

# SERS 协同机器学习在生物医药检测中的应用

陈云帆<sup>1</sup>, 王也晨<sup>1</sup>, 王龙<sup>2\*</sup>, 安琪<sup>1\*</sup>, 冯泽国<sup>2</sup>

(1.中国地质大学(北京)材料科学与工程学院,北京 100083; 2.中国人民解放军总医院第一医学中心疼痛科,北京 100853)

**摘要:**表面增强拉曼光谱(surface-enhanced Raman spectroscopy, SERS)能够捕获分子的化学键振动特征,实现高灵敏度和特异性的无标记检测,在检测生物样品方面具有显著的优势和应用潜力,但面临数据特征复杂、光谱波动性、重复性、特征峰重叠等挑战。机器学习(machine learning, ML)通过算法和模型使计算机从数据中自动学习并进行预测或决策,在解析生物样品的复杂 SERS 光谱,推进其临床应用中展现出广阔前景。本文综述 ML 算法如何提升 SERS 性能,介绍用于 ML 的 SERS 数据的预处理方法、探讨 ML 处理 SERS 数据的功能类型和基本流程,并重点介绍其在分类和定量分析以及疾病诊断中的应用。此外,本文还讨论 ML 辅助 SERS 在分子结构预测、拉曼光谱数据库构建及 DNA 与 RNA 区分方面的潜力。最后,对 ML 集成 SERS 的挑战和未来发展方向进行展望。

**关键词:**表面增强拉曼散射;机器学习;生物医药;特征提取;疾病诊断

**中图分类号:**O657.37 **文献标志码:**A

**引用格式:**陈云帆,王也晨,王龙,等. SERS 协同机器学习在生物医药检测中的应用[J]. 山东大学学报(理学版),2025,60(10):23-41.

## Application of SERS collaborative machine learning in biomedical detection

CHEN Yunfan<sup>1</sup>, WANG Yeichen<sup>1</sup>, WANG Long<sup>2\*</sup>, AN Qi<sup>1\*</sup>, FENG Zeguo<sup>2</sup>

(1. School of Materials Science and Technology, China University of Geosciences, Beijing 100083, China; 2. Department of Pain Medicine, The First Medical Center of Chinese PLA General Hospital, Beijing 100853, China)

**Abstract:** Surface-enhanced Raman spectroscopy (SERS) enables highly sensitive and specific label-free detection by capturing the vibrational signatures of molecular chemical bonds. It offers significant advantages and broad application potential in the analysis of biological samples. However, its clinical translation remains limited due to challenges such as spectral complexity, variability, limited reproducibility, and overlapping characteristic peaks. Machine learning (ML), which allows computational models to learn patterns from data and make informed predictions or decisions, has shown great promise in deciphering complex SERS spectra and advancing their biomedical applications. This review highlights recent advances in the integration of ML with SERS, detailing how ML algorithms enhanced analytical performance. It also outlines common data preprocessing techniques for SERS, describes core ML workflows, and examines their roles in classification, quantitative analysis, and disease diagnosis. Furthermore, the review explores emerging applications such as molecular structure prediction, Raman spectral database development, and discrimination between DNA and RNA. Finally, the current challenges and future directions of ML-assisted SERS are discussed, with an emphasis on improving robustness, interpretability, and clinical applicability.

**Key words:** SERS; machine learning; biomedicine; feature extraction; disease diagnosis

## 0 引言

在临床诊断中,生物检测常作为疾病诊断的重要组成部分。然而,传统分析方法通常存在耗时、成本高昂且对实验操作要求严格的局限性,因此,开发快速、低成本且即时的检测技术具有重要的临床价值。生物

收稿日期:2025-04-15; 网络出版时间:2025-10-16 16:26:33

基金项目:国家自然科学基金资助项目(22272152)

第一作者:陈云帆(2000—),男,博士研究生,研究方向为表面增强拉曼光谱. E-mail:3003230001@email.cugb.edu.cn

\*通信作者:王龙(1988—),男,副主任医师,博士,研究方向为慢性疼痛的机制研究. E-mail:Flynn.xu@163.com

安琪(1983—),女,教授,博士,研究方向为自组装功能材料. E-mail:an@cugb.edu.cn

系统包含多种生物分子,如核酸、蛋白质和代谢物,这些分子的组成或浓度变化可用于识别不同的分子疾病亚型。因此,它们的变化可作为理想的疾病诊断标记物和治疗靶点<sup>[1]</sup>。拉曼光谱(Raman spectroscopy)是一种基于分子振动的技术,可通过分析与入射光频率不同的散射光谱获取分子振动和转动信息,已广泛应用于检测领域(图1(a))。然而,拉曼散射信号较弱,其典型散射截面分别比红外吸收和荧光过程低约 $10^6$ 和 $10^{14}$ 倍<sup>[2]</sup>。1974年,Fleischmann等<sup>[3]</sup>首次报道了表面增强拉曼光谱(surface-enhanced Raman spectroscopy, SERS)效应,发现吡啶分子在粗糙银电极上吸附后,其拉曼信号显著增强,并将其归因于粗糙银电极的面积增加所导致的吸附增强。这一结果于1977年分别被Van Duyne及Creighton研究团队验证<sup>[4-5]</sup>。他们发现,吡啶分子在粗糙银电极上的拉曼增强因子(enhancement factor, EF)可达 $10^5$ 到 $10^6$ 。然而,由于电极粗糙度仅使表面积增加约10倍,这一数值远低于实验观测到的拉曼信号增强幅度,因此研究人员推测,可能存在尚未发现的增强机制。Albrecht等<sup>[5]</sup>提出,该现象涉及等离激元激发的共振拉曼效应。这一独特的信号增强机制被称为SERS。SERS的发现极大推动了表面光谱学和分析科学的发展,突破了传统拉曼光谱灵敏度低的固有限制,并引领该领域进入了快速发展阶段。

贵金属(如金、银和铜)是最早用于SERS研究的材料。自胶体银和胶体金被首次发现五年后, Lee和Meisel<sup>[6]</sup>开发了一种广泛应用的合成方法,使其能够用于拉曼信号增强。1997年, Kneipp等<sup>[7]</sup>和Nie等<sup>[8]</sup>独立实现了一项重要突破,即单分子SERS(single-molecule SERS)。进一步研究表明,SERS信号的显著增强主要源于银纳米团簇之间的极小间隙所产生的局域电磁场增强效应。由于出色的增强性能,银和金纳米粒子已成为SERS研究和实际应用中最为常用的基底材料。金属纳米粒子的优越性能主要归因于其高EF、良好的光谱再现性、长期稳定性和优异的生物相容性<sup>[9]</sup>。此外,不同尺寸和形貌的金属纳米颗粒展现出不同的增强能力,因此,合理的纳米结构设计可进一步优化SERS基底的性能。例如,金、银纳米立方体、纳米棒、纳米星、纳米锥等多种纳米结构,以及具有纳米尖端或纳米间隙的多分支纳米结构(图1(b))均可显著提升SERS效果<sup>[10-11]</sup>。与金属基底相比,半导体材料具有额外的光学和电学特性,可通过电荷转移跃迁和化学增强机制在SERS中发挥重要作用。在SERS研究的早期阶段,活性半导体材料主要包括金属氧化物、卤化银及单一元素半导体(如硅和锗)<sup>[12-13]</sup>。近年来,石墨烯、过渡金属二硫族化物(transition metal dichalcogenides, TMDs)及有机半导体等新型半导体SERS基底相继被报道,并受到广泛关注<sup>[14-16]</sup>。尽管这些新型材料展现出良好的应用前景,但其EF仍然偏低(通常不超过 $10^3$ )。此外,已有研究尝试通过调控形貌和优化尺寸来提升半导体纳米结构的拉曼增强性能。在优化条件下,某些半导体材料可实现更高的EF( $10^5 \sim 10^6$ ),如二维非晶 $\text{TiO}_2$ 和 $\text{ZnO}$ 超结构,这些材料在某些应用中已展现出可与金属纳米基底相媲美的性能<sup>[17]</sup>。与此同时,SERS机理的研究也在不断深入。目前,SERS的增强机制主要分为两类:电磁增强机理(electromagnetic enhancement mechanism, EM)和化学增强机理(chemical enhancement mechanism, CM)。物理增强机理的核心在于电磁机制,其中贵金属纳米结构的等离激元共振起着关键作用。当入射光与等离激元金属纳米结构相互作用时,导电电子被激发并形成集体振荡,从而放大局部电磁场。当分子处于该局部电磁场中时,其拉曼信号显著增强(图1(c))<sup>[18]</sup>。此外,研究表明,电荷转移(charge transfer, CT)模型可能是化学增强的适用理论。当入射光子的能量与SERS基底和目标分子之间的费米能级差匹配时,基底与分子之间的电荷转移动力学过程会被激发,从而提高分子极化率,并增强SERS信号(图1(d))<sup>[19]</sup>。在一般情况下,物理增强机制可提供约 $10^3$ 至 $10^{12}$ 的EF,相比之下,化学增强机制的增强能力有限,其EF通常在 $10^3$ 左右<sup>[20]</sup>。因此,物理增强通常被认为是SERS效应的主要贡献机制。

SERS基底及增强机理的进步,使SERS成为生物医学领域中重要的定性与定量检测技术。生物样品如脂质、蛋白质和核酸等的SERS光谱包含丰富的分子信息<sup>[21]</sup>。此外,不同分子的共存及相互作用也可体现在SERS光谱中,甚至分子环境的微小变化都可能导致SERS光谱的显著变化。例如,当环境中的电荷、pH或分子浓度发生变化时,不同选择性及SERS增强效应的变化可能导致SERS光谱相对强度的变化,或某些特定官能团振动峰的缺失。因此,SERS具有极高的灵敏度及特异性,在多种疾病诊断中发挥着重要作用<sup>[22]</sup>。此外,在生物系统实验中,采用相同的SERS基底及激发条件可有助于解析复杂生物材料的SERS光谱。然而,随着SERS应用范围的扩大,其传统技术局限性日益凸显。首先,在许多应用中,处理大量光谱数据极具挑战性且耗时。例如,在多种生物标志物或病毒检测,以及复杂生物过程中反应监测中,需要分析海量的SERS光谱数据<sup>[23]</sup>。尤其在医学检测中,往往涉及大量患者样本,导致SERS光谱数据规模不断增长,

增加数据处理的难度。更为严重的是,SERS 光谱中不同分子信号的重叠问题,在很大程度上限制了其应用范围。例如,在小鼠血清样本中,不同分子的光谱信号高度重叠,使多个分子的特征峰同时出现在相同区域,从而导致 SERS 光谱难以解析<sup>[24]</sup>。尽管 SERS 技术在生物样品检测中具有无标记、高灵敏度和特异性强等显著优势,但其临床实际应用仍面临诸多挑战。一方面,SERS 光谱高度依赖于基底的物理和化学特性,导致其重复性和稳定性较差;另一方面,生物样品自身成分复杂,不同分子间的信号重叠、背景干扰和光谱漂移等问题严重影响数据的可解析性和分析准确性。此外,在大规模样本分析和高通量检测场景中,手动处理海量 SERS 数据不仅效率低下,而且难以准确提取关键特征,限制了其在精确诊断中的应用。

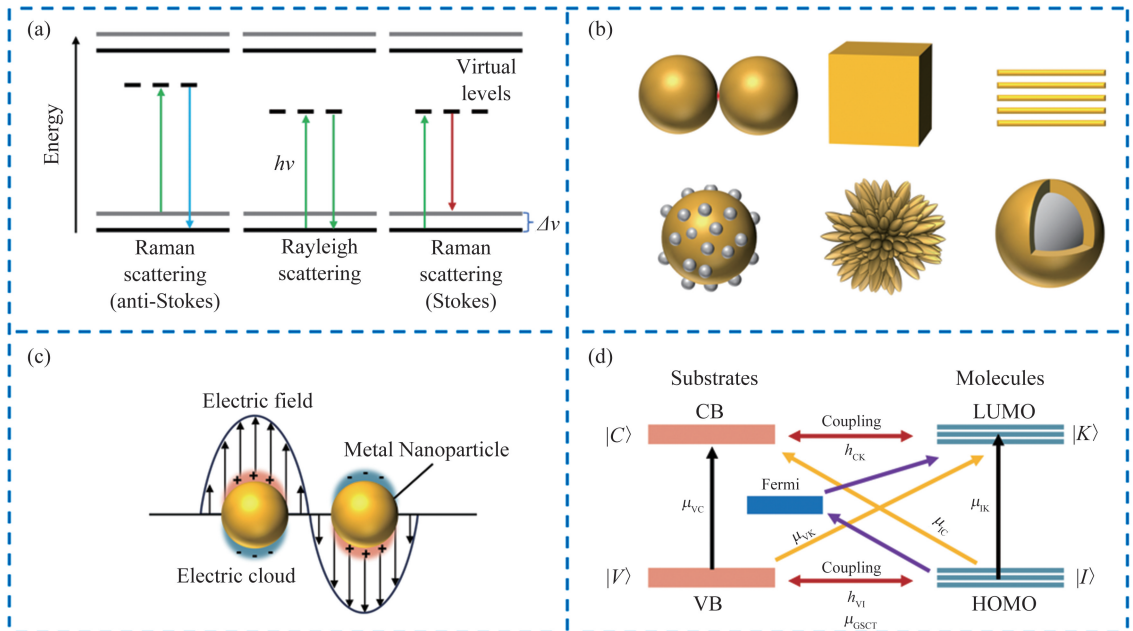


图 1 (a) 拉曼散射原理; (b) 典型的 SERS 基底; (c) 电磁增强机理; (d) 化学增强机理

Fig.1 (a) Raman scattering principle; (b) typical sers substrate; (c) electromagnetic enhancement mechanism; (d) chemical enhancement mechanism

面对上述多层次挑战,迫切需要一种高效、智能的数据分析工具来辅助 SERS 技术的深入发展。在此背景下,机器学习(machine learning, ML)技术的引入显得尤为关键和迫切。机器学习不仅能够自动识别和提取 SERS 光谱中的潜在规律,还能有效应对高维、非线性和噪声干扰等复杂数据特征,从而提升 SERS 数据分析的准确性、鲁棒性和效率。在疾病分类、分子识别、定量分析及分子结构预测等多个环节中,机器学习都展现出强大的建模能力和推广潜力。可以说,SERS 与机器学习的深度融合是推动其向临床实际应用转化的关键一步。通过将化学计量学应用于以矩阵形式呈现的数据(如 SERS 数据),研究人员能够构建机器学习预测算法,以用于化学数据的分类与分离。具体而言,这些预测算法能够学习不同类别数据之间的相似性与差异,例如从健康供体和患病供体的生物样本中提取的信息。一旦基于已知数据训练了模型,该模型便可用于对未知数据进行分类预测,从而实现特定且精准的医学筛查与诊断。机器学习不仅能够自主学习数据模式并优化预测模型,而且随着数据量的增加,模型的预测精度也不断提升。此外,机器学习还能从原始数据中提取关键特征,进而利用这些特征进行决策,以解决复杂的实际问题<sup>[25]</sup>。在数学维度上,机器学习依赖于模型的预测性能及其泛化能力,即利用不同格式的变量相关特征子集构建模型,并通过模式分析与降维方法训练最优模型<sup>[26]</sup>。在获取大规模数据后,通常需将数据划分为训练集与测试集。训练集用于机器学习模型的构建,而测试集则用于评估模型的泛化能力与预测性能。随着机器学习与 SERS 技术的结合,SERS 光谱可用于痕量物质或小分子的分析,并进一步实现疾病预测及病原体、生物标志物的分类。当需要对复杂光谱数据进行分类时,机器学习可提供有效的解决方案。此外,机器学习还可帮助研究人员提升 SERS 在定量分析中的应用价值。机器学习在 SERS 分析中的应用依赖于高质量的 SERS 基底。为确保机器学习模型能够准确捕捉不同组之间的真实差异,而非由基底不均匀性引入的偏差,SERS 基底需具备高度的均匀性、可重复性和稳定性。这些特性不仅有助于提高信号的可比性,还能增强模型的鲁棒性和泛化能力,从而提升分析结果的可靠性。

本综述将介绍机器学习与 SERS 结合使用的基本流程,包括样品预处理、数据预处理以及模型验证与评估。此外,本文总结机器学习结合 SERS 在生物医药检测中的分类与定量应用,并举例说明其在生物医学领域的典型应用,重点探讨其在预测分子结构、建立拉曼光谱数据库以及区分 DNA 和 RNA 方面的潜力。最后,本文讨论 SERS 与机器学习结合的优势与挑战,并展望其未来发展方向。

## 1 SERS 与机器学习的结合应用

### 1.1 SERS 与机器学习的应用流程

将 ML 与 SERS 结合的基本流程如图 2(a) 所示,主要包括样品预处理、数据预处理、模型训练与优化以及模型验证与评估 4 个阶段,具体说明如下。

#### (1) 样品预处理

生物样本(如血液、尿液、唾液、组织液)通常含有复杂背景,需通过稀释、离心或去蛋白等方式去除干扰,提升 SERS 信号质量。例如,血清稀释至  $10^{-6}$  以下有助于提高光谱稳定<sup>[27-28]</sup>;尿液则常采用两步离心法去除碎片与大分子蛋白,从而获得可用于机器学习分析的高质量信号(图 2(b))<sup>[29]</sup>。样品预处理的优化对于 SERS 结合 ML 进行数据分析至关重要,能够有效提高检测的灵敏度、特异性和可重复性。

#### (2) 数据预处理

为提高模型性能,需对 SERS 原始光谱进行基线校正、去噪、归一化等处理<sup>[30-31]</sup>。常用方法包括 Savitzky-Golay 平滑算法去除基线漂移,以及深度神经网络去除宇宙射线干扰和背景噪声<sup>[32]</sup>。合理预处理可有效提取目标分子特征信号,并避免模型训练过程中的误差积累<sup>[33]</sup>。

#### (3) 模型训练与优化

机器学习模型的训练过程包括训练集、验证集和测试集三部分。数据集的划分至关重要,通常采用  $k$  倍交叉验证( $k$ -fold cross-validation)以及独立外部验证来评估模型的泛化能力。

① 交叉验证:数据集随机划分为  $k$  组,其中  $k-1$  组用于模型训练,剩余 1 组用于评估。该过程重复  $k$  次,最终计算平均性能指标。常见方法包括留一法(leave-one-out)、留组法(leave-one-group-out)等。

② 外部验证:使用未参与模型训练的数据进行测试,以评估模型在真实应用中的适用性,并检测是否存在过拟合问题。外部验证结果对于机器学习模型在临床应用中的推广至关重要<sup>[34]</sup>。

#### (4) 模型验证与评估

在训练模型后,需要对其性能进行评估,以确保其准确性和鲁棒性。以下是常用的模型评估指标:

① 损失曲线(loss curve):用于监测训练过程中模型的误差变化,如图 2(c) 所示,损失值越低,模型性能越好<sup>[35]</sup>。

② 受试者工作特性曲线(receiver operating characteristic curve, ROC)及曲线下的面积(area under curve, AUC)值:ROC 曲线用于评估分类模型在不同阈值下的性能,AUC 可衡量模型的综合性能,如图 2(d) 所示<sup>[36]</sup>。

③ 混淆矩阵(confusion matrix):用于分析分类模型的预测正确率,矩阵的行列表示实际类别与预测类别,如图 2(e) 所示<sup>[37]</sup>。

④ 其他评估指标:包括准确率(accuracy)、精确率(precision)、召回率(recall)、特异性(specificity)等。此外,在回归问题中,常用均方误差(common mean square error, MSE)、均方根误差(root mean square error, RMSE)、平均绝对误差(mean absolute error, MAE)和标准偏差(standard deviation, SDM)等指标来衡量预测误差。

### 1.2 分类检测

在获得更精确的拉曼光谱后,提高拉曼光谱分析的效率和准确性一直是研究人员关注的重点。机器学习作为一种强大的数据分析工具,能够处理大规模、复杂且多样化的数据,并可根据任务需求分为分类、回归、聚类和降维这 4 大类。机器学习涵盖多种具体的化学计量学方法,每种方法均可针对特定问题提供有效解决方案。根据学习方式的不同,机器学习可进一步分为无监督学习(unsupervised learning)和监督学习(supervised learning)两大类。其中,无监督学习适用于数据间关系未知的情况,通过模式识别或聚类方法挖

掘数据间的潜在结构系<sup>[38]</sup>。而监督学习则依赖于已有的标注数据,通过学习输入特征与输出类别之间的映射关系,构建最优模型以识别拉曼光谱中的生物标志物或疾病状态。表 1 展示 SERS 结合机器学习在不同分析任务中的方法类型与监督方式的对应关系。以下将分别介绍这两类算法在 SERS 数据分析中的具体应用。

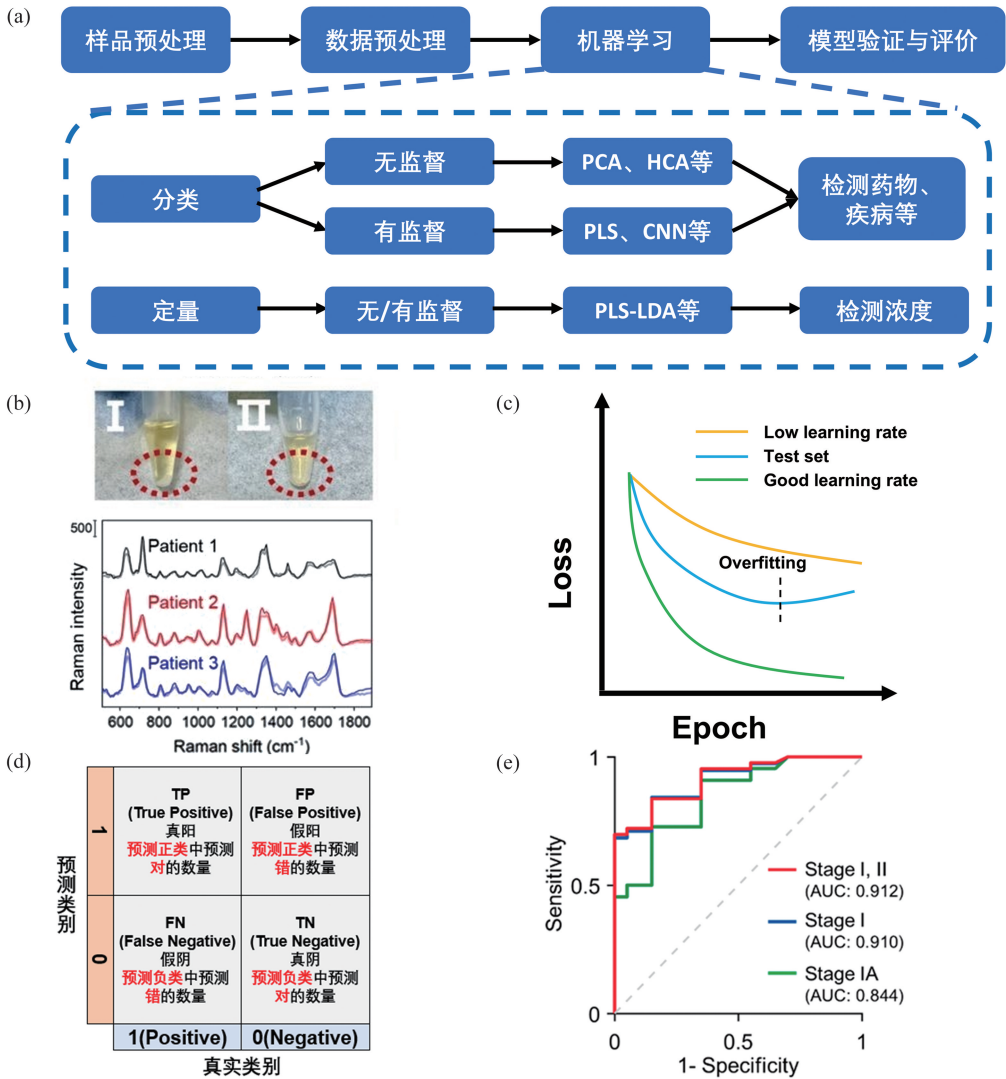


图 2 (a) 基于机器学习的 SERS 数据训练和测试的一般过程<sup>[29]</sup>; (b) 上图为第一步(I)和第二步离心(II)后的人类尿液样本照片,下图为超滤前后尿液上清液样品的光谱比较<sup>[29]</sup>; (c) 训练过程中的损失曲线,当过拟合时,测试集的损失会增加,表明应在虚线内停止拟合<sup>[36]</sup>; (d) 显示测试数据分类模型性能的混淆矩阵表示<sup>[36]</sup>; (e) ROC 曲线和相应的 AUC 值<sup>[36]</sup>

Fig.2 (a) General process of SERS data training and testing based on machine learning<sup>[29]</sup>; (b) the upper image shows human urine samples after the first step ( I ) and the second step of centrifugation ( II ), while the lower image compares the spectra of urine supernatant samples before and after ultrafiltration<sup>[29]</sup>; (c) loss curve during the training process, when overfitting occurs, the loss of the test set increases, indicating that fitting should be stopped within the dashed line<sup>[36]</sup>; (d) confusion matrix representation showing the classification model's performance on test data<sup>[36]</sup>; (e) ROC curve and corresponding AUC values<sup>[36]</sup>

表 1 SERS 结合机器学习的任务类型与监督方式的对应关系

Table 1 Correspondence between task types and supervision modes in SERS combined with machine learning

| 类型 | 描述            | 典型算法            | 监督类型             | SERS 中的应用                 | 参考文献    |
|----|---------------|-----------------|------------------|---------------------------|---------|
| 分类 | 将样本归类到不同类别    | SVM、RF、CNN      | 有监督              | 识别不同类别的生物分子或疾病状态          | [39-40] |
| 回归 | 预测连续变量        | PLSR、MLR、深度学习回归 | 有监督              | 定量分析目标分子浓度,如药物代谢监测        | [41-42] |
| 聚类 | 按相似性对数据分组     | k-means、层次聚类    | 无监督              | 发现未知样本类别,如不同类型细菌分类        | [43]    |
| 降维 | 提取关键特征,减少数据维度 | PCA、t-SNE、LDA   | 无监督/<br>有监督(LDA) | 提取 SERS 主要特征,提高分类与定量分析的精度 | [44-45] |

### (1) 无监督学习在 SERS 分析中的应用

无监督机器学习因其在模式识别、数据探索和数据可视化方面的优势,在 SERS 分析中得到了广泛应用<sup>[46]</sup>。目前,常用的无监督学习算法包括主成分分析(principal component analysis, PCA)和分层聚类分析(hierarchical cluster analysis, HCA)等。PCA 是 SERS 数据处理中最常用的降维方法,能够在保留数据主要信息的同时减少变量数量,从而提取用于区分不同类别分子的关键光谱特征<sup>[47]</sup>。通过降维,PCA 可以将高维数据转换为二维或三维评分图,以可视化不同类别数据的聚类情况,并评估光谱差异的统计学显著性。例如,Ding 等<sup>[48]</sup>利用三维 PCA 显示芬太尼(fentanyl, FE)、吗啡(morphine, HER)、氯胺酮(ketamine, KET)及其混合物的 SERS 光谱聚类情况,验证了 SERS 平台在毒品检测与鉴别中的应用潜力(图 3(a))。此外,Lim 等<sup>[49]</sup>结合 SERS 与 PCA 对感染不同流感病毒的细胞进行了无标记鉴定。HCA 通过构建树状聚类图(dendrogram),直观地展示光谱之间的相似性或差异性,使研究人员能够更清晰地理解数据的内部关系<sup>[50]</sup>。相较于 PCA,HCA 具有无需降维即可直接处理完整光谱的优势,因此在需要保留所有光谱信息的生物学应用中具有较大潜力。例如,在 SERS 细菌分类研究中,HCA 被用于分析 6 种细菌的 600 组光谱数据,最终成功将其分为不同聚类,并揭示了细菌分类的三级特征:表面电荷;总体生化特征;功能基团的类型和数量(图 3(b))<sup>[43]</sup>。尽管无监督学习在 SERS 数据分析中表现出色,但其在处理干扰因素较多的复杂样本时仍存在一定局限性。例如,实验中的背景噪声、干扰物质及数据采集条件的变化可能影响聚类结果的可靠性。此外,由于无监督模型缺乏明确的标签信息,其在特定任务(如疾病诊断)中的适用性有限。针对这些挑战,研究人员通常结合监督学习方法,以提高模型的分类能力和泛化性能。

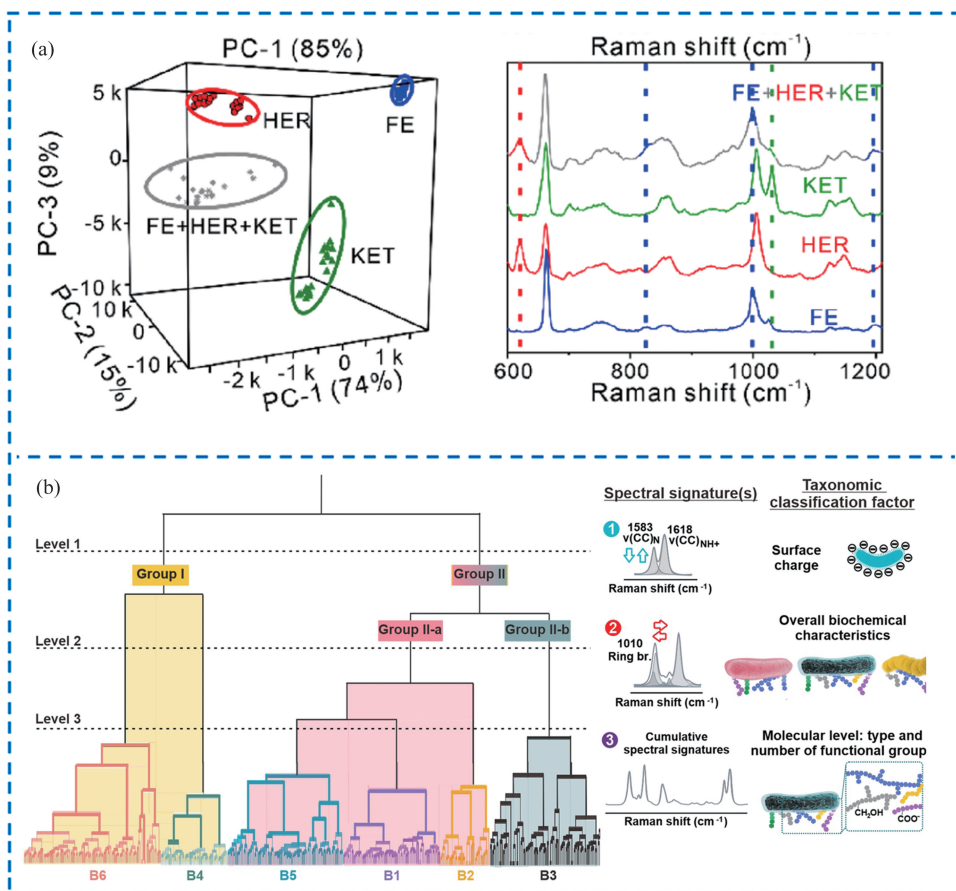


图3 无监督学习在 SERS 中的应用: (a) 三元混合物( $c_{FE}$ : 0.1  $\mu\text{g}/\text{mL}$ ,  $c_{HER}$ : 10  $\mu\text{g}/\text{mL}$ ,  $c_{KET}$ : 10  $\mu\text{g}/\text{mL}$ )及其单一组分的 3D PCA 得分图和相应的 SERS 光谱<sup>[48]</sup>; (b) 对测量的 4-巯基吡啶(4-mercaptopyridine, MPY) SERS 进行系统聚类分析,使用基于细胞外基质(extracellular matrices, ECM)表面化学类型的分类因子即表面电荷、整体 ECM 化学特征以及相互作用功能群的类型和数量,将细菌物种分为 3 个主要级别<sup>[43]</sup>

Fig.3 Application of unsupervised learning in SERS: (a) 3D PCA score plot and corresponding SERS spectra of a ternary mixture ( $c_{FE}$ : 0.1  $\mu\text{g}/\text{mL}$ ,  $c_{HER}$ : 10  $\mu\text{g}/\text{mL}$ ,  $c_{KET}$ : 10  $\mu\text{g}/\text{mL}$ ) and its individual components<sup>[48]</sup>; (b) systematic cluster analysis of the measured MPY SERS, using classification factors based on ECM surface chemistry, including surface charge, overall ECM chemical characteristics, and the type and number of interacting functional groups, categorizing bacterial species into three major levels<sup>[43]</sup>

## (2) 有监督学习在 SERS 分析中的应用

监督机器学习算法依赖于标注数据集(例如已知分析物的分子结构、浓度或样本的疾病状态),能够显著提高 SERS 分析的准确性<sup>[51]</sup>。目前,已开发出多种监督学习算法,其复杂度和灵活性各异,范围涵盖从简单的偏最小二乘法(partial least squares method, PLS)到更复杂的支持向量机(support vector machine, SVM)、基于树的算法以及基于神经网络(neural network, NN)的算法<sup>[51-54]</sup>。通常,在样本数量有限且 SERS 频移与分析物浓度之间的关系呈线性时,简单算法表现更优;而对于大规模数据集或非线性光谱—分析物浓度关系,则更适合采用复杂算法<sup>[46]</sup>。值得注意的是,这些监督学习算法不仅适用于定性分析(如疾病分类),也可用于定量分析(如生物标志物浓度测定)。PLS 是最常用的回归方法之一,其核心目标是寻找一组潜在因子,以最大程度地解释 SERS 光谱与标注数据之间的多维协方差,并据此构建回归或分类模型<sup>[55]</sup>。例如,Leong 等<sup>[55]</sup>使用纳米多孔金纳米粒子作为 SERS 基底,采用 PLS 回归建立校准曲线,以量化 D-色氨酸和 L-色氨酸的混合物浓度(图 4(a))。所构建的模型具有良好的拟合度(交叉验证  $R^2 = 0.96$ ),并可在绝对误差 1% 的范围内准确测定未知混合物的成分比例。在另一项研究中,Jančí 等<sup>[41]</sup>开发了一种用于检测鱼类中组胺含量的 SERS 定量方法,基于 PLS 回归模型评估了 SERS 光谱与组胺浓度的关系。该方法可在 0~200 mg/kg 的浓度范围内实现可靠测定,并在  $1\ 139.9 \sim 1\ 643.7\ \text{cm}^{-1}$  的光谱区域内表现出良好的线性趋势(预测  $R^2 = 0.962$ , 偏离度为 7.250)。由于 PLS 主要适用于线性关系分析,在处理非线性或更复杂的数据时,通常需要采用更高级的机器学习算法。SVM 适用于小样本数据集,并能有效处理非线性光谱—分析物关系。其基本原理是将数据投影至高维空间,以寻找最优超平面,从而实现类别划分<sup>[56]</sup>。Diao 等<sup>[57]</sup>基于 SERS 光谱信号结合 SVM 算法,对不同细胞周期阶段(G0/G1、S 和 G2/M 期)的外泌体进行了分类,准确率达到 85%。其中,处于 G0/G1 期的外泌体预测准确率接近 95%,而 S 期和 G2/M 期的预测准确率分别为 81.8% 和 64.7%。此外,通过 ROC 曲线评估 SVM 预测性能,G0/G1、S 和 G2/M 期的 AUC 值分别为 0.98、0.87 和 0.84(图 4(b))。Kim 等<sup>[58]</sup>进一步利用 SVM 对健康个体、感染性脓毒症患者(有/无脓毒性休克)进行了分类分析,分类准确率达到 95.0%,精确度为 95.8%。这些研究结果表明,SVM 在疾病分类任务中展现出卓越性能,并在 SERS 诊断平台中具有广泛应用潜力。

树的算法(如随机森林(random forest, RF))通过构建多个决策树并聚合其结果,提高模型准确性并减少过拟合。例如,Shu 等<sup>[59]</sup>将无标记 SERS 和激光解吸电离质谱(laser desorption ionization mass spectrometry, LDI-MS)结合,开发了一种针对中风患者血清样本的双模态光谱分析方法。研究人员基于 RF 和最小绝对收缩与选择算子回归(least absolute shrinkage and selection operator, LASSO),构建了 SERS/LDI-MS 融合的中风筛查模型。与单模态方法相比,该模型在发现队列和验证队列中的 AUC 分别为 0.949 (95%CI:0.917~0.977) 和 0.911 (95%CI:0.812~0.984) (图 4(c))。此外,该方法的特异性(0.86)显著优于单一模态 SERS(0.64) 和 LDI-MS(0.73),表明双模态光谱在早期中风检测中的优势。基于卷积神经网络(convolutional neural network, CNN)的算法能够有效处理复杂数据,尤其适用于大规模 SERS 数据分析。其核心优势在于能够通过卷积层和池化层提取关键特征,从而增强光谱分析能力<sup>[60-61]</sup>。Chen 等<sup>[62]</sup>利用柠檬酸钠还原的 Ag 纳米粒子获取芬太尼成瘾和吗啡成瘾小鼠血清的 SERS 光谱,并采用 CNN 进行分类,最终准确率分别达到 100% 和 87% (图 4(d))。研究人员使用 F1-score 评估模型性能,CNN 在 5 种有监督学习模型中表现最佳(F1-score=0.885)。类似地,Jiang 等<sup>[63]</sup>开发了一种轻量级 CNN(MLS-CNN),用于分析 90 例健康对照和 585 例癫痫患者血清样本的 SERS 数据,并成功实现癫痫病因的精准分类。该模型的整体诊断准确率达 100%,病因分类准确率超过 89%,表明 CNN 在 SERS 领域的广泛应用潜力。尽管监督学习在 SERS 分析中展现出巨大潜力,但仍面临过拟合(overfitting)问题,尤其是在光谱样本数量不足的情况下。特别是对于基于树和神经网络的算法,通常需要数百至数万个光谱数据来构建稳健模型。此外,不充分的模型评估与验证也可能导致模型泛化能力下降。因此,在数据采集、模型构建和评估等各个环节,需要采取严格的预防措施,以确保 SERS 分析的鲁棒性和可靠性。

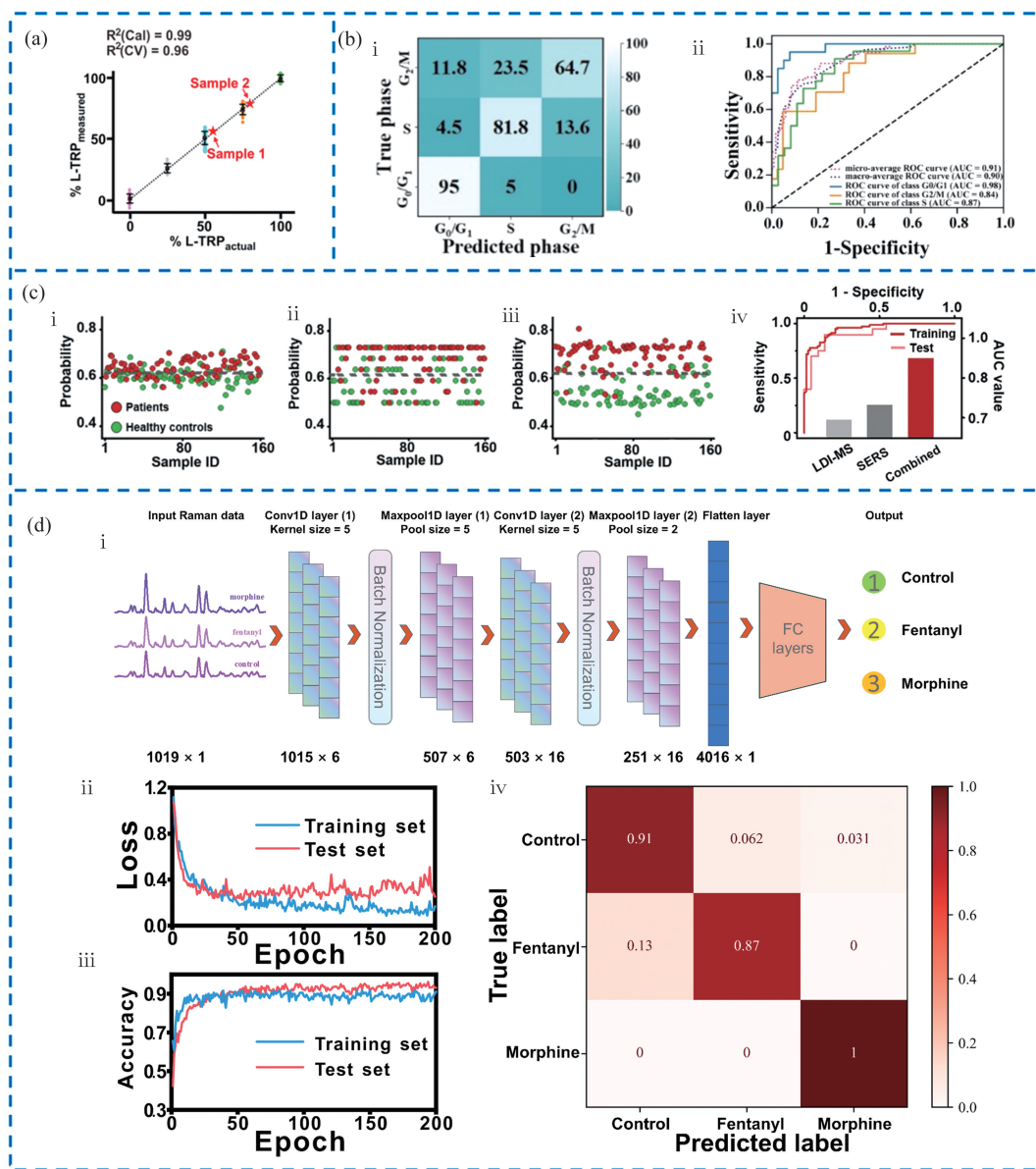


图4 有监督学习在SERS中的应用: (a) 测量值摩尔百分比与实际值摩尔百分比的L-色氨酸的PLS回归图<sup>[55]</sup>; (b) (i) 来自3个不同细胞周期阶段的外来体的SERS光谱获得的SVM预测结果的混淆矩阵, (ii) SVM模型的ROC曲线<sup>[57]</sup>; (c) 基于(i) LDI-MS、(ii) SERS和(iii) SERS/LDI-MS的训练队列的中风患者(红色)和健康对照(绿色)的预测概率分层, (iv) 基于LDI-MS、SERS和SERS/LDI-MS数据的诊断模型的AUC值和ROC曲线<sup>[59]</sup>; (d) (i) CNN模型框架, 学习过程中, (ii) 准确性曲线和(iii) 训练和测试数据集的损失曲线, (iv) 使用CNN进行成瘾分类的混淆矩阵<sup>[62]</sup>

Fig.4 Application of supervised learning in SERS: (a) PLS regression plot of measured values molar % versus actual values molar % of L-tryptophan<sup>[55]</sup>; (b) (i) confusion matrix of SVM prediction results obtained from SERS spectra of exosomes from three different cell cycle stages, (ii) ROC curve of the SVM model<sup>[57]</sup>; (c) stratified prediction probabilities of stroke patients (red) and healthy controls (green) in the training cohort based on (i) LDIMS, (ii) SERS, and (iii) SERS/LDI-MS; (iv) AUC values and ROC curves of diagnostic models based on LDI-MS, SERS, and SERS/LDI-MS data<sup>[59]</sup>; (d) (i) framework of the CNN model, (ii) accuracy curve and (iii) loss curve of the training and testing datasets during the learning process; (iv) confusion matrix for addiction classification using the CNN model<sup>[62]</sup>

### 1.3 定量检测

随着机器学习技术的进步,化学计量学方法得到了广泛应用,使SERS在复杂生物基质中的量化数据处理能力得到显著提升。其中,PLS多元校准方法尤为常见,因为它可以综合考虑峰值信息或整个拉曼光谱数据,从而实现更可靠的定量分析。例如,Leong等<sup>[64]</sup>在一项病例对照研究中,利用偏最小二乘判别分析(partial least squares discriminant analysis, PLS-DA)模型对COVID-19患者和健康个体的呼吸样本进行分

类,并借助一系列不同分子受体功能化的 SERS 平台捕获和检测呼吸样本中的挥发性有机化合物(图 5(a))。研究表明,该 PLS-DA 模型在区分无症状和有症状 COVID-19 患者方面的分类敏感性达 96.2%,特异性高达 99.9%。此外,尿酸(uric acid, UA)是嘌呤代谢的终产物,已被证明是先兆子痫、痛风和心血管疾病的重要生物标志物。而肌酐(creatinine, CRN)则是评估肾脏滤过功能的关键指标。在相关研究中, Jones 等<sup>[65]</sup>采用 E-SERS 结合机器学习方法,对 UA 和 CRN 进行了定量复用检测。研究发现,UA 和 CRN 的 SERS 信号强度表现出不同的电位依赖性,从而能够构建具有重要意义的三维 E-SERS 数据集。随后,研究人员采用两步偏最小二乘回归-多层感知器(partial least squares regression-multi layer perceptron, PLSR-MLP)机器学习算法,对 E-SERS 数据进行分析,并成功预测了未见光谱中的分析物浓度,预测精度高达 0.94(图 5(b))。这一方法可推广至尿液或其他体液中的一系列生物标志物检测,为复杂生物样本的高精度分析提供了可能。除了 PLS,高斯过程回归(gaussian process regression, GPR)也被广泛应用于定量分析。Garg 等<sup>[66]</sup>开发了一种多变量非参数回归模型,将 SERS 测量值映射至 SARS-CoV-2 浓度范围( $10^3$  至  $10^6$  pfu/mL)。研究利用探地雷达方法,通过裂解的 SARS-CoV-2 样本在  $10^2 \sim 10^6$  pfu/mL 浓度范围内建立 SERS 数据集,并基于有理二次核函数构建 GPR 模型(图 5(c))。结果表明,该模型在  $10^3 \sim 10^6$  pfu/mL 范围内的预测值与实际浓度具有较高一致性, RMSE 为 0.52, 决定系数(coefficient of determination,  $R^2$ )达 0.84。在农药残留检测方面, Wan 等<sup>[67]</sup>采用 SVM(支持向量机)模型,不仅对不同农药种类进行了分类,还实现了对农药浓度的预测。研究针对噻苯咪唑(thiabendazole, TBZ)浓度( $10^{-5} \sim 10^{-10}$  mg/L)的 SERS 光谱数据进行分类,并以 40%的数据用于训练,剩余 60%用于模型测试。所得混淆矩阵(图 5(d))表明, SVM 与 SERS 结合可以有效区分不同浓度梯度下的农药残留,进一步验证了该方法在农药残留分析中的应用潜力。随着复杂数据分析能力的提升,以及对多变量响应模式的深入挖掘, SERS 定量检测已成为一个集成多元数据处理的综合方法。这些先进的机器学习算法显著提高了 SERS 在化合物识别、分类和定量分析中的性能,为生物医药检测提供了更加精确和高效的解决方案。

为更好地理解 ML 在 SERS 分析中的优势,有必要将其与传统化学计量学方法进行区分。传统化学计量学(如 PCA、PLS 等)通常依赖于线性假设和人为设定的特征选择规则,适用于数据结构清晰、变量间关系线性化的情形。这些方法强调可解释性强、计算效率高,适合处理中小规模数据。然而,在 SERS 光谱分析中,生物样本往往具有高维度、非线性和强噪声的特点,导致传统化学计量学难以捕捉其中潜在复杂模式。相比之下, ML 方法(如 SVM、RF、NN 等)不依赖先验的线性假设,具备强大的非线性建模能力与自动特征提取能力。以 SVM 为例,其通过核函数映射可有效区分高维空间中的非线性数据类别;而深度学习如 CNN 还能直接从原始光谱中学习多层次特征,实现端到端的诊断判别。这些方法能够处理大规模、高噪声数据,适应复杂样本中的隐含结构,从而显著提升模型的泛化能力和预测精度。因此,尽管化学计量学在早期 SERS 数据分析中发挥过重要作用,但面对当前 SERS 在生物医学中的高通量应用需求,机器学习提供了更加灵活、强大和智能的解决方案,成为推动 SERS 技术临床转化的关键工具。

#### 1.4 SERS-ML 中机器学习算法创新

SERS-ML 的融合为疾病诊断、环境监测与药物筛选等应用场景开辟了全新路径,而机器学习算法的持续创新则是支撑该技术体系高效、稳定运行的核心要素。目前, SERS-ML 系统中常用的算法包括 PCA、SVM、KNN、RF 等传统方法,它们在小样本拉曼光谱的分类与特征提取中已展现出良好性能。然而,随着数据规模扩展与样本复杂性增加,传统算法在特征自动化提取、非线性建模与泛化能力方面逐渐暴露出局限。近年来,深度学习(deep learning, DL)方法因其端到端特征学习能力,在 SERS 数据建模中受到广泛关注。典型结构如 CNN、ResNet 等在肿瘤细胞识别、细菌分类等任务中表现出色,特别适用于复杂背景干扰下微弱光谱信号的提取。同时,自监督学习、迁移学习等策略也被引入以缓解标注样本稀缺、数据分布不均等问题<sup>[68]</sup>。例如, Tan 等<sup>[69]</sup>将 SERS 的分子特异性与自适应迁移学习策略结合,成功实现对 11 种不同肉碱混合物中各组分的识别与定量,达到 100%的识别准确率和 3%的平均定量误差。该模型仅依赖纯化合物光谱训练,便可准确解析复杂未知混合物,显著降低数据采集成本,展现出迁移学习在复杂体系分析中的巨大潜力。Zheng 等<sup>[70]</sup>开发的“胶体数字 SERS”平台,则利用光谱信号的二值化处理(“开/关”模式)与深度神经网络结合,实现了对单分子事件的精准可视化,显著降低了假阳性率。该平台已成功应用于哺乳动物细胞培养基中痕量成分的检测,表现出优异的准确性与可重复性,展示其在生物制造和过程监控中的可转化潜力。在另

一项研究中, Sun 等<sup>[71]</sup>创新性地将 SERS 与结构强化的深度神经网络融合, 构建了 CNN-GRU-Attention 混合模型用于溴氰菊酯的高灵敏度定量检测。该模型引入门控循环单元(gate controlled cycle unit, GRU)与注意力机制, 显著提升了特征提取效率和非线性建模能力。在实验中, 其预测结果达到了  $R^2 = 0.9827$ 、 $RMSE = 0.3896$ , 较传统方法准确率提升了 8%, 误差降低 40%。研究表明, 该融合策略通过多维特征聚焦与时间依赖建模, 突破了传统方法在痕量农药检测中的灵敏度瓶颈, 为环境污染物监测提供了创新路径。一方面, 随着 SERS-ML 向实际应用推进, 模型的“黑箱”特性也成为制约因素。在临床场景中, 决策过程的不透明性、对相关性的过度依赖以及合规性要求(如 FDA/EMA 认证)构成了显著障碍。对此, 研究者提出了如不确定性量化技术以识别低置信度预测, 以及波数注意力机制以提升关键光谱区域关注度等方法, 从而增强模型的可解释性与临床信任度<sup>[72]</sup>。另一方面, SERS-ML 系统的部署也面临多重现实挑战, 包括对高质量样本的依赖、训练计算成本高、模型迁移困难等<sup>[73]</sup>。未来研究应聚焦于开发低成本、可批量制造的 SERS 基底, 结合联邦学习和边缘计算技术, 减少对中心化数据与云端资源的依赖, 以推动 SERS-ML 从实验室研究向临床与产业级别的实用化转化。

综上所述, 算法的持续演进正驱动 SERS 技术从信号获取向智能解析迈进。通过提升模型表达能力、降低数据依赖、增强可解释性与可部署性, SERS-ML 在多个实际应用中展现出变革潜力。算法创新不仅在理论性能上取得显著突破, 更是推动 SERS 技术实现从原型验证到真实世界落地的关键一环。

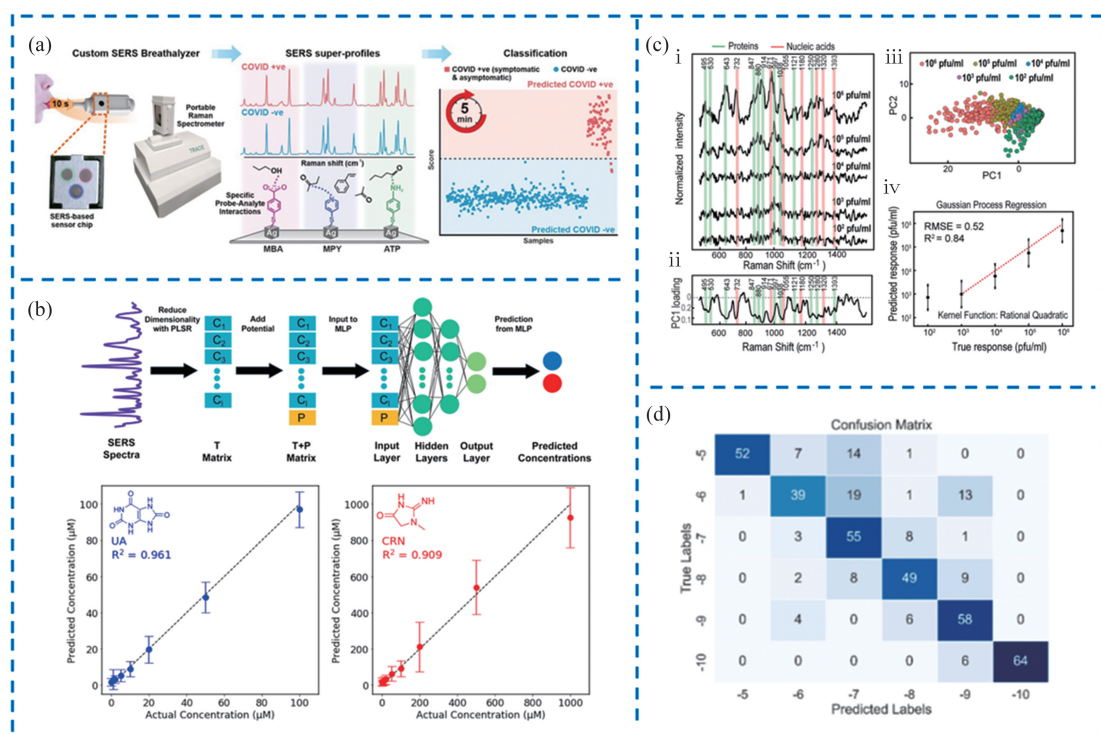


图5 SERS 在定量检测中的实例: (a) 利用多分子受体功能化的 SERS 基底捕获和检测与 COVID-19 相关的呼吸挥发性有机化合物, 然后根据患者的 COVID-19 感染状况利用 PLS-DA 对 SERS 光谱进行分类<sup>[64]</sup>; (b) 利用未见 E-SERS 光谱预测 UA 和 CRN 浓度的两步 PLSR-MLP 算法示意图, 测试数据集 UA 和 CRN 的实际浓度与两步 PLSR-ANN 模型预测的浓度的比较<sup>[65]</sup>; (c) (i) SARS-CoV-2 多变量回归定量分析, 裂解的 SARS-CoV-2 的平均校准 SERS 光谱, (ii) 主成分(principal component, PC)加载, (iii) 浓度范围为  $10^2 \sim 10^6$  pfu/mL 的裂解 SARS-CoV-2 的 SERS 光谱的 PCA 分析所得的 PC 评分散点图, (iv) 使用具有有理二次核函数的 GPR 模型预测 SARS-CoV-2 浓度<sup>[66]</sup>; (d) TBZ 集中训练集的混淆矩阵<sup>[67]</sup>

Fig.5 Representative applications of SERS in quantitative analysis: (a) detection and capture of COVID-19-related respiratory volatile organic compounds using SERS substrates functionalized with multimolecular receptors, followed by classification of SERS spectra based on patients' COVID-19 infection status using PLS-DA<sup>[64]</sup>; (b) schematic diagram of the two-step PLSR-MLP algorithm for predicting UA and CRN concentrations using unseen E-SERS spectra, along with a comparison of actual concentrations of UA and CRN in the test dataset versus predictions from the two-step PLSR-ANN model<sup>[65]</sup>; (c) (i) multivariate regression quantitative analysis of SARS-CoV-2, showing the mean calibrated SERS spectra of lysed SARS-CoV-2, (ii) PC loadings, (iii) PCA score scatter plot obtained from the SERS spectra of lysed SARS-CoV-2 within a concentration range of  $10^2$  to  $10^6$  pfu/mL, (iv) Prediction of SARS-CoV-2 concentration using a GPR model with a rational quadratic kernel function<sup>[66]</sup>; (d) confusion matrix of TBZ concentration training set<sup>[67]</sup>

### 1.5 SERS-ML 在疾病诊断中的应用

SERS-ML 在疾病诊断领域展现出巨大潜力,尤其是在多种癌症、感染性疾病、神经系统疾病及代谢相关疾病的检测中,已成为一种高灵敏、非侵入、自动化潜力强的创新分析工具。

目前,SERS-ML 在疾病诊断中的应用研究主要集中在利用不同类型的生物样本检测各种形式的癌症。在一项研究中,肿瘤抑制蛋白 p53 的野生型被认为是一种重要的癌症生物标志物。Hernánovel-Vidales 等<sup>[74]</sup>通过 PCA 和 SVM 算法分析其 SERS 光谱,发现野生型与突变型 p53 的分离率可达 94%。使用留一法交叉验证后,模型准确率保持较高,检测限为 0.946 mmol/L,为使用自发拉曼光谱开展癌症相关研究提供了可能。癌症转移是导致患者死亡的主要原因之一。在癌症早期,从原发性或转移性肿瘤中脱落的循环肿瘤细胞(circulating tumor cell, CTC)可能进入外周血,并扩散至远处组织。CTC 可通过体外培养、表型及基因型分析等方式实现精确检测,已成为肿瘤早期诊断、术后评估、治疗监测及复发/转移判断中的关键生物标志物<sup>[23]</sup>。近年来,SERS 技术在 CTC 检测方面取得了显著进展。例如,通过 AuNP-MBA-rBSA-FA 复合物制备的一种高灵敏度、高特异性的 SERS 生物探针,在 5~500 个癌细胞/mL 的范围内,其 SERS 峰强度与癌细胞数之间呈良好线性关系( $R^2=0.9935$ ),表明该探针可实现外周血中 CTC 的定量检测<sup>[75]</sup>。此外,Pang 等<sup>[76]</sup>开发了一种基于  $Fe_3O_4@$ 核壳结构的双信号增强策略,其在检测外周血中肝癌 CTC 时,灵敏度可达 1 个细胞/mL。尽管检测灵敏度较高,但该类 SERS 系统在光谱稳定性、生物相容性及抗干扰能力方面仍有待进一步优化。Fang 等<sup>[77]</sup>使用银胶膜作为 SERS 基底,测量了血细胞与多种肿瘤细胞的表面拉曼光谱(图 6(a)),结果显示,核酸相关特征峰在肿瘤细胞与血细胞之间存在显著差异。

除肿瘤外,SERS-ML 在病原体检测中的应用亦逐渐深入。近年来,多功能 SERS 平台<sup>[78]</sup>与微纳米结构平台<sup>[79]</sup>被广泛用于提升拉曼信号强度、提高病原体富集效率及加快检测速度。尽管上述技术取得了积极成果,但在面对复杂临床样本时,仍存在定性定量分析准确性不足、结果可重复性差等挑战。为提升性能,研究者尝试结合机器学习以增强分析能力、提升灵敏度并缩短数据处理时间。例如,有研究利用便携式拉曼光谱仪结合 PCA 方法,对白色念珠菌的表型变异进行快速识别,并通过构建的 PCA 模型预测不同作用机制下已知和未知抗真菌药物的药效响应<sup>[80]</sup>。在细菌检测方面,Hu 等<sup>[81]</sup>对来自 9 种临床常见 *Staphylococcus* 菌属的 117 株细菌共 2 752 个 SERS 光谱数据进行了系统研究。结果显示,在 3 种无监督学习算法中,基于密度的空间聚类表现最优,Rand 指数为 0.973 3;在 10 种监督学习算法中,卷积神经网络准确率达到 98.21%,AUC 高达 99.93%(图 6(b))。这项研究不仅验证了 SERS-ML 模型在复杂细菌种类识别中的高效性,也体现了其在临床病原体快速诊断中的应用前景。

在神经系统疾病中,SERS-ML 在中风的早期识别与分类中展现出变革潜力。Freitas 等<sup>[82]</sup>开发了一种基于银纳米星与血浆共孵育的低成本 SERS 方案,通过检测胶质细胞酸性蛋白等中风标志物,结合机器学习模型,实现了出血性与缺血性中风的快速区分,整个流程在 15 min 内完成,模型可在几秒内输出预测结果,为床旁即时决策提供了可能。此外,Shu 等<sup>[59]</sup>提出的 SERS 与激光解吸电离质谱联合双模分析平台,在仅 0.1  $\mu$ L 血清样本中实现了高通量、低预处理、双指纹信息提取,通过机器学习模型建立的中风筛查系统,AUC 分别达 0.949 和 0.911。在代谢性疾病诊断中,SERS-ML 也呈现出强大适应性。Chen 等<sup>[62]</sup>通过分析血清中代谢物变化,成功识别了阿片类药物成瘾模型小鼠。Kao 等<sup>[83]</sup>进一步将实验室级 SERS 平台应用于超痕量代谢物的临床多重检测。面对低拉曼截面、弱结合力等挑战,他们提出“限制与捕获”策略,利用超疏水纳米平台靶向富集尿液中的流产相关代谢物,结合 PCA 和 PLSR 模型,成功将分子指纹转化为可量化读数,并在 40 例先兆流产患者中验证其与妊娠结局的显著相关性。整个流程仅需 30 min,具备实际筛查价值。

尽管 SERS-ML 在多个疾病方向上表现出强劲性能,其实际临床推广仍面临诸多挑战:包括临床样本的高度异质性、缺乏统一的数据标准化流程、模型跨平台泛化能力不足,以及分析结果的可解释性与可验证性欠佳。未来研究应从以下方面重点发力:构建高质量、可共享的标准化 SERS 数据库;发展多模态融合策略,将拉曼数据与影像、组学、临床指标等多源数据联合建模以提升鲁棒性;开发自动化、高通量的便携式检测平台,并引入可解释性 AI 方法与模型不确定性评估机制,增强 SERS-ML 在真实世界医疗环境中的可靠性与可操作性。

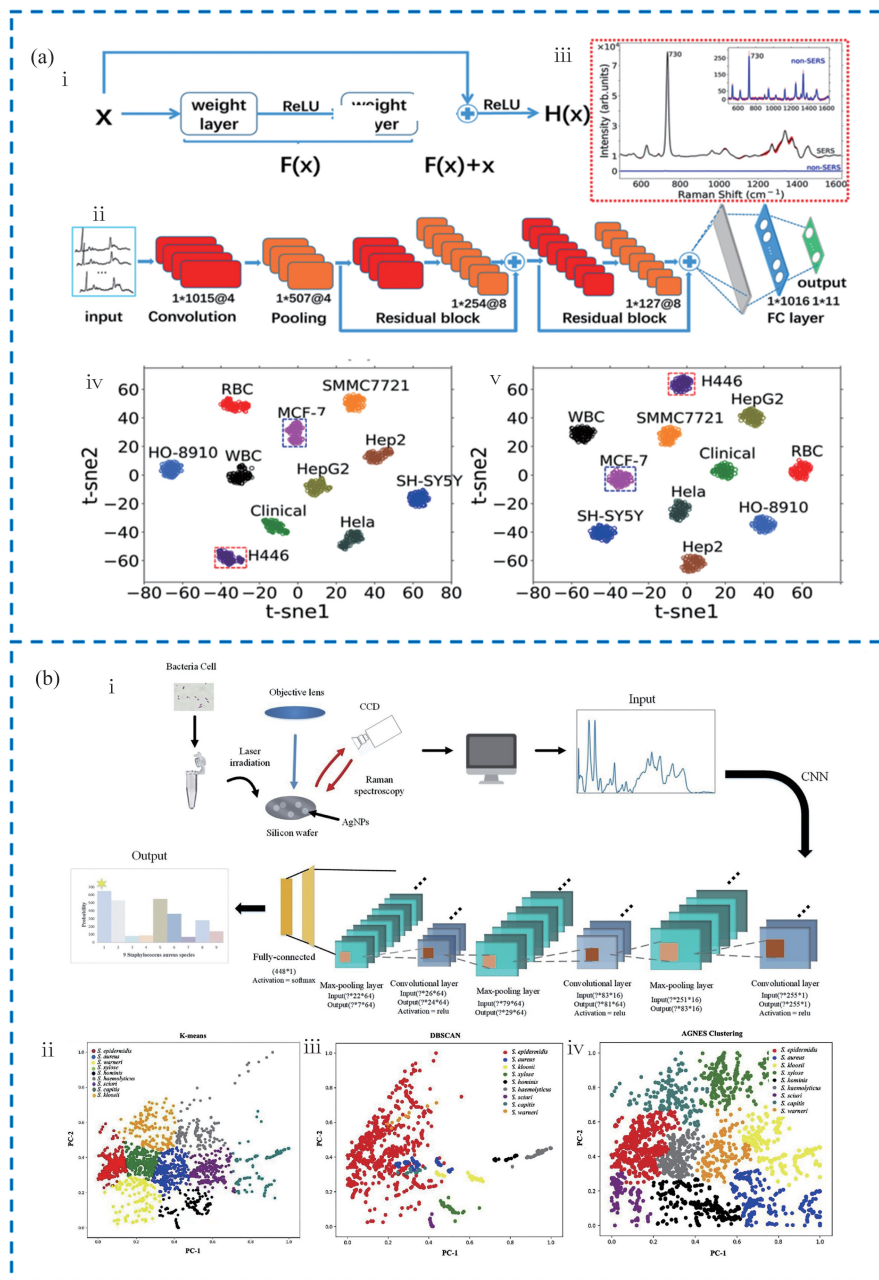


图6 SERS-ML在疾病检测中的应用:(a) SERS联合深度学习实现肿瘤细胞快速识别的流程与结果,包括残差网络(residual network, ResNet)单元结构((i)由卷积层、池化层、两个残差块及两个全连接层构成的基本网络架构,(ii)在银膜与铝板基底上获得的肿瘤细胞拉曼增强光谱,(iii)不同基底采集的光谱反映了肿瘤细胞间的分子振动差异,为模型特征提取提供依据,通过分别包含两个(iv)和5个(v)残差块的ResNet模型提取特征后的t-SNE可视化结果,显示深度特征提取在细胞分类中的有效性)<sup>[77]</sup>;(b)机器学习方法在临床分离葡萄球菌种的SERS光谱分析中的应用(包括CNN处理SERS数据的信号流示意图(i),并分别通过k-means(ii)、DBSCAN(iii)和凝聚嵌套聚类(iv)算法对9种葡萄球菌种进行无监督聚类分析,验证了各算法在细菌分类中的适用性和效果)<sup>[81]</sup>

Fig.6 Applications of SERS-ML in disease detection: (a) illustration of label-free surface-enhanced Raman spectroscopy (SERS) combined with deep learning for rapid tumor cell identification, schematic of a ResNet unit ((i) architecture of the ResNet model, consisting of convolutional layers, pooling layers, two residual blocks, and two fully connected layers, (ii) enhanced Raman spectra of tumor cells obtained using silver film and aluminum substrates, (iii) spectra collected from different substrates reflect molecular vibration differences among tumor cells, providing the basis for feature extraction by the model, t-SNE visualization of features extracted by a ResNet model with two residual blocks (iv) and five residual blocks (v) demonstrating improved feature separation for tumor cell classification)<sup>[77]</sup>; (b) application of machine learning algorithms in the SERS-based classification of *Staphylococcus* species ((i) schematic representation of the data flow in a CNN during the processing of *Staphylococcus* SERS spectra, unsupervised clustering results for nine *Staphylococcus* species using *k*-means (ii), DBSCAN (iii) and agglomerative nesting clustering (iv), highlighting the effectiveness of different algorithms in bacterial species classification)<sup>[81]</sup>

## 2 机器学习用于 SERS 数据挖掘的实例展示

除了分类和定量检测,机器学习在 SERS 中的另一重要应用是基于 SERS 光谱推测分子结构,以及根据分子结构预测其拉曼光谱。Chen 等<sup>[84]</sup>提出了一种基于机器学习的 SERS 化学空间模型,该模型利用关键的光谱-结构相关性,实现了对未经训练分子的双向光谱-结构和结构-光谱预测,其预测准确率超过 90%。如图 7 所示,该研究采用两步机器学习框架。在第一步,构建了 4 个平行的 RF 分类器,以识别分子中的特定官能团,包括醇、醛、胺和羧酸。第二步则利用 PLS 回归,通过分子  $\text{CH}_2$  剪切振动的相关性预测碳链长度。这一框架成功解析了辛烷醛的完整分子结构(图 7(ii)),验证了其在分子结构解析中的有效性。具体而言,在第一步中,4 个 RF 分类器准确识别出目标分子仅含有醛官能团,而不含醇、胺或羧酸。对 50 个单独的 SERS 光谱进行分析时,该分类器的总体预测精度达到 100%。在第二步,通过醛类分子的 PLS 回归模型,进一步精确预测了目标分子的碳链长度为  $8.1 \pm 0.1$ 。最终,结合步骤 1(官能团鉴定)和步骤 2(碳链长度预测)的信息,成功推测出该未训练分子的结构为辛烷醛,其预测结果与实际分子结构完全一致。相反,一步单 RF 模型预测未知测试分子为现有训练数据集中最接近的模拟物,但不能准确预测未训练分子的实际碳链长度(图 7(i))。基于机器学习的 SERS 分子结构预测能力,标志着分子鉴定方法从传统的光谱匹配转向光谱-结构关系建模的重大突破。然而,当目标分子同时具有多种官能团时,如何实现准确区分仍然是一个挑战。因此,未来需要进一步探索机器学习与 SERS 结合的方法,以提升多官能团分子的识别能力。

Forward predicting the chemical identity of an untrained molecule

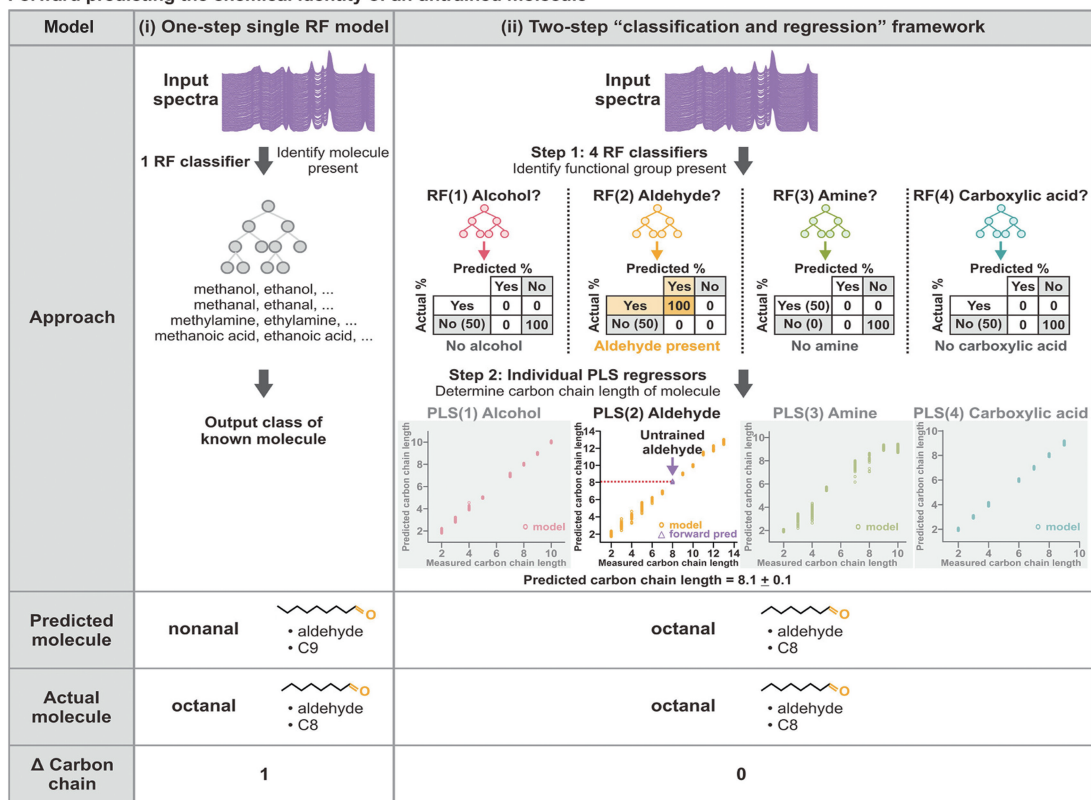


图 7 (i)一步单 RF 模型和(ii)两步监督“分类和回归”模型进行正向光谱-结构预测时的模型性能<sup>[84]</sup>

Fig.7 Model performance in forward spectrum-to-structure prediction using (i) one-step single random forest (RF) model and (ii) a two-step supervised "classification and regression" framework<sup>[84]</sup>

不同类型的纳米颗粒或纳米结构金属 SERS 基底在提供 SERS 光谱时可能存在显著差异。这些不一致性甚至引发了一种假设,即 SERS 光谱的应用可能需要特定于基底的 SERS 光谱库。针对这一问题, Ju 等<sup>[85]</sup>提出了一种基于 ML 的方法,利用标准拉曼光谱库从 SERS 光谱中识别未知化学物质。该方法通过计算未知 SERS“查询”光谱与拉曼光谱库中各光谱之间的相似性分数,并基于分数的相对大小识别未知化学物质(图 8(a))。这一过程被称为“光谱识别”,类似于人脸识别。研究者利用特征峰提取技术,以适应数据中常见的峰移(通常为  $1 \sim 8 \text{ cm}^{-1}$ ),并提出了一种相似性指标——特征峰相似性(characteristic peak similarity,

CaPSim)。在计算相似性分数时, CaPSim 仅关注特征峰位置, 因此对 SERS 光谱中的大多数干扰因素具有较强的鲁棒性。基于 CaPSim, 研究者开发了一种光谱识别算法, 该算法能够高精度地将未知分析物的 SERS 光谱与标准拉曼光谱库进行匹配(图 8(b))。这一方法为整合不同研究小组获得的 SERS 基底特定光谱提供了更严格的标准, 并有助于将 SERS 技术更广泛地应用于化学分析和传感领域。

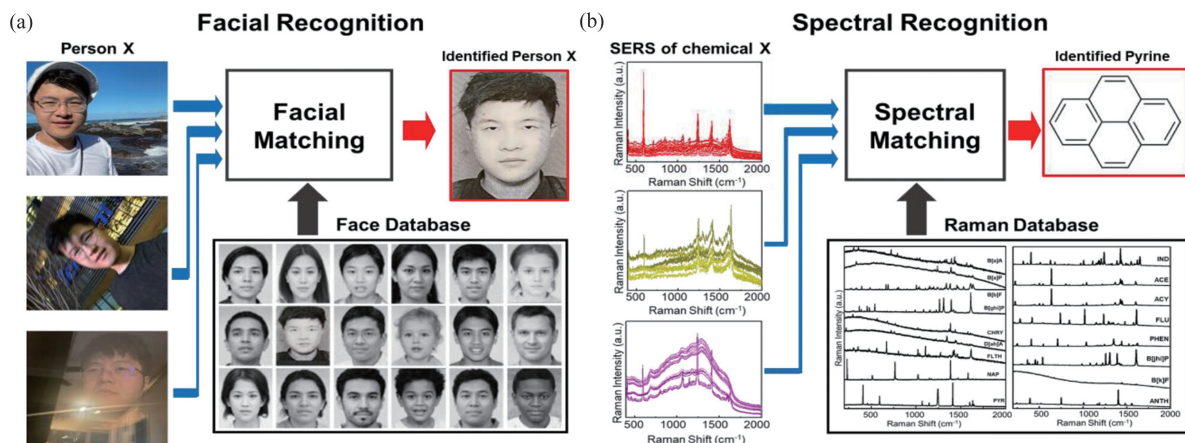


图 8 面部和光谱识别的示意图: (a) 用于识别人  $X$  的提取面部特征的面部匹配算法和用于面部匹配的面部数据库<sup>[85]</sup>; (b) 同样使用其 SERS 光谱识别未知分子  $X$  涉及的提取光谱特征的光谱匹配算法和拉曼数据库<sup>[85]</sup>

Fig.8 Schematic diagram of facial and spectral recognition: (a) facial recognition of individual  $X$  involves a facial matching algorithm that extracts facial features and compares them with a facial database<sup>[85]</sup>; (b) similarly, identifying an unknown molecule  $X$  using its SERS spectrum involves a spectral matching algorithm that extracts spectral features and compares them with a Raman database<sup>[85]</sup>

在医学检测领域, SERS 与 ML 的结合展现出广阔的应用前景。特别是, ML 已被用于分析 DNA 和 RNA 的无标记 SERS 光谱, 以区分疾病相关的遗传生物标志物。例如, Chheda 等<sup>[86]</sup> 利用精胺包覆的银纳米颗粒 (AgNPs) 作为带正电的 SERS 受体, 以吸引带负电的单链 DNA 和 RNA 磷酸骨架, 从而增强其 SERS 信号。研究者首先分析了一系列具有序列修饰 (如碱基取代、添加及缺失) 的样本, 随后进一步分析了前列腺癌生物标志物 miR-21 及其突变变体。为了解释所得 SERS 光谱, 研究者开发了一种基于功能数据分析 (functional data analysis, FDA) 的框架, 用于检测突变 DNA 和 RNA 寡核苷酸。该框架包含 4 个关键步骤: 光谱采集与增强; 光谱配对; 基于 GP 的建模; 基于 GP 的假设检验 (图 9)。

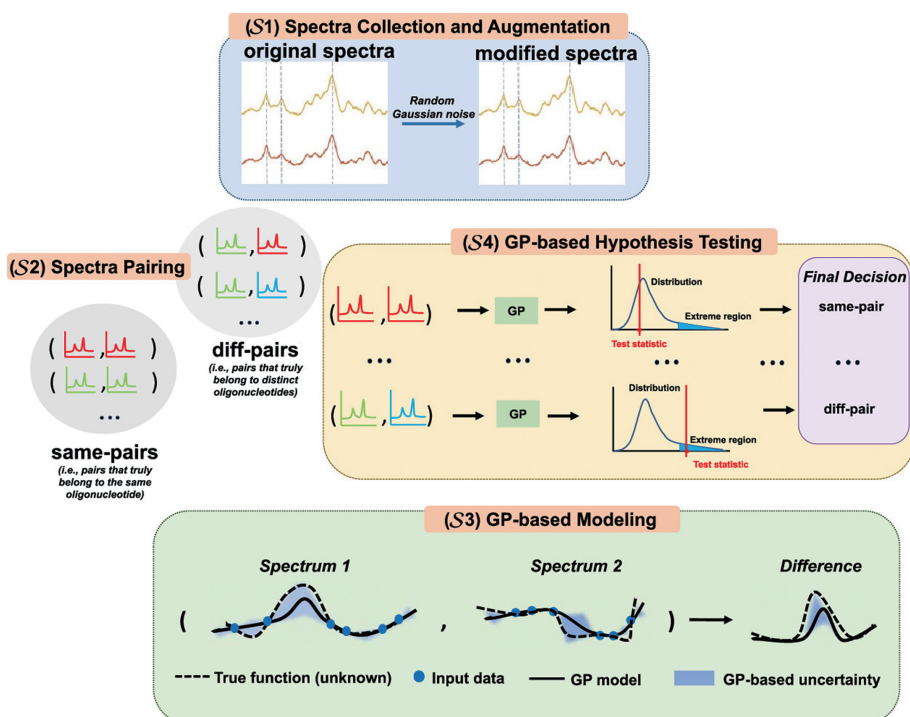


图 9 SERS 结合 ML 区分 DNA 和 RNA 过程图<sup>[86]</sup>

Fig.9 Process diagram of differentiation of DNA and RNA using SERS combined with machine learning<sup>[86]</sup>

实验结果表明,该方法能够准确区分不同寡核苷酸的 SERS 光谱,并在准确性、灵敏度和特异性等多个指标上优于传统数据驱动方法。研究者认为,SERS 与 ML 的结合可为疾病诊断提供强有力的工具,并具有潜在的临床应用价值。

### 3 总结与展望

本综述概述了表面增强拉曼光谱(SERS)与机器学习(ML)在生物医药检测中的应用。我们首先介绍了 SERS 与 ML 结合的基本步骤,包括样品预处理、数据预处理、模型训练、验证及评价。随后,总结了 ML 在 SERS 生物医药检测中的实际应用,尤其是在分类和定量分析方面的优势。此外,还讨论了 SERS 结合 ML 在预测分子结构、建立拉曼识别库、以及区分 DNA 和 RNA 等方面的潜力。

SERS 与 ML 的融合为生物医药领域,尤其是疾病早期诊断与分类提供了一种高效、灵敏、非侵入且具有高度自动化潜力的分析工具。与传统临床检测方法相比,SERS-ML 系统不仅在操作简便性、检测速度和成本控制方面具有显著优势,还展现出在微量样本识别与复杂生物体系中信息提取的卓越能力。尤其是在处理复杂、高维拉曼数据集时,机器学习算法可有效提取潜在特征,挖掘样本之间的本质差异,从而实现精准分类与定量分析。例如,在阿片类药物(如芬太尼)早期成瘾筛查中,通过采集血清样本的 SERS 指纹图谱并结合机器学习算法,可实现对滥用风险的快速识别与分型判断;在感染性疾病早期诊断方面,SERS-ML 系统能够对病原微生物特异性拉曼信号进行快速捕捉与分类,显著缩短检测时间、提升检测灵敏度;此外,SERS 结合 ML 还可应用于液体活检中肿瘤标志物的无创检测,如通过识别外泌体或循环 miRNA 的光谱特征,实现对乳腺癌、肺癌等疾病的早期预警与动态监测。随着便携式拉曼设备与嵌入式 AI 芯片的发展,该技术也有望在床旁检测设备或可穿戴终端中实现集成应用,推动 SERS-ML 向真实世界的临床和公共健康场景加速转化。然而,尽管该技术路径展现出巨大前景,其在实际应用中仍面临一系列关键挑战,亟待系统性研究和技术突破:

(1) 数据质量与模型泛化能力问题。当前 SERS 结合 ML 的研究在数据获取方面存在显著瓶颈,样本量有限且数据质量不一,这不仅影响模型的训练效果,还限制了其在实际场景中的泛化能力。尤其是数据来源的异质性与批间差异可能导致模型过拟合,在跨平台、跨样本类型时表现不稳定。此外,特征提取与降维方法的选择对分类结果具有关键影响。例如,主成分分析(PCA)作为 SERS 中常用的降维工具,尽管在降噪和特征压缩方面表现良好,但其在结合支持向量机(SVM)等分类器时的最优适配性仍需进一步探讨,尤其是在处理非线性、高噪声数据时的鲁棒性有待提高。

(2) 模型可解释性与疾病机制的衔接不足。目前,大多数 SERS-ML 研究聚焦于基于光谱模式的分类与回归任务,但往往忽略了分类背后的生物机制解释。这种“黑箱式”分析方法虽能提供较高准确率,但缺乏对疾病发生机制与分子变化过程的深入解析,限制了其在临床可应用性与科研深度上的进一步推进。因此,构建具备生物学可解释性的机器学习框架,如引入基于特征重要性评估的模型(如 SHAP、LIME)或整合多组学信息,将是未来提升 SERS 在疾病诊断中可信度的关键方向。

(3) 模型选择的适配性与场景依赖性问题。不同类型的 ML 算法(如决策树、神经网络、随机森林、卷积神经网络等)对数据维度、样本量及任务目标具有不同的适配性。目前的研究多采用经验法则进行模型构建,缺乏系统性评估与模型-任务匹配策略。为提升模型效果与可迁移性,应建立标准化评估指标与算法选择指南,明确“何种算法适用于何种 SERS 任务”的理论依据和实践流程。

(4) SERS 基底的可机器学习化设计不足。在当前研究中,ML 主要被用于数据分析层面,而对 SERS 基底的材料设计和结构调控参与有限。事实上,基底性能(如 EF 均一性、信号稳定性、表面化学环境等)直接影响光谱数据的质量与分析可行性。因此,未来有必要从“数据可分析性”的角度出发,构建面向 ML 优化的 SERS 基底设计理念,如开发信号重现性高、响应机制明确、适用于高通量测量的智能基底,真正实现“端到端”的分析系统构建。SERS 与机器学习的协同融合尚处于发展阶段,未来应在数据质量控制、模型可解释性、算法适配策略及智能基底设计等方面进行深度融合与系统化发展,以推动该技术从实验室走向临床应用的实际转化。

尽管 SERS 结合 ML 的研究面临挑战,但其在生物分析领域的应用前景依然广阔。目前,SERS 已被证

明是一种实用且可靠的生物分析工具,但由于生物体系与纳米材料之间的复杂相互作用,其在光谱分析和解读方面仍需谨慎。未来研究应重点关注改进数据质量与模型优化,提高训练数据的多样性,以增强 ML 模型的泛化能力,同时探索适用于 SERS 分析的最优 ML 算法,以进一步提高数据解读的准确性和可靠性。此外,为了提升 SERS 在疾病诊断中的临床可用性,可以结合最大似然估计等统计模型,对 SERS 光谱进行反卷积,以提取关键生物标志物信息,从而提高疾病检测的准确性和灵敏度。然而,在 SERS 分析中使用 ML 时,应始终对数据进行严格验证,而非单纯依赖算法输出。总之,SERS 结合 ML 的应用极大提升了生物标志物检测和疾病诊断的准确性,为未来生物医学分析提供了新的机遇。尽管 ML 算法的引入增强了 SERS 数据分析的智能化水平,但并不能完全替代传统光谱分析方法,仍需在确保数据质量的基础上谨慎使用,并结合实验验证。未来,随着 ML 技术的发展及大规模数据集的积累,SERS 在精准医学和综合组学分析领域的应用将进一步拓展,为下一代智能诊断工具的开发提供有力支撑。

#### 参考文献:

- [1] CHEN Y F, AN Q, TENG K X, et al. Application of SERS in *in-vitro* biomedical detection[J]. *Chemistry-An Asian Journal*, 2023, 18(4): e202201194.
- [2] YI J, YOU E M, HU R, et al. Surface-enhanced Raman spectroscopy: a half-century historical perspective[J]. *Chemical Society Reviews*, 2025, 54(3): 1453-1551.
- [3] FLEISCHMANN M, HENDRA P J, MCQUILLAN A J. Raman spectra of pyridine adsorbed at a silver electrode[J]. *Chemical Physics Letters*, 1974, 26(2): 163-166.
- [4] JEANMAIRE D L, VAN DUYN R P. Surface Raman spectroelectrochemistry part I. Heterocyclic, aromatic, and aliphatic amines adsorbed on the anodized silver electrode[J]. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, 1977, 84(1): 1-20.
- [5] ALBRECHT M G, CREIGHTON J A. Anomalously intense Raman spectra of pyridine at a silver electrode[J]. *Journal of the American Chemical Society*, 1977, 99(15): 5215-5217.
- [6] LEE P C, MEISEL D. Adsorption and surface-enhanced Raman of dyes on silver and gold sols[J]. *The Journal of Physical Chemistry*, 1982, 86(17): 3391-3395.
- [7] KNEIPP K, WANG Y, KNEIPP H, et al. Single molecule detection using surface-enhanced Raman scattering (SERS)[J]. *Physical Review Letters*, 1997, 78(9): 1667-1670.
- [8] NIE S, EMORY S R. Probing single molecules and single nanoparticles by surface-enhanced Raman scattering[J]. *Science*, 1997, 275(5303): 1102-1106.
- [9] BENZ F, CHIKKARADDY R, SALMON A, et al. SERS of individual nanoparticles on a mirror: size does matter, but so does shape[J]. *The Journal of Physical Chemistry Letters*, 2016, 7(12): 2264-2269.
- [10] TIAN Y, FANG G, WU F X, et al. Raman spectroscopic technologies for chiral discrimination: current status and new frontiers[J]. *Coordination Chemistry Reviews*, 2025, 526: 216375.
- [11] LIN S, DONG M Y, LI C, et al. Machine learning-assisted ultrasensitive SERS immunoassays across wide concentration ranges toward clinical ovarian cancer diagnosis[J]. *Advanced Functional Materials*, 2025, e09813.
- [12] HAN X X, JI W, ZHAO B, et al. Semiconductor-enhanced Raman scattering: active nanomaterials and applications[J]. *Nanoscale*, 2017, 9(15): 4847-4861.
- [13] HAN X X, KÖHLER C, KOZUCH J, et al. Potential-dependent surface-enhanced resonance Raman spectroscopy at nanostructured TiO<sub>2</sub>: a case study on cytochrome b5[J]. *Small*, 2013, 9(24): 4175-4181.
- [14] LAI H S, XU F G, ZHANG Y, et al. Recent progress on graphene-based substrates for surface-enhanced Raman scattering applications[J]. *Journal of Materials Chemistry B*, 2018, 6(24): 4008-4028.
- [15] LEE Y, KIM H, LEE J, et al. Enhanced Raman scattering of rhodamine 6G films on two-dimensional transition metal dichalcogenides correlated to photoinduced charge transfer[J]. *Chemistry of Materials*, 2016, 28(1): 180-187.
- [16] YILMAZ M, BABUR E, OZDEMIR M, et al. Nanostructured organic semiconductor films for molecular detection with surface-enhanced Raman spectroscopy[J]. *Nature Materials*, 2017, 16(9): 918-924.
- [17] WANG X T, SHI W X, WANG S X, et al. Two-dimensional amorphous TiO<sub>2</sub> nanosheets enabling high-efficiency photoinduced charge transfer for excellent SERS activity[J]. *Journal of the American Chemical Society*, 2019, 141(14): 5856-5862.

- [18] DING S Y, YOU E M, TIAN Z Q, et al. Electromagnetic theories of surface-enhanced Raman spectroscopy[J]. *Chemical Society Reviews*, 2017, 46(13):4042-4076.
- [19] PARK W H, KIM Z H. Charge transfer enhancement in the SERS of a single molecule[J]. *Nano Letters*, 2010, 10(10):4040-4048.
- [20] VALLEY N, GREENELTCH N, VAN DUYN R P, et al. A look at the origin and magnitude of the chemical contribution to the enhancement mechanism of surface-enhanced Raman spectroscopy (SERS): theory and experiment[J]. *The Journal of Physical Chemistry Letters*, 2013, 4(16):2599-2604.
- [21] KNEIPP J, SEIFERT S, GÄRBER F. SERS microscopy as a tool for comprehensive biochemical characterization in complex samples[J]. *Chemical Society Reviews*, 2024, 53(15):7641-7656.
- [22] CUTSHAW G, UTHAMAN S, HASSAN N, et al. The emerging role of Raman spectroscopy as an omics approach for metabolic profiling and biomarker detection toward precision medicine[J]. *Chemical Reviews*, 2023, 123(13):8297-8346.
- [23] ZHOU H, XU L G, REN Z H, et al. Machine learning-augmented surface-enhanced spectroscopy toward next-generation molecular diagnostics[J]. *Nanoscale Advances*, 2023, 5(3):538-570.
- [24] CAO D W, LIN H C, LIU Z Y, et al. Serum-based surface-enhanced Raman spectroscopy combined with PCA-RCKNCN for rapid and accurate identification of lung cancer[J]. *Analytica Chimica Acta*, 2022, 1236:340574.
- [25] LOO J, CAI C X, CHOONG J, et al. Deep learning-based classification and segmentation of retinal cavitations on optical coherence tomography images of macular telangiectasia type 2[J]. *British Journal of Ophthalmology*, 2022, 106(3):396-402.
- [26] CHEN P C, LIU Y, PENG L. How to develop machine learning models for healthcare[J]. *Nature Materials*, 2019, 18(5):410-414.
- [27] LI H M, WANG Q, TANG J, et al. Establishment of a reliable scheme for obtaining highly stable SERS signal of biological serum[J]. *Biosensors and Bioelectronics*, 2021, 189:113315.
- [28] WANG Y P, YU C W, JI H Y, et al. Label-free therapeutic drug monitoring in human serum by the 3-step surface enhanced Raman spectroscopy and multivariate analysis[J]. *Chemical Engineering Journal*, 2023, 452:139588.
- [29] LU Y, WANG J Y, BI X Y, et al. Non-invasive and rapid diagnosis of low-grade bladder cancer *via* SERSomes of urine[J]. *Nanoscale*, 2025, 17(12):7303-7312.
- [30] DONG Y L, HU J Y, JIN J L, et al. Advances in machine learning-assisted SERS sensing towards food safety and biomedical analysis[J]. *TrAC Trends in Analytical Chemistry*, 2024, 180:117974.
- [31] DIJKSTRA R J, SCHEENEN W J J M, DAM N, et al. Monitoring neurotransmitter release using surface-enhanced Raman spectroscopy[J]. *Journal of Neuroscience Methods*, 2007, 159(1):43-50.
- [32] SHEN J H, LI M, LI Z F, et al. Single convolutional neural network model for multiple preprocessing of Raman spectra[J]. *Vibrational Spectroscopy*, 2022, 121:103391.
- [33] THRIFT W J, RAGAN R. Quantification of analyte concentration in the single molecule regime using convolutional neural networks[J]. *Analytical Chemistry*, 2019, 91(21):13337-13342.
- [34] RALBOVSKY N M, LEDNEV I K. Towards development of a novel universal medical diagnostic method: Raman spectroscopy and machine learning[J]. *Chemical Society Reviews*, 2020, 49(20):7428-7453.
- [35] YANG L, SHAMI A. On hyperparameter optimization of machine learning algorithms: theory and practice[J]. *Neurocomputing*, 2020, 415:295-316.
- [36] SHIN H, OH S, HONG S, et al. Early-stage lung cancer diagnosis by deep learning-based spectroscopic analysis of circulating exosomes[J]. *ACS Nano*, 2020, 14(5):5435-5444.
- [37] FENG J Z, WANG Y, PENG J, et al. Comparison between logistic regression and machine learning algorithms on survival prediction of traumatic brain injuries[J]. *Journal of Critical Care*, 2019, 54:110-116.
- [38] DING Y, SUN Y, LIU C, et al. SERS-based biosensors combined with machine learning for medical application[J]. *ChemistryOpen*, 2023, 12(1):e202200192
- [39] YADAV A, NAIK R, GUPTA E, et al. Single-shot, receptor-free, rapid detection and classification of five respiratory viruses by machine learning integrated SERS sensing platform[J]. *Biosensors and Bioelectronics*, 2025, 279:117394.
- [40] HAN S, PARK J, MOON S, et al. Label-free and liquid state SERS detection of multi-scaled bioanalytes via light-induced pinpoint colloidal assembly[J]. *Biosensors and Bioelectronics*, 2024, 264:116663.
- [41] JANČI T, VALINGER D, GAJDOŠ KLJUSURIĆ J, et al. Determination of histamine in fish by surface enhanced Raman spectroscopy using silver colloid SERS substrates[J]. *Food Chemistry*, 2017, 224:48-54.
- [42] ZHU J J, JIANG X, RONG Y W, et al. Label-free detection of trace level zearalenone in corn oil by surface-enhanced Raman

- spectroscopy (SERS) coupled with deep learning models[J]. *Food Chemistry*, 2023, 414:135705.
- [43] LEONG S X, TAN E X, HAN X M, et al. Surface-enhanced Raman scattering-based surface chemotaxonomy: combining bacteria extracellular matrices and machine learning for rapid and universal species identification[J]. *ACS Nano*, 2023, 17(22):23132-23143.
- [44] ZHAO J L, CHEN J F, TANG J, et al. Artificial intelligence assisted label-free surface-enhanced Raman scattering detection of early-stage cancer-derived exosomes based on g-C<sub>3</sub>N<sub>4</sub>/Ag hybrid substrate prepared by electro-synthesis[J]. *Chemical Engineering Journal*, 2024, 498:155526.
- [45] MATSCHULAT A, DRESCHER D, KNEIPP J. Surface-enhanced Raman scattering hybrid nanoprobe multiplexing and imaging in biological systems[J]. *ACS Nano*, 2010, 4(6):3259-3269.
- [46] LIN L L, ALVAREZ-PUEBLA R, LIZ-MARZÁN L M, et al. Surface-enhanced Raman spectroscopy for biomedical applications: recent advances and future challenges[J]. *ACS Applied Materials & Interfaces*, 2025, 17(11):16287-16379.
- [47] SHIN H, JEONG H, PARK J, et al. Correlation between cancerous exosomes and protein markers based on surface-enhanced Raman spectroscopy (SERS) and principal component analysis (PCA)[J]. *ACS Sensors*, 2018, 3(12):2637-2643.
- [48] DING Z X, WANG C, SONG X, et al. Strong  $\pi$ -metal interaction enables liquid interfacial nanoarray-molecule co-assembly for Raman sensing of ultratrace fentanyl doped in heroin, ketamine, morphine, and real urine[J]. *ACS Applied Materials & Interfaces*, 2023, 15(9):12570-12579.
- [49] LIM J Y, NAM J S, SHIN H, et al. Identification of newly emerging influenza viruses by detecting the virally infected cells based on surface enhanced Raman spectroscopy and principal component analysis[J]. *Analytical Chemistry*, 2019, 91(9):5677-5684.
- [50] MARTIN F L, KELLY J G, LLABJANI V, et al. Distinguishing cell types or populations based on the computational analysis of their infrared spectra[J]. *Nature Protocols*, 2010, 5(11):1748-1760.
- [51] LEONG S X, KOH L K, KOH C S L, et al. In situ differentiation of multiplex noncovalent interactions using SERS and chemometrics[J]. *ACS Applied Materials & Interfaces*, 2020, 12(29):33421-33427.
- [52] NGUYEN C Q, THRIFT W J, BHATTACHARJEE A, et al. Longitudinal monitoring of biofilm formation via robust surface-enhanced Raman scattering quantification of pseudomonas aeruginosa-produced metabolites[J]. *ACS Applied Materials & Interfaces*, 2018, 10(15):12364-12373.
- [53] MEUNIER C J, MCCARTY G S, SOMBERS L A. Drift subtraction for fast-scan cyclic voltammetry using double-waveform partial-least-squares regression[J]. *Analytical Chemistry*, 2019, 91(11):7319-7327.
- [54] ACOSTA C M, OGOSHI E, SOUZA J A, et al. Machine learning study of the magnetic ordering in 2D materials[J]. *ACS Applied Materials & Interfaces*, 2022, 14(7):9418-9432.
- [55] LEONG S X, KOH C S L, SIM H Y F, et al. Enantiospecific molecular fingerprinting using potential-modulated surface-enhanced Raman scattering to achieve label-free chiral differentiation[J]. *ACS Nano*, 2021, 15(1):1817-1825.
- [56] ALOBAIDI M, MALIK K M, SABRA S. Linked open data-based framework for automatic biomedical ontology generation[J]. *BMC Bioinformatics*, 2018, 19(1):319.
- [57] DIAO X K, QI G H, LI X L, et al. Label-free exosomal SERS detection assisted by machine learning for accurately discriminating cell cycle stages and revealing the molecular mechanisms during the mitotic process[J]. *Analytical Chemistry*, 2025, 97(9):5093-5101.
- [58] KIM W H, LEE S, JEON M J, et al. Rapid and differential diagnosis of sepsis stages using an advanced 3D plasmonic bimetallic alloy nanoarchitecture-based SERS biosensor combined with machine learning for multiple analyte identification[J]. *Advanced Science*, 2025, 2414688.
- [59] SHU W K, ZHANG M J, ZHANG C Q, et al. An alloy platform of dual-fingerprints for high-performance stroke screening[J]. *Advanced Functional Materials*, 2023, 33(5):2210267.
- [60] DAS S, SAXENA K, TINGUELY J C, et al. SERS nanowire chip and machine learning-enabled classification of wild-type and antibiotic-resistant bacteria at species and strain levels[J]. *ACS Applied Materials & Interfaces*, 2023, 15(20):24047-24058.
- [61] TAMTAJI M, GUO X Y, TYAGI A, et al. Machine learning-aided design of gold core-shell nanocatalysts toward enhanced and selective photooxygenation[J]. *ACS Applied Materials & Interfaces*, 2022, 14(41):46471-46480.
- [62] CHEN Y F, LI H T, SHI J, et al. Diagnosis of early opioids addiction by label-free serum SERS fingerprints with machine learning[J]. *Chemical Engineering Journal*, 2025, 505:159412.
- [63] JIANG H Y, UNIVERSITY C S, ZHANG Y B, et al. Comprehensive serum analysis via an AI-assisted magnetically driven SERS platform for the diagnosis and etiological differentiation of childhood epilepsy[J]. *ACS Applied Materials & Interfaces*,

- 2025, 17(8):11731-11741.
- [64] LEONG S X, LEONG Y X, TAN E X, et al. Noninvasive and point-of-care surface-enhanced Raman scattering (SERS)-based breathalyzer for mass screening of coronavirus disease 2019 (COVID-19) under 5 min[J]. *ACS Nano*, 2022, 16(2):2629-2639.
- [65] JONES T, ZHOU D, LIU J, et al. Quantitative multiplexing of uric acid and creatinine using polydisperse plasmonic nanoparticles enabled by electrochemical-SERS and machine learning[J]. *Journal of Materials Chemistry B*, 2024, 12(41):10563-10572.
- [66] GARG A, HAWKS S, PAN J, et al. Machine learning-driven SERS fingerprinting of disintegrated viral components for rapid detection of SARS-CoV-2 in environmental dust[J]. *Biosensors and Bioelectronics*, 2024, 247:115946.
- [67] WAN Y, WEI Q, SUN H, et al. Machine learning assisted biomimetic flexible SERS sensor from seashells for pesticide classification and concentration prediction[J]. *Chemical Engineering Journal*, 2025, 507:160813.
- [68] KHONDAKAR K R, MAZUMDAR H, DAS S, et al. Machine learning (ML)-assisted surface-enhanced Raman spectroscopy (SERS) technologies for sustainable health[J]. *Advances in Colloid and Interface Science*, 2025, 344:103594.
- [69] TAN E X, CHEN J R T, PANG D W C, et al. Transfer learning-assisted SERS: predicting molecular identity and concentration in mixtures using pure compound spectra[J]. *Angewandte Chemie International Edition*, 2025, e202508717.
- [70] ZHENG P, WU L T, LEE M K H, et al. Deep learning-powered colloidal digital SERS for precise monitoring of cell culture media[J]. *Nano Letters*, 2025, 25(15):6284-6291.
- [71] SUN X R, XUAN L R, LIU C L, et al. Quantitative analysis of deltamethrin residues in water using surface-enhanced Raman spectroscopy[J]. *Spectrochimica Acta Part A, Molecular and Biomolecular Spectroscopy*, 2025, 343:126545.
- [72] TANG J W, YUAN Q, ZHANG L, et al. Application of machine learning-assisted surface-enhanced Raman spectroscopy in medical laboratories: principles, opportunities, and challenges[J]. *TrAC Trends in Analytical Chemistry*, 2025, 184:118135.
- [73] MAHMOUD A Y F, TEIXEIRA A, ARANDA M, et al. Will data analytics revolution finally bring SERS to the clinic? [J]. *TrAC Trends in Analytical Chemistry*, 2023, 169:117311.
- [74] HERNÁNDEZ-VIDALES K, GUEVARA E, OLIVARES-ILLANA V, et al. Characterization of wild-type and mutant p53 protein by Raman spectroscopy and multivariate methods[J]. *Journal of Raman Spectroscopy*, 2019, 50(10):1388-1394.
- [75] WU X X, XIA Y Z, HUANG Y J, et al. Improved SERS-active nanoparticles with various shapes for CTC detection without enrichment process with supersensitivity and high specificity[J]. *ACS Applied Materials & Interfaces*, 2016, 8(31):19928-19938.
- [76] PANG Y F, WANG C W, XIAO R, et al. Dual-selective and dual-enhanced SERS nanoprobe strategy for circulating hepatocellular carcinoma cells detection[J]. *Chemistry—A European Journal*, 2018, 24(27):7060-7067.
- [77] FANG X L, ZENG Q Y, YAN X L, et al. Fast discrimination of tumor and blood cells by label-free surface-enhanced Raman scattering spectra and deep learning[J]. *Journal of Applied Physics*, 2021, 129(12):123103.
- [78] RIPPA M, CASTAGNA R, PANNICO M, et al. Octupolar metastructures for a highly sensitive, rapid, and reproducible phage-based detection of bacterial pathogens by surface-enhanced Raman scattering[J]. *ACS Sensors*, 2017, 2(7):947-954.
- [79] RODRÍGUEZ-LORENZO L, GARRIDO-MAESTU A, BHUNIA A K, et al. Gold nanostars for the detection of foodborne pathogens via surface-enhanced Raman scattering combined with microfluidics[J]. *ACS Applied Nano Materials*, 2019, 2(10):6081-6086.
- [80] LI H, CAO Y B, LU F. Differentiation of different antifungals with various mechanisms using dynamic surface-enhanced Raman spectroscopy combined with machine learning[J]. *Journal of Innovative Optical Health Sciences*, 2021, 14(4):2141002.
- [81] HU X, ZENG Q Z, XIAO J, et al. Herpes simplex virus 1 induces microglia gasdermin D-dependent pyroptosis through activating the NLR family pyrin domain containing 3 inflammasome[J]. *Frontiers in Microbiology*, 2022, 13:838808.
- [82] FREITAS C, ELEUTÉRIO J, SOARES G, et al. Towards rapid and low-cost stroke detection using SERS and machine learning[J]. *Biosensors*, 2025, 15(3):136.
- [83] KAO Y C, HAN X M, LEE Y H, et al. Multiplex surface-enhanced Raman scattering identification and quantification of urine metabolites in patient samples within 30 min[J]. *ACS Nano*, 2020, 14(2):2542-2552.
- [84] CHEN J R T, TAN E X, TANG J X, et al. Machine learning-based SERS chemical space for two-way prediction of structures and spectra of untrained molecules[J]. *Journal of the American Chemical Society*, 2025, 147(8):6654-6664.
- [85] JU Y, NEUMANN O, BAJOMO M, et al. Identifying surface-enhanced Raman spectra with a Raman library using machine learning[J]. *ACS Nano*, 2023, 17(21):21251-21261.
- [86] CHHEDA J, FANG Y T, DERIU C, et al. Discrimination of genetic biomarkers of disease through machine-learning-based hypothesis testing of direct SERS spectra of DNA and RNA[J]. *ACS Sensors*, 2024, 9(5):2488-2498.