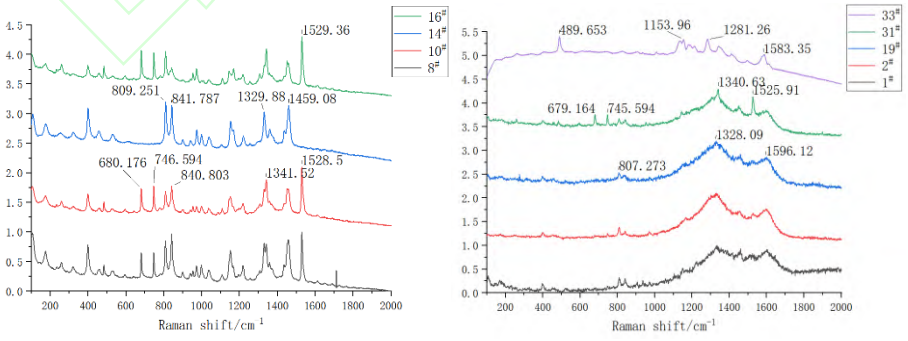


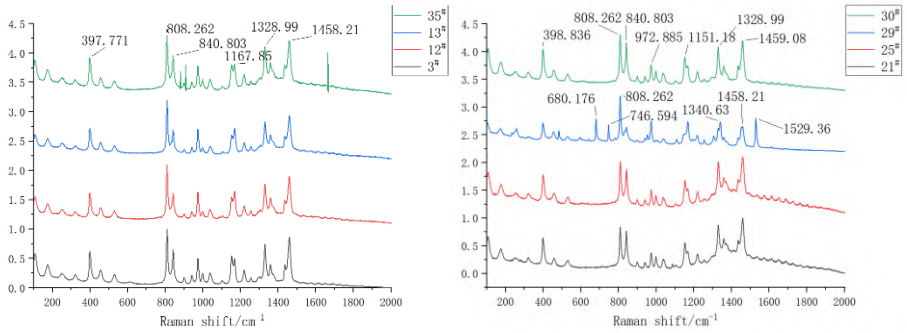
图 6 主成分分析可视化展示

Fig. 6 Visual display of Principal Component Analysis

由图 7a 可知，14#样本在 1329 cm^{-1}、1459 cm^{-1} 附近的特征峰与 8#、10#、16#不同，可单独归为一类；图 7b 中，33#样品在 489 cm^{-1}、1153 cm^{-1}、1281 cm^{-1}、1583 cm^{-1} 附近特征峰与其他样品不同，31#样品在 679 cm^{-1}、745 cm^{-1}、1525 cm^{-1} 附近特征峰与其他样品不同，1#、2#、19#特征峰相同，可归为不同类别。图 7c 中，3#、12#、13#、35#样品在 397 cm^{-1}、808 cm^{-1}、1328 cm^{-1}、1458 cm^{-1} 附近特征峰相同，可归为一类；图 7d 中，29#样本在 680 cm^{-1}、746 cm^{-1}、1529 cm^{-1} 附近特征峰与 21#、25#、30#明显不同，可归为一类，对比图 6 可知，将口罩样品分成 6 类准确可靠。



(a) (b)



(c) (d)

图 7 口罩拉曼光谱特征峰比对分析

Fig. 7 Comparison and analysis of characteristic peaks of Raman spectrum of mask

3.2 机器学习方法建模

为实现口罩样品的自动分类识别，输入未知光谱数据就能判别其种类，将 37 种口罩样品对应唯一类别标签（见表 3）并建立机器学习分类模型。

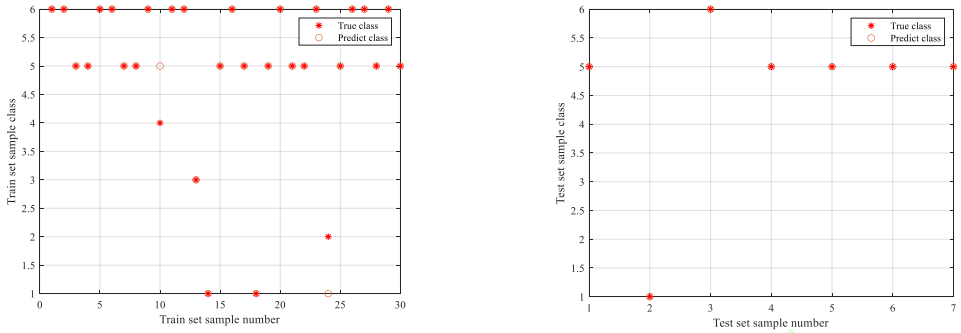
表 3 基于 PCA 和拉曼特征光谱的口罩分类

Table 3 classification of masks based on PCA and Raman spectra

Label	Sample number
1	1#、2#、19#
2	31#
3	33#
4	14#
5	4#、7#、8#、9#、10#、11#、16#、17#、18#、20#、22#、23#、26#、27#、29#、34#、37#
6	3#、5#、6#、12#、13#、15#、21#、24#、25#、28#、30#、32#、35#、36#

对于支持向量机模型，实验设置初始值 C =1、γ =0.1，终止模型运行阈值 10^{-4}，采用交叉验证法寻找最佳参数 C 为 1.5157，γ 为 0.0078，37 个口罩样品随机划分为两部分，30 个样本训练模型，7 个样本用来预测分类标签，模型训练和测试结果（见图 8），由图 8a 可知，训练集 30 个样品有两个预测错误，训练准确率 93.3%，图 8b 中测试集 7 个样品预测正确，测试准确率 100%，模型所用时间 20 s。为验证不同核函数对 SVM 模型预测准确率的影响，分别选用线性核函数、多项式核函数、RBF 核函数和 sigmoid 核函数，得到对比结果（见表 4），由表 4 可知，不同核函数的运行时间和精确度各不相同，其中，选用 RBF

核函数得到的运行时间较长，为 20 s，但其训练集和测试集准确率高于其他核函数，从而验证了选用 RBF 核函数的有效性。



(a) (b)

图 8 SVM 训练集和测试集预测结果

Fig. 8 Prediction results of SVM training set and test set

表 4 SVM 模型不同核函数对比结果

Table 4 Comparison results of different kernel functions of SVM model

Kernel function type	Training set accuracy(%)	Test set accuracy(%)	Model run time(s)
Linear	50.00	28.57	5
Polynomial	46.67	42.85	10
RBF	93.30	100	20
Sigmoid	96.67	85.00	15

对于贝叶斯判别分析模型，实验用 30 个口罩样品作为先验认识并得到判别函数，用判别函数验证剩下 7 个口罩所属类别，分析结果(见表 5).

表5 口罩样品贝叶斯判别函数汇总

Table 5 Summary of Bayes discriminant function for mask samples

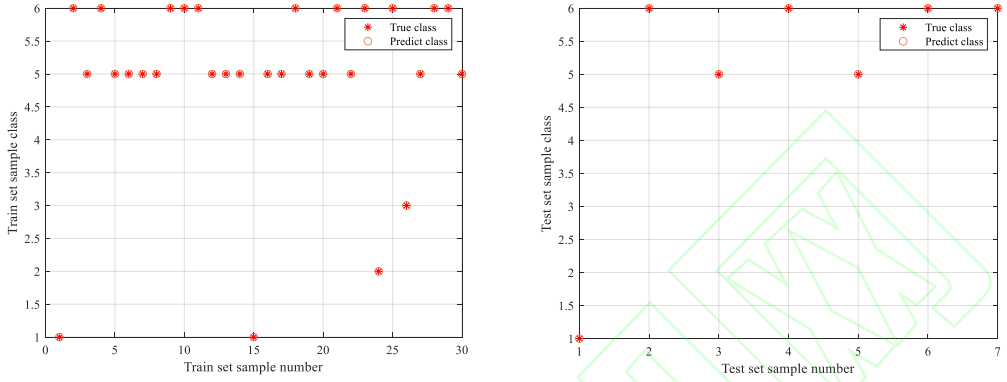
Function	Eigenvalue	Variance (%)	Cumulative variance(%)	Canonical correlation	Function test	Wilks' Lambda	Sig.
1	15064.487	66.2	66.2	1	1 though 5	0.00	0.00
2	6222.540	27.4	93.6	1	2 though 5	0.00	0.00
3	1303.951	5.7	99.3	1	3 though 5	0.00	0.00
4	126.670	0.6	99.9	0.996	4 though 5	0.00	0.00
5	24.217	0.1	100	0.98	5	0.04	0.00

由表 5 可知，贝叶斯判别函数 1 的特征值为 15064.487，方差百分比 66.2%，函数 2 的特征值为 62222.540，方差百分比 27.4%，两者累计方差百分比达 93.6%，正相关性。同时对 5 个判别函数进行显著性检验，函数 1、2 的 Sig 值远小于 0.05，具有统计学意义。根据以上分析可确定前两个判别函数作为未知样本类别的预测函数，函数表达公式如下：

$$W_1(x) = 27.712x_1 - 19.636x_2 + \ldots - 0.461x_{31} \qquad (3-5)$$

$$W_2(x) = -13.792x_1 - 7.105x_2 + \ldots + 0.212x_{31} \qquad (3-6)$$

其中，$x_i(i=1,2,\ldots,31)$ 表示第 x 个口罩样品的 i 维指标。用建立的判别函数式对口罩样品训练和分类预测（见图 9），由图 9a 可知，30 个训练样品分类准确率为 100%，图 9b 中 7 个测试样本的分类准确率为 100%，模型运行时间 10 s。



(a) (b)

图 9 贝叶斯判别分析训练集和测试集预测结果

Fig. 9 Prediction results of training set and test set of Bayesian discriminant analysis

对于 BP 神经网络模型，建模时随机选择 20 个样品数据作为训练集，10 样品为验证集，7 个样品作为测试集，设置学习率 0.001，8 个隐藏层神经元，选定 sigmoid 为激活函数，交叉熵为损失函数，模型迭代次数 15 次收敛，得到交叉熵损失函数变化图（见图 10），由图 10 可知，验证集和测试集的交叉熵趋于平缓，训练集的交叉熵处于下降趋势，说明模型在测试集上趋于稳定，未出现过拟合现象。基于此，可得到最优测试集交叉熵值为 0.060186，并根据混淆矩阵可知，模型训练集准确率 93.9%，验证集准确率 60%，测试集准确率 60%，模型运行时间 5 s。

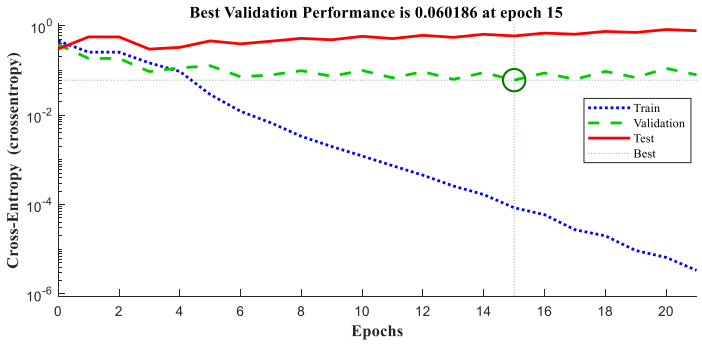


图 10 迭代次数与交叉熵的变化关系

Fig. 10 The relationship between iteration times and cross entropy

通过准确率对比三个分类模型发现：贝叶斯判别模型的训练和测试准确率都达到了100%，可作为口罩分类识别的最佳模型；SVM 模型的训练准确率为 93.3%，测试准确率达到100%，在有一定偏差容忍度的情况下可以选择；BP 神经网络的训练、验证和测试准确率都比较低，可能原因是神经网络适合多维大数据的分析，而文章所用口罩样品数据少，不能满足其权重更新优化的需要，而贝叶斯判别模型和 SVM 模型适合小数据集分类识别，如果数据量过大，训练速度会很慢，但更适合本课题研究要求。

4 结论

本文以口罩样品的拉曼光谱为基础，构建了基于特征提取和多模型结合优选的方案对口罩种类进行识别，并最终确定了基于贝叶斯判别分析的分类模型，模型训练集和测试集分类准确率都达到 100%且模型运行时间 10 s，符合快速、准确的实验要求。实验的设计方案可为犯罪现场提取的口罩物证种类识别提供借鉴，丰富法庭科学理化检验方法。但值得思考的是本文口罩样品的种类仅处于研究阶段，若想实现任意口罩种类识别还需增加样品种类，构建样品数据库，接下来将丰富实验样品种类和数量，构建更具有实战性意义的分类模型。

参考文献

[1] Zhang H X, Cao H H, Zhang X , *et al*. Near Infrared Quantitative Analysis of Fiber Contents of Polyamide/Spandex Knitted Underwear Fabric [J].Journal of Textile Research,2020(6):82-85.
张海煊,曹海辉,张续,等.锦氨内衣纤维含量近红外光谱法快速定量分析[J].针织工业,2020(6):82-85.

[2] Li H Y, Liu W, Cheng X Q, *et al*. Identification of elastic fibers by infrared spectroscopy, Raman spectroscopy, and pyrolysis gas chromatography-mass spectrometry[J]. Textile Testing and Standard,2020,6(1):13-16.
黎海洋,刘旺,程鑫桥,等.红外、拉曼光谱和裂解气-质联用技术鉴别弹性纤维[J].纺织检测与标准,2020,6(1):13-16.

[3] Wei Z H, Li W X, Du Y J, *et al*. Establishment and application of fabrics attenuated total reflection Fourier transform infrared spectroscopy spectrum library[J]. Journal of Textile Research,2019,40(8):64-68.
魏子涵,李文霞,杜宇君,等.织物傅里叶变换衰减全反射红外光谱库的建立及应用[J].纺织学报,2019,40(8):64-68.

[4] Jin X K, Tian W, Zhu Y J, *et al*. Qualitative identification of textile chemical composition based on hyperspectral imaging system[J]. Journal of Textile Research,2018,39(10):50-57.
金肖克,田伟,朱炜婧,等.基于高光谱成像系统的纺织品成分定性鉴别[J].纺织学

报,2018,39(10):50-57.

[5] Hu W, Ye S, Zhang Y, *et al*. Machine Learning Protocol for Surface-Enhanced Raman Spectroscopy[J].The journal of physical chemistry letters,2019,10(20):6026-6031.

[6] Liu J X, Du B, Deng Y Q, *et al*. Terahertz-Spectral Identification of Organic Compounds Based on Differential PCA-SVM Method[J]. Chinese Journal of Lasers, 2019,46(6): 0614039.
刘俊秀,杜彬,邓玉强,等.基于差分-主成分分析-支持向量机的有机化合物太赫兹吸收光谱识别方法[J].中国激光,2019,46(6):0614039.

[7] Li G W, Gao X H, Xiao N W, *et al*. Estimation of Soil Organic Matter Content Based on Characteristic Variable Selection and Regression Methods[J]. Acta Optica Sinica,2019,39(9):0930002.
李冠稳,高小红,肖能文,等.特征变量选择和回归方法相结合的土壤有机质含量估算[J].光学学报,2019,39(9):0930002.

[8] Xu H D, Lin L L, Li Z, *et al*. Nephrite Origin Identification Based on Raman Spectroscopy and Pattern Recognition Algorithms[J]. Acta Optica Sinica,2019,39(3):0330001.
徐荟迪,林露璐,李征,等.基于拉曼光谱和模式识别算法的软玉产地鉴别[J].光学学报,2019,39(3):0330001.

[9] Study on the identification of X-ray fluorescent spectral paper ashes based on Support Vector Machine algorithm Laser & Optoelectronics Progress

[10] Pedro S S, Ana C, Ana S A, *et al*. Identification of rice flour types with near-infrared spectroscopy associated with PLS-DA and SVM methods[J]. European Food Research and Technology,2020,246(3):527-537.

[11] Hu X, Wu R M, Zhu X Y, *et al*. Fast Detection of Chlorpyrifos Residues in Tea via Surface-Enhanced Ra man Spectroscopy Combined with Two-Dimensional Correlation Spectroscopy[J]. Acta Optica Sinica,2019,39(7):0730001.
胡潇,吴瑞梅,朱晓宇，等.表面增强拉曼光谱结合二维相关谱快速检测茶叶中的毒死蜱残留[J].光学学报,2019,39(7):0730001.

[12] He X L, Wang J F, He Y, *et al*. Infrared spectroscopy identification of plastic steel windows based on Bayes discrimination analysis[J]. Laser Journal,2019,40(11):33-37.
何欣龙,王继芬,何亚,等.Bayes判别的塑钢窗红外光谱快速识别[J].激光杂志,2019,40(11):33-37.

[13] Wen C P, Bai Y Y, Zeng J J, *et al*. Bayes discriminant analysis of natural grassland classification[J]. Chinese Journal of Grassland,2016,38(3):50-55.
文畅平,白银涌,曾娟娟,等.天然草地分类的 Bayes 判别分析法[J].中国草地学报,2016,38(3):50-55.

[14] Zhou S, Shen C Y, Zhang L, *et al*.Dual-optimized adaptive Kalman filtering algorithm based on BP neural network and variance compensation for laser absorption spectroscopy.[J]. Optics express,2019,27(22):874-888.

[15] Ershat A, Baidengsha M, Mamat S, *et al*. Combined Estimation of Chlorophyll Content in Cotton Canopy Based on Hyperspectral Parameters and Back Propagation Neural Network[J]. Acta Optica Sinica,2019,39(9):0930003.
依尔夏提·阿不来提,白灯莎·买买提艾力,买买提·沙吾提,等.基于高光谱和BP神经网络的棉花冠层叶绿素含量联合估算[J].光学学报,2019,39(9):0930003.

[16] Song H S, Ma L Z, Wang Y F, *et al*. Recognition of Formaldehyde, Methanol Based on PCA-BP Neural Network[J]. Laser & Optoelectronics Progress,2020, 57(7):071201.
宋海声,麻林召,王一帆,等.基于 PCA-BP 神经网络对甲醛和甲醇的识别研究[J].激光与光

电子学进展,2020, 57(7):071201.

[17] Cai Y, Su M X, Cai X S. Method for Mixed-Particle Classification Based on Convolutional Neural Network[J]. Acta Optica Sinica,2019,39(7):0712002.
蔡杨,苏明旭,蔡小舒.基于卷积神经网络的混合颗粒分类法研究[J].光学学报,2019,39(7):0712002.