

RamanNexus：拉曼光谱定性定量分析软件的设计与实现

钟晨锐¹, 田亚平^{2*}

1. 江苏师范大学物理与电子工程学院, 江苏徐州, 221116

2. 江苏师范大学文学院, 江苏徐州, 221132

摘要 | 针对拉曼光谱分析中定性识别与定量预测流程分散、模型接入方式不统一及结果管理衔接不足的问题, 基于MATLAB R2025b设计并实现RamanNexus桌面分析软件。系统采用分层架构, 将光谱导入、前处理、分类识别、浓度预测和结果导出集成为连续分析流程, 并通过统一模型接口支持网络模型与核极限学习机模型。以分类模型保存的测试记录评价识别性能, 以GA-CNN-LSTM-Attention模型的测试结果评价定量性能, 并采用直接模型调用对照检验软件推理的一致性。结果显示, 360条分类测试记录的Top1准确率为99.72%, 宏平均F1为0.9972; 蒽、荧蒽、芴和芘定量模型的決定系数分别为0.9884、0.9585、0.9988和0.9964。该软件实现了智能光谱模型的统一接入与交互式应用, 可为实验室批量光谱分析及结果管理提供工程支持。

关键词 | 拉曼光谱; 智能分析软件; MATLAB

Copyright © 2026 by the author(s).

This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/).

<https://creativecommons.org/licenses/by-nc/4.0/>



1 引言

拉曼光谱具有样品前处理要求低、检测速度快、可反映分子振动特征等优点, 已在环境监测^[1]、材料识别^[2]、食品安全^[3]、医学生物分析^[4,5]等场景中得到广泛应用。随着机器学习、深度学习与化学计量学方法的发展, 拉曼光谱分析逐渐从单条谱图判读演化为面向批量样本的数据处理与模型推理流程^[6,7], 其中通常涉及光谱导入、位移匹

配、预处理^[8]、分类识别、定量预测^[9]、结果解释与结果归档等多个环节。为了保证分析过程具有可重复性和可操作性^[10], 面向实际应用的软件化实现已经成为拉曼分析研究的重要支撑方向^[11,12,13]。

已有Python光谱分析工具支持预处理、建模与流程集成^[11,13]; 科学计算实践强调数据组织、过程记录与可重复执行^[14]。当定性定量模型分散存储、处理步骤缺少统一调度时, 需要在应用层建立连续的计算与结果管理流程。拉曼光谱数据协调研究关注测量差异、数据处理及格式一致性^[15,16],

作者简介: 钟晨锐, 江苏师范大学物理与电子工程学院光电信息科学与工程专业本科生, 研究方向为深度学习算法与光谱检测技术; 田亚平 (*通讯作者), 中国矿业大学信息与控制工程学院博士研究生, 江苏师范大学文学院副教授, 研究方向为表面增强拉曼散射传感检测与人工智能算法。

文章引用: 钟晨, 田亚平. RamanNexus: 拉曼光谱定性定量分析软件的设计与实现[J]. 计算机与智能教育技术, 2026, 1(3): 89-101.

<https://doi.org/10.68070/2108b068>

为系统的数据定义与接口组织提供参考。因此，如何将既有模型组织为可重复执行的连续分析流程，是光谱软件工程的重要问题。

RamanSPy 提供 Python 分析接口^[11]，PyRamanGUI采用图形交互^[12]，PyFasma支持模块化处理与多变量分析^[13]，为光谱分析的软件化提供了参考。本文面向既有MATLAB模型，研究异构分类与回归模型接入、批量推理及样本级结果关联，构建贯穿数据处理与结果管理的桌面分析流程。

光谱分析软件兼具科学计算与应用系统的特征。输入列维度对应波数位置，处理参数影响特征分布，输出需要保持样本关系与浓度量纲。因此，系统既需满足操作流程要求，也需维持模型变换参数与计算路径的一致性。本文分别评价模型预测性能和软件数值一致性，以区分任务误差与工程实现差异。

本文基于分层架构设计RamanNexus桌面系统，集成数据导入、前处理、批量分类、浓度预测及结果导出，并通过分类结果驱动回归模型选择。定量接口兼容网络模型和核极限学习机（kernel extreme learning machine, KELM）模型^[17]。研究重点为既有模型的统一接入、样本结果关联及调用一致性验证，以支持实验室批量光谱分析。

2 系统需求分析与总体架构

2.1 系统需求分析

系统需求围绕数据处理、模型接入、流程组织及部署适配四个方面展开。已有拉曼预处理软件与方法综述为处理流程提供参考^[18,19]，数据处理层

据此集成批量光谱矩阵导入、基线校正、去尖峰和平滑等操作。模型接入层需要识别分类与回归模型的结构差异，统一管理网络、变换参数与输出定义。现有分析工具展示了处理与分析步骤的组织方式^[11,13]；本文在流程组织层连接数据准备、前处理、识别预测及结果归档，使不同阶段的样本关系保持一致。部署适配层需要处理资源定位、目录权限及输出路径变化，以支持从研究环境向独立应用环境的迁移。上述需求共同构成系统架构及接口设计的依据。

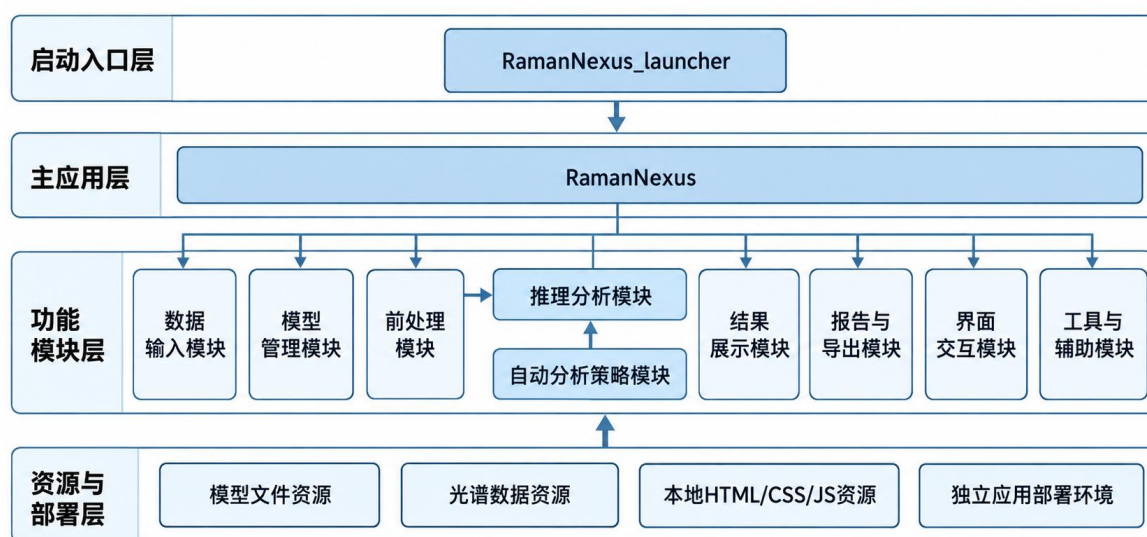
系统在计算层保留模型类型和处理参数的可配置性，在交互层统一组织光谱浏览与结果查看^[12,16]。分析日志、样本索引与导出结构共同关联操作状态和计算结果，为模型调试、结果复核及部署适配提供基础。

2.2 总体架构设计

总体架构包括启动入口、主应用、功能模块与界面资源四层。启动入口负责初始化，主应用维护状态并调度计算，功能模块封装数据读取、模型管理、前处理、推理及导出，界面资源层组织本地HTML、CSS、JavaScript与图标。明确的层间接口将界面控制与计算逻辑分离，减少跨模块依赖，便于维护与扩展^[20]。

总体架构如图1所示。主应用集中维护光谱矩阵、模型集合、当前样本与结果，功能模块据此执行计算，界面组件同步更新显示。共享状态避免不同操作路径分别维护样本与结果，并为资源定位和应用打包提供组织边界。

RamanNexus总体架构图



数据导入 → 模型加载 → 预处理 → 分析推理 → 结果展示 → 导出部署

图1 RamanNexus 总体架构图

2.3 模块组成与业务流程

功能模块按数据流划分。读取模块导入光谱与位移，模型模块完成类型识别和参数检查，推理模块统一调用分类及回归模型，自动策略模块根据类别选择定量模型。前处理模块完成光谱变换，结果模块组织预测表、日志与谱图输出，界面辅助模块

生成uihtml数据载荷。模块职责及输入输出见表1。

业务流程依次包括模型加载、数据导入、分析执行、结果查看及文件导出。系统根据分析模式进行分类、回归或联合分析，结果同步写入谱图、概览与结果表。各阶段共享样本索引，使输入光谱、模型输出及归档文件保持对应关系，具体流程如图2所示。

表 1 RamanNexus 主要功能模块及职责

模块	主要功能	输入	输出
io	光谱与位移导入、数据清理	xlsx/csv	矩阵或位移向量
models	定性/定量模型加载、类型识别与自检	mat 模型文件	模型结构体
infer	定性分类、定量预测、维度适配	光谱矩阵、模型	结果表
policy	依据定性结果筛选定量模型	Top1类别、模型键	筛选结果
pre	裁剪、去尖峰、基线校正、平滑、归一化	光谱、位移、参数	预处理光谱
report	导出 Excel、日志、预处理数据和图像	结果表、日志、光谱	xlsx/txt/png
ui/util	结果载荷转换、路径管理和资源定位	结果数据、路径	界面与部署辅助对象

RamanNexus业务流程图

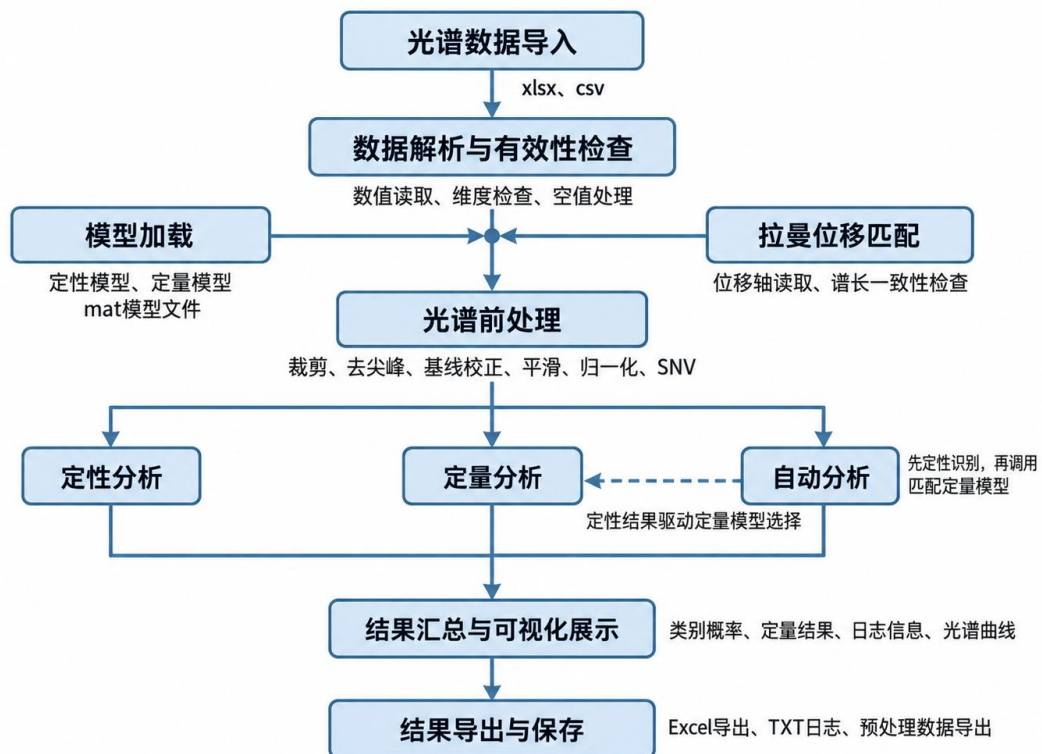


图 2 RamanNexus 业务流程图

3 关键功能设计与实现

3.1 数据导入与内部数据组织

系统支持xlsx与csv格式的光谱矩阵及拉曼位移

文件，采用readmatrix读取数值数据^[21]，并统一转换为double类型。光谱矩阵以行为样本、列为谱点组织，位移向量用于描述各列对应的拉曼位移。统一的数据结构为批量前处理、输入维度检查及模型推理提供基础。

数据导入后，读取模块清理空白行列，并检查

位移向量的数值有效性及其长度与谱点数的一致性。长度检查用于约束光谱与坐标的维度对应关系，波数范围和网格间隔则需要与模型训练条件匹配。对于谱线内部的缺失列，直接删除可能改变特征位置，因此文件外围空白与内部缺失数据应分别处理，以保持光谱坐标的物理含义。

检查完成后，光谱矩阵、位移向量、样本数量与索引写入统一应用状态，为前处理、模型推理及结果显示提供共同输入。原始数据与处理结果分别保存，以支持预览与分析输入切换。

3.2 模型管理与统一接入机制

分类模型以mat文件保存网络、归一化参数及类

别定义。加载阶段检查必要字段，并通过试推理验证输入张量、网络接口和输出结构的兼容性。试推理评价接口可执行性，预测性能则通过真实标签对应的测试结果评价。

定量模型根据字段识别为网络或KELM模型^[17]，封装为包含模型类型、变换参数与预测入口的标准结构体。上层模块调用统一接口，结果模块读取一致的输出结构，以减少模型差异对展示和导出逻辑的影响。

回归模型加载后以零输入执行自检，检查输出是否非空且为有限值。自检失败的模型不进入预测流程。该机制检查接口及数值有效性，不承担样本适用性或检测限评价。系统支持的数据与模型类型见表2。

表2 RamanNexus 支持的数据与模型类型

类别	支持类型	处理方式	说明
光谱数据	xlsx、csv	读取、选择、清理	实验光谱数据的主要输入形式
拉曼位移	xlsx、csv	合法性与长度校验	匹配后参与绘图与联动显示
定性模型	mat	提取分类网络、归一化参数、类别信息	用于定性模型加载
定量模型	mat	统一封装后预测	兼容 NET/KELM 模型
导出结果	xlsx、txt、png	统一导出	支持目录回退保存

3.3 光谱前处理模块

拉曼光谱采集易受到随机尖峰、背景漂移及噪声干扰^[22]。系统采用可组合前处理流程，依次组织位移区间裁剪、去尖峰、基线校正、平滑与归一化。去尖峰采用稳健异常值识别思路作为方法参考^[23]，本软件具体基于中值与中位数绝对偏差（median absolute deviation，MAD）进行异常检测与修正。惩罚最小二乘平滑^[24]、改进非对称最小二乘方法^[25]及非对称重加权惩罚最小二乘方法^[26]为基线估计提供相关研究背景；本文实际实现采用非对称最小二乘（asymmetric least squares，AsLS）基线校正。随后采用Savitzky-Golay（SG）平滑^[27]，并按配置执行向量归一化或标准正态变量变换（standard normal variate，SNV）^[28]。各步骤通过参数配置控制是否执行及执行条件，使不同模型的

处理需求能够在同一流程中表达。

前处理结果首先用于谱图预览。当处理前后特征维度一致时，系统允许将处理结果切换为模型分析输入。前处理次序、参数及波数范围共同决定输入分布，因此预测阶段应采用与模型训练一致的处理条件。维度检查能够约束张量结构，而背景校正和平滑对特征强度的影响仍需结合具体模型评价。

自动前处理配置通过中位谱、差分噪声及信号幅值估计谱线的噪声特征，并据此调整SG窗口、AsLS参数与去尖峰阈值，为人工参数选择提供初始配置。该策略属于规则驱动的参数估计，其用途是提高前处理预览的操作效率。图3给出了前处理前后的谱线对比，用于展示背景与噪声处理效果；处理参数在模型预测中的适用性取决于训练阶段的输入定义。

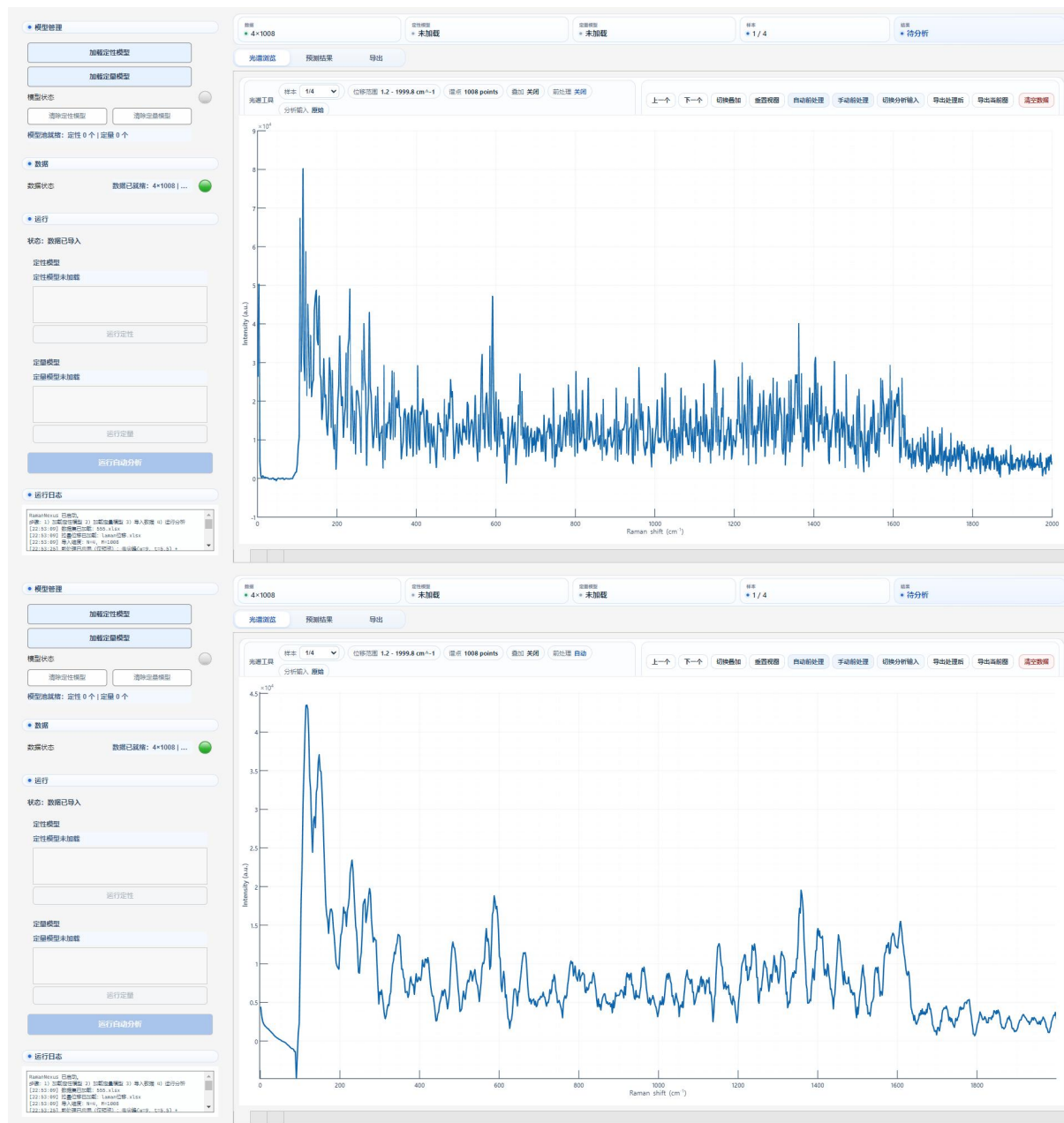


图 3 光谱前处理流程效果对比图

3.4 定性分析、定量分析与自动分析策略

定性分析以当前光谱矩阵为输入，利用模型保存的归一化参数完成数据变换，并依据网络结构构造输入张量，随后执行批量分类。结果表保存样本编号、Top1类别及分数、Top2类别及分数，保留前两位输出有助于观察类别间的相对响应。网络输出分数用于描述分类结果及排序，其概率校准性质需通过专门验证确定。

对于多分类模型接入，系统采用最大Top1分数规则汇总结果，并保留产生该结果的模型标识。该机制实现了不同分类器输出的统一管理，但未经校

准的模型分数不一定处于相同尺度，因而不据此推断融合增益。本文分类性能及数值一致性分析采用单个分类网络，避免将多模型汇总规则与单模型评价混合。

定量分析通过统一预测入口调用网络模型及KELM模型^[17]。网络模型依次执行输入变换、张量适配、网络预测和输出反归一化；KELM模型依据保存的特征列及参数完成核回归。两类计算路径返回统一格式的浓度结果，使样本编号、模型键及预测值能够由相同的展示和导出逻辑处理。

多环芳烃（polycyclic aromatic hydrocarbons, PAHs）定量任务采用GA-CNN-LSTM-Attention模

型。遗传算法(GA)用于模型优化,卷积神经网络(CNN)提取局部特征,长短期记忆网络(LSTM)建模谱点序列依赖,注意力机制对特征加权。软件加载训练完成的网络与变换参数执行推理,模型测试评价回归性能,直接调用对照评价软件实现的一致性。

定量计算采用模型保存的输入与输出变换参数,避免依据当前批次重新估计归一化参数所引起的输出变化。

自动分析通过分类结果建立定性与定量计算的路由关系。当Top1类别为PAHs时,系统调用ant、fla、flu和pyr定量模型;当类别与已加载模型键匹配时,调用对应模型;其余情形采用全部可用定量模型的回退规则。该策略将类别识别与浓度预测组织为连续流程,并保留模型来源信息,以便对不同输出进行结果追溯。

3.5 结果展示、导出机制与可部署适配

结果界面采用MATLAB桌面组件与本地HTML组件协同实现。uihtml加载本地HTML、CSS及JavaScript资源^[29],桌面端通过数据载荷传递模型状态、样本信息和结果表。显示层仅负责信息呈现与交互响应,模型运算由功能模块完成,使计算路径与展示逻辑保持分离。该设计支持结果列动态构造及样本联动,同时维持桌面应用的统一运行入口。

结果管理以分析日志、预测结果表和前处理数

据为主要归档对象,其设计参考了计算研究的过程记录与可重复性原则^[10,14],并支持当前谱图输出。MATLAB Runtime为编译后的MATLAB应用提供运行环境^[30]。本文在工程层面通过统一资源路径管理定位模型和界面文件;工程目录无写入权限时,导出路径回退至用户文档目录,以适应应用打包后的文件权限变化。路径适配与计算接口分离,使部署环境调整不会改变模型预测的执行逻辑。

4 交互界面与结果管理

4.1 界面总体布局

主界面采用操作区与数据展示区分离的布局。左侧操作区集成模型加载、分析执行及日志显示,右侧展示区由状态栏、光谱浏览页、结果页和导出页组成。状态栏呈现数据规模、模型加载及分析状态,标签页组织不同任务的信息视图。界面布局与系统分析流程相对应,使模型配置、数据浏览和结果管理能够在同一应用中完成,具体界面见图4。

系统以当前样本索引连接谱图绘制、结果概览和结果表定位。结果表选中事件更新样本索引并刷新相应谱图,样本切换事件则同步更新结果概览。两类操作共享同一索引变量,从而维持界面显示与模型输出的样本对应关系。该机制将交互状态与数据状态集中管理,减少批量分析时不同结果视图之间的定位偏差。

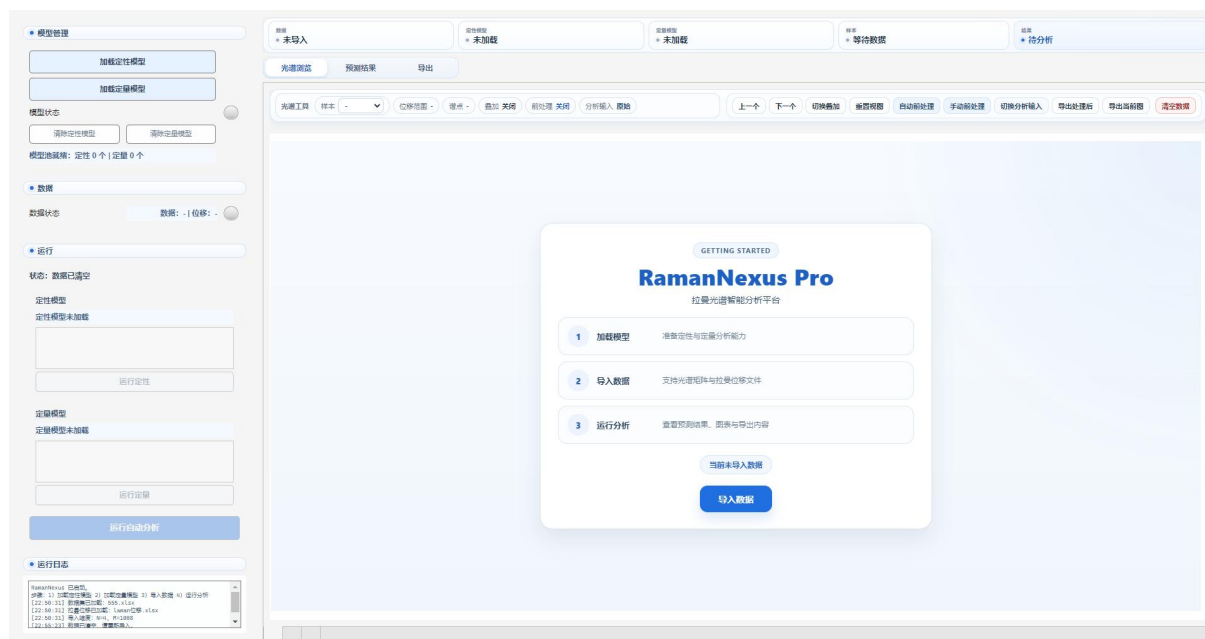


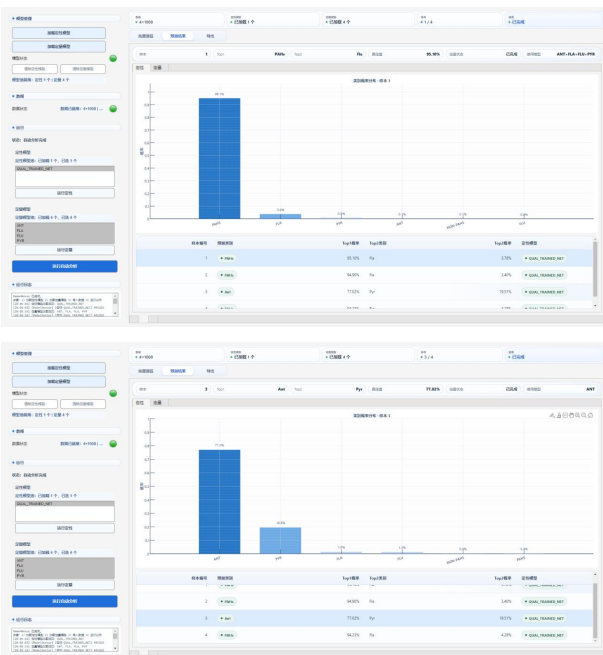
图4 RamanNexus主界面图

4.2 结果展示与导出页面

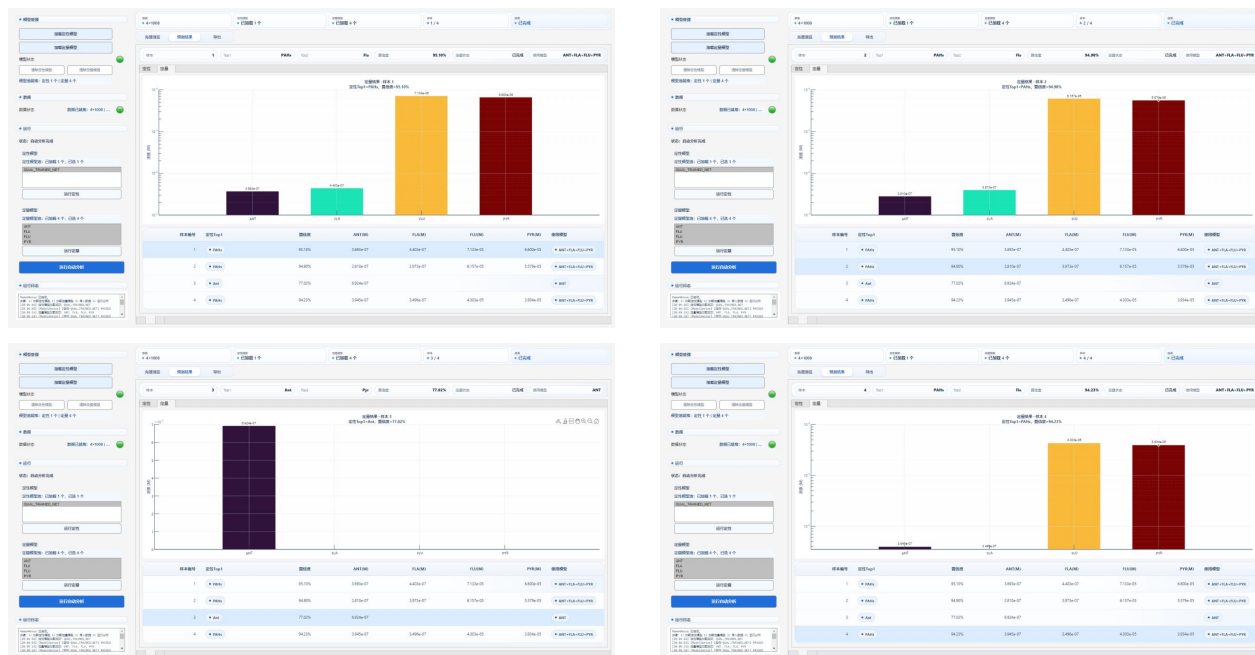
结果页分为定性与定量两个区域（见图5）。定性区域显示类别排序及模型分数，定量区域以动态列呈现不同回归模型的预测值，核心列保留样本编号及分类输出信息。模型数量或名称变化时，系统依据模型集合重新构造结果列，并将相同列定义用于界面与文件导出。统一列结构有助于维持可视化

结果与归档数据的一致性。

导出页统一组织结果表、日志、前处理数据及谱图保存操作（见图6）。集中式导出口将分析状态与输出文件关联，使批量计算后的结果归档具有一致的操作路径，并为后续结果复核提供数据依据。



定性页面



定量页面

图 5 定性与定量结果展示图

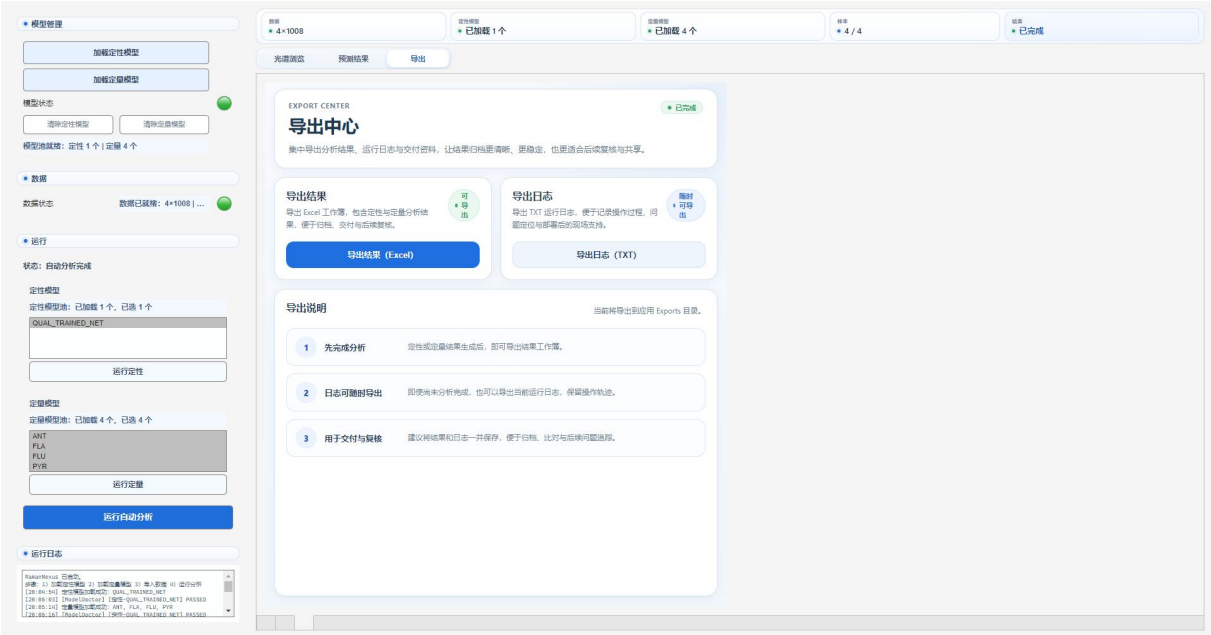


图 6 导出中心与结果文件示意图

5 实验设计与结果分析

5.1 测试环境与评价方案

实验在Windows 11环境下开展，处理器为Intel Core i7-12700H，内存为16GB，数值验证平台为MATLAB R2025b。评价对象包括一个PAHs分类网络及四个GA-CNN-LSTM-Attention回归模型。分类性能依据模型保存的360条测试标签与预测标签计算，回归性能采用葱、茺葱、茺和茺模型的测试评价结果。软件调用验证分别使用20条分类输入和20条回归输入，对照统一输入条件下的软件输出与直

接调用结果。模型性能与调用一致性分别统计，两组接口验证输入不用于联合识别与定量准确率计算。

5.2 功能测试

功能测试覆盖数据读取、模型加载、光谱前处理、定性识别、定量预测、自动分析及文件导出。测试通过检查各阶段输出及状态转换，评价分析流程能否完成预定操作。预测性能由分类与回归指标分别描述，计算接口的实现正确性则通过数值对照检验。功能测试项目及结果见表3，评价范围限于所采用的输入与操作路径。

表 3 RamanNexus 主要测试项目和结果

测试项	测试内容	结果
数据导入测试	xlsx、csv 光谱数据导入	所测流程完成
模型加载测试	mat 定性/定量模型加载	所测流程完成
前处理测试	裁剪、去尖峰、基线校正、平滑等	所测流程完成
定性分析测试	类别识别及概率输出	所测流程完成
定量分析测试	目标组分预测	所测流程完成
自动分析测试	先定性后定量流程	所测流程完成
导出测试	Excel、日志、预处理数据导出	所测流程完成

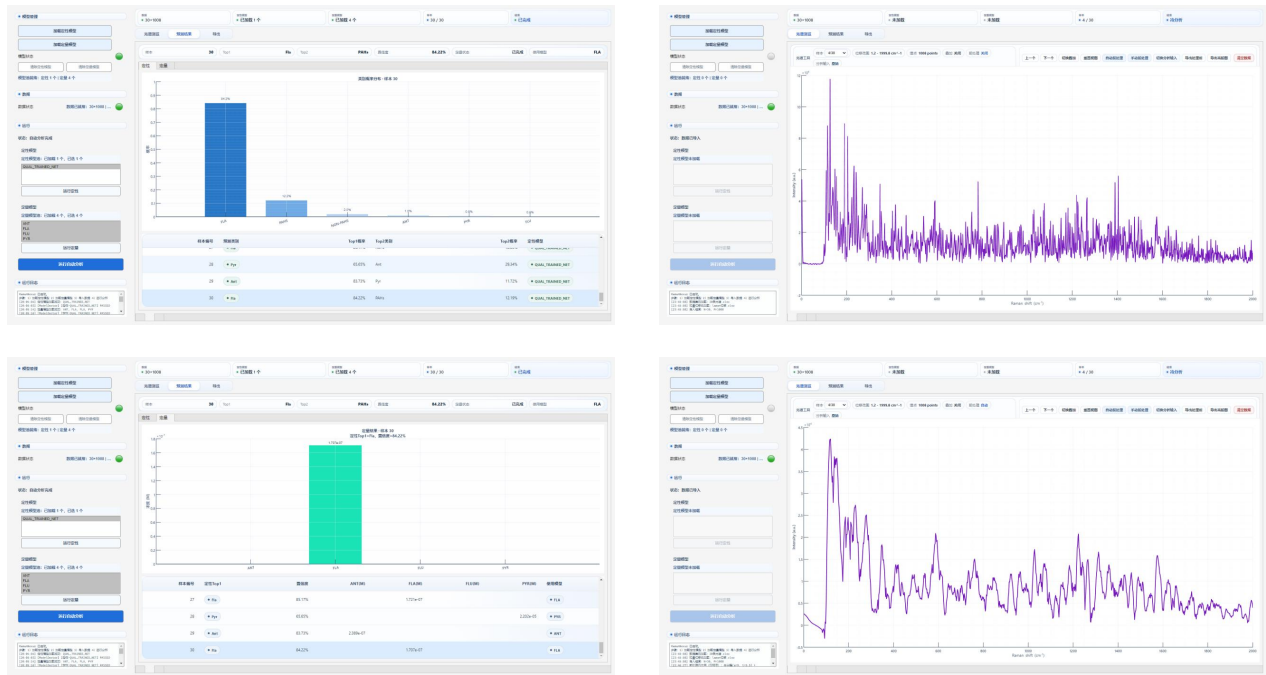


图 7 测试结果展示图

功能测试结果见图7。系统完成光谱矩阵读取与模型加载后，能够执行裁剪、去尖峰、基线校正、平滑及归一化等前处理操作。定性分析返回样本类别和模型分数，定量分析返回对应组分预测值，自动分析完成分类结果驱动的回归调用。各阶段结果

写入统一数据结构，并同步更新谱图与结果表，表明输入、计算及显示模块之间的数据链路能够完成预定任务。

名称	修改日期	类型	大小
 Figures	2026/3/17 21:18	文件夹	
 RamanNexus_Log_20260317_2117.txt	2026/3/17 23:59	文本文档	2 KB
 RamanNexus_PreprocessedSpectra_2...	2026/3/17 23:59	Microsoft Excel ...	420 KB
 RamanNexus_Result_20260317_2117....	2026/3/17 23:59	Microsoft Excel ...	13 KB

图 8 输出结果图

结果输出如图8所示。分析完成后，系统生成结果表、日志及前处理数据文件，并保存当前谱图。界面结果与导出结果共享样本编号和模型列定义，避免二者分别生成列结构所引起的对应差异。上述测试验证了所测条件下的流程完整性与结果归档功能，长时间运行、独立部署及大规模负载性能仍属于后续系统评价范围。

在功能测试基础上，进一步从分类性能、回归性能和数值一致性三个方面进行量化分析。分类指标描述标签识别结果，回归指标描述浓度预测误差，数值一致性指标描述软件与参考计算之间的实现差异，三类指标分别对应不同评价目标。

5.3 分类模型性能评价

分类性能采用Top1准确率及宏平均精确率、召回率和F1值评价。准确率为正确预测数占测试记录总数的比例；宏平均指标分别对各类别计算后取算术平均，使各类别具有相同权重。测试记录包含六个编码类别，每类60条，共360条；混淆矩阵中359条预测正确，一条真实编码3被预测为编码4。分类性能统计结果见表4。

表4 分类模型测试性能

评价指标	复核数值	记录范围
Top1准确率	99.72%	359/360
宏平均 Precision	0.9973	六个编码类别
宏平均 Recall	0.9972	六个编码类别
宏平均 F1	0.9972	六个编码类别

分类模型测试记录的Top1准确率为99.72%，宏平均精确率、召回率及F1分别为0.9973、0.9972和0.9972。840条训练记录中839条预测正确，训练准确率为99.88%。这些结果表明分类网络在保存的训练与测试记录上具有较高识别率。由于评价依据为保存的标签与预测结果，其结论适用于该模型的既有数据划分，不扩展为新采集样本或其他测量条件下的识别性能。

5.4 模型调用与前处理的数值一致性验证

模型调用验证采用相同输入比较软件封装接口与直接网络调用。20条分类输入的预测类别完全一致；不同执行方式下的最大分数差为 2.563×10^{-5} ，归一化数据差为0。回归验证保持模型、变换参数与后处理规则一致，并统一执行方式。浓度及其差值的单位统一为 $\text{mol} \cdot \text{L}^{-1}$ （ $1 \text{ M} = 1 \text{ mol} \cdot \text{L}^{-1}$ ）。四个回归模型各20条输入的软件预测值与参考结果一致，最大绝对差均为0。上述一致性反映计算路径的实现关系，不等同于对真实类别或浓度的预测误差。回归模型调用一致性验证结果见表5。

表5 回归模型调用一致性验证结果

模型键	输入条数	最大绝对差/($\text{mol} \cdot \text{L}^{-1}$)	显式双精度对照差/($\text{mol} \cdot \text{L}^{-1}$)
ant	20	0	1.252×10^{-14}
fla	20	0	3.576×10^{-14}
flu	20	0	1.394×10^{-13}
pyr	20	0	1.205×10^{-13}

为分析数值精度对调用结果的影响，采用显式双精度计算完成输出反归一化及缩放恢复，对照中的最大绝对差约为 $1.40 \times 10^{-13} \text{ mol} \cdot \text{L}^{-1}$ 。该结果说明在所测输入和执行条件下，软件封装未引入可辨识的额外数值偏差。直接调用对照与真实浓度评价具有不同参照对象，前者验证推理实现，后者评价模型误差，二者应分别解释。

前处理验证采用具有已知数学性质的构造输入。SG平滑对三次多项式的最大误差为 3.55×10^{-15} ，内部点与参考函数的最大差为 2.91×10^{-11} ；SNV变换后样本均值的最大绝对值为 2.23×10^{-16} ，标准差相对1的最大偏差为 4.44×10^{-16} ；向量归一化的模长偏差为0；AsLS在线性背景输入上的最大残差约为 $4.02 \times$

10^{-9} 。结果支持这些基础运算在特定输入上的数值一致性，而实测光谱的前处理效果还取决于谱线特征及参数配置。

5.5 定量模型性能评价

定量性能采用平均绝对误差（MAE）、均方根误差（RMSE）、均方误差（MSE）、决定系数(R^2)及平均绝对百分比误差（MAPE）评价。MAE反映平均绝对偏差，RMSE与MSE对较大误差具有更高敏感性， R^2 描述模型对响应变量变异的解释程度，MAPE描述相对误差。蒽（ant）、荧蒽（fla）、苊（flu）和芘（pyr）的GA-CNN-LSTM-Attention测试指标见表6。

表6 GA-CNN-LSTM-Attention 模型测试指标

目标物	MAE	RMSE	R	MAPE	MSE
ant	0.026450	0.038894	0.9884	42.8708	0.0015127
fla	0.046080	0.073174	0.9585	50.3531	0.0053545
flu	0.026134	0.051229	0.9988	68.7804	0.0026244
pyr	0.062431	0.087972	0.9964	20.9868	0.0077390

综合上述指标,模型在四种目标物的测试数据上均表现出较高的拟合程度,但相对预测误差仍存在差异,其定量分析能力应结合实际浓度范围和应用所需的误差限进行评价。

5.6 讨论

实验结果分别从模型预测与软件实现两个层面支持系统的分析功能。分类与回归指标评价既定数据条件下的模型性能,直接调用对照则验证封装接口的计算一致性。较高 R^2 不必然对应较小的低浓度相对误差,尤其当真实浓度接近零时,MAPE对分母变化具有较高敏感性。因此,定量性能应结合绝对误差、相对误差及浓度区间综合分析。软件设计通过统一模型描述、样本索引和结果结构组织计算流程,其工程价值主要体现在模型接入与结果管理的一致性。

系统的适用性仍受输入波数网格、处理条件及模型训练范围约束。边界输入验证显示,局部NaN未在读取阶段被拒绝,内部全缺失列可能被删除,全零背景叠加单点尖峰时MAD阈值未识别该异常。这些结果表明缺失值处理与去尖峰规则需要结合异常输入类型进一步完善。对于未知类别回退及分类、回归输入长度差异,后续应增加模型适用性约束及物理波数网格检查。现有测试以既定模型及保存记录为基础,模型泛化性能还需通过物理样本分组、跨批次及外部数据验证。

6 结论

本文面向拉曼光谱定性定量分析的软件集成需求,设计并实现了RamanNexus桌面系统。分层架构及统一模型接口将光谱导入、前处理、分类识别、浓度预测和结果导出组织为连续工作流,支持网络模型与KELM模型的统一管理。样本索引联动及动态结果列使谱图、预测表与输出文件保持对应关系,为批量光谱分析及结果复核提供了实现基础。

分类模型测试记录的Top1准确率为99.72%,宏平均F1为0.9972;GA-CNN-LSTM-Attention在葱、茺菹、茺和茺任务上的 R^2 分别为0.9884、0.9585、0.9988和0.9964。在统一执行条件下,四个回归模型各20条输入的软件预测值与直接调用结果一致。上述结果支持既有模型在桌面环境中的集成应用及推理实现的一致性。后续研究将重点开展浓度分层评价、跨批次及跨测量条件验证,并完善异常输入处理、模型适用范围约束和独立部署测试。

伦理声明

本研究不涉及人体或动物实验,也不涉及个人敏感信息,无需相关伦理审批。

利益冲突声明

全体作者声明不存在与本研究相关的利益冲突。

数据可用性声明

本研究相关数据可根据合理申请由通讯作者提供。

基金资助声明

本研究未获得基金项目资助。

参考文献

- [1] CHEN J, HUANG Y W, ZHAO Y. Characterization of polycyclic aromatic hydrocarbons using Raman and surface - enhanced Raman spectroscopy[J]. Journal of Raman Spectroscopy, 2015, 46(1): 64-69.
- [2] LIU J, OSADCHY M, ASHTON L, et al. Deep convolutional neural networks for Raman spectrum recognition: a unified solution[J]. Analyst, 2017, 142(21): 4067-4074.
- [3] XU Y, ZHONG P, JIANG A, et al. Raman spectroscopy coupled with chemometrics for food authentication: A review[J]. TrAC Trends in Analytical Chemistry, 2020, 131: 116017.
- [4] BUTLER H J, ASHTON L, BIRD B, et al. Using Raman spectroscopy to characterize biological materials[J]. Nature Protocols, 2016, 11(4): 664-687.
- [5] HO C S, JEAN N, HOGAN C A, et al. Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning[J]. Nature Communications, 2019, 10(1): 4927.
- [6] QI Y, HU D, JIANG Y, et al. Recent Progresses in Machine Learning Assisted Raman Spectroscopy[J]. Advanced Optical Materials, 2023, 11(14): 2203104.
- [7] ACQUARELLI J, VAN LAARHOVEN T, GERRETZEN J, et al. Convolutional neural networks for vibrational spectroscopic data analysis[J]. Analytica Chimica Acta, 2017, 954: 22-31.

- [8] LILAND K H, ALMØY T, MEVIK B H. Optimal Choice of Baseline Correction for Multivariate Calibration of Spectra[J]. *Applied Spectroscopy*, 2010, 64(9): 1007-1016.
- [9] MA C, ZHAI L, DING J, et al. Raman spectroscopy combined with partial least squares (PLS) based on hybrid spectral preprocessing and backward interval PLS (biPLS) for quantitative analysis of four PAHs in oil sludge[J]. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 2024, 310: 123953.
- [10] SANDVE G K, NEKRUTENKO A, TAYLOR J, et al. Ten Simple Rules for Reproducible Computational Research[J]. *PLOS Computational Biology*, 2013, 9(10): e1003285.
- [11] GEORGIEV D, PEDERSEN S V, XIE R, et al. RamanSPy: An Open-Source Python Package for Integrative Raman Spectroscopy Data Analysis[J]. *Analytical Chemistry*, 2024, 96(21): 8492-8500.
- [12] BREHM S, HIMCINSCHI C, KRAUS J, et al. PyRamanGUI: Open-source graphical user interface for analyzing Raman spectra[J]. *SoftwareX*, 2023, 23: 101486.
- [13] PAVLOU E, KOURKOUMELIS N. PyFasma: an open-source, modular Python package for preprocessing and multivariate analysis of Raman spectroscopy data[J]. *Analyst*, 2025, 150(14): 3112-3122.
- [14] WILSON G, BRYAN J, CRANSTON K, et al. Good enough practices in scientific computing[J]. *PLOS Computational Biology*, 2017, 13(6): e1005510.
- [15] BARTON B, THOMSON J, LOZANO DIZ E, et al. Chemometrics for Raman Spectroscopy Harmonization[J]. *Applied Spectroscopy*, 2022, 76(9): 1021-1041.
- [16] GEORGIEV G, COCA - LOPEZ N, LELLINGER D, et al. Open Source for Raman Spectroscopy Data Harmonization[J]. *Journal of Raman Spectroscopy*, 2025, 56(9): 878-881.
- [17] HUANG G B, ZHOU H, DING X, et al. Extreme Learning Machine for Regression and Multiclass Classification[J]. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 2012, 42(2): 513-529.
- [18] SHEEHY G, PICOT F, DALLAIRE F, et al. Open-sourced Raman spectroscopy data processing package implementing a baseline removal algorithm validated from multiple datasets acquired in human tissue and biofluids[J]. *Journal of Biomedical Optics*, 2023, 28(2): 025002.
- [19] YAN C. A review on spectral data preprocessing techniques for machine learning and quantitative analysis[J]. *iScience*, 2025, 28(7): 112759.
- [20] WILSON G, ARULIAH D A, BROWN C T, et al. Best Practices for Scientific Computing[J]. *PLOS Biology*, 2014, 12(1): e1001745.
- [21] MATHWORKS. readmatrix: 从文件中读取矩阵 [EB/OL]. [2026-10-11]. <https://ww2.mathworks.cn/help/matlab/ref/readmatrix.html>.
- [22] VULCHI R T, MORGUNOV V, JUNJURI R, et al. Artifacts and Anomalies in Raman Spectroscopy: A Review on Origins and Correction Procedures[J]. *Molecules*, 2024, 29(19): 4748.
- [23] WHITAKER D A, HAYES K. A simple algorithm for despiking Raman spectra[J]. *Chemometrics and Intelligent Laboratory Systems*, 2018, 179: 82-84.
- [24] EILERS P H C. A Perfect Smoother[J]. *Analytical Chemistry*, 2003, 75(14): 3631-3636.
- [25] HE S, ZHANG W, LIU L, et al. Baseline correction for Raman spectra using an improved asymmetric least squares method[J]. *Analytical Methods*, 2014, 6(12): 4402-4407.
- [26] BAEK S J, PARK A, AHN Y J, et al. Baseline correction using asymmetrically reweighted penalized least squares smoothing[J]. *Analyst*, 2015, 140(1): 250-257.
- [27] SAVITZKY A, GOLAY M J E. Smoothing and Differentiation of Data by Simplified Least Squares Procedures[J]. *Analytical Chemistry*, 1964, 36(8): 1627-1639.
- [28] BARNES R J, DHANOA M S, LISTER S J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra[J]. *Applied Spectroscopy*, 1989, 43(5): 772-777.
- [29] MATHWORKS. uihtml: 创建 HTML UI 组件 [EB/OL]. [2026-10-11]. <https://ww2.mathworks.cn/help/matlab/ref/uihtml.html>.
- [30] MATHWORKS. MATLAB Runtime[EB/OL]. [2026-10-11]. <https://ww2.mathworks.cn/products/compiler/matlab-runtime.html>.

Design and Implementation of RamanNexus: A Software for Qualitative and Quantitative Raman Spectral Analysis

Zhong Chenrui¹, Tian Yaping^{2*}

1. School of Physics and Electronic Engineering, Jiangsu Normal University, Xuzhou 221116

2. Kewen College, Jiangsu Normal University, Xuzhou 221132

Abstract: Raman spectral analysis often involves separate workflows for classification, concentration prediction, and result management. RamanNexus was developed in MATLAB R2025b to integrate these operations into a desktop application. A layered architecture and unified model interface support data import, preprocessing, batch inference, visualization, and export for neural network and kernel extreme learning machine models. Classification performance was evaluated using stored test labels and predictions, regression performance using GA-CNN-LSTM-Attention model test results, and implementation consistency by comparison with direct model calls. The 360 classification test records yielded a Top1 accuracy of 99.72% and a macro-averaged F1 score of 0.9972. Regression R^2 values were 0.9884, 0.9585, 0.9988, and 0.9964 for anthracene, fluoranthene, fluorene, and pyrene, respectively. RamanNexus provides an integrated environment for applying trained spectral models and managing batch analysis results.

Keywords: Raman spectroscopy; intelligent analysis software; MATLAB