# DEVELOPMENT AND APPLICATION OF AN AUTOMATED RAMAN SPECTROSCOPY–MACHINE LEARNING SYSTEM FOR QUALITY CONTROL OF BIOLOGICAL SAMPLE PRESERVATION SOLUTIONS

Chu Van Ket, Do Tuan Men, Trinh Minh Viet

*Department of Biochemistry, Institute 69 - Hoang Hoa Tham Street, Ngoc Ha Ward, Hanoi City, Vietnam*

Received: 14/07/2026
Reviced: 19/08/2026; Accepted: 08/09/2026

## ABSTRACT

**Objective:** To develop a machine learning model integrated with Raman spectroscopy for the rapid quantification of potassium acetate in preservation solutions.

**Methods:** The study was conducted using a LabRam Soleil Raman spectrometer (532 nm) and a dataset comprising 240 Raman spectra obtained from potassium acetate/glycerol samples. The data were divided into training, validation, and test sets. Five-fold cross-validation and Grid Search were applied for hyperparameter optimization. Four models (MLR, PCR, PLS, and ANN) were developed and evaluated using $R^2$, RMSE, and MAPE. Model performance was further compared with potentiometric titration using 10 real-world samples.

**Results:** The PLS model with six latent variables achieved the best overall performance ($R^2$ = 0.9991, RMSE = 0.52%, and MAPE = 1.24%), with a training time of 2.5 seconds and high interpretability. The automated workflow reduced analysis time by 85% and enabled the processing of 96 samples per day. A strong correlation was observed between the PLS-Raman method and potentiometric titration ($r = 0.985$, $p < 0.001$), demonstrating high specificity for potassium acetate determination.

**Conclusion:** The Raman-PLS system provides a rapid, accurate, and reliable analytical tool with strong potential for quality control of biological sample preservation solutions.

***Keywords:*** Raman spectroscopy; machine learning; partial least squares regression; potassium acetate; preservation solution.

*Corresponding author
**Email:** dr.minhviet@gmail.com **Phone**: (+84) 988663688 **DOI:** 10.52163/yhc.v67iCD14.6491

# XÂY DỰNG VÀ ỨNG DỤNG HỆ THỐNG PHÂN TÍCH TỰ ĐỘNG DỰA TRÊN PHỔ RAMAN VÀ HỌC MÁY NHẰM KIỂM SOÁT CHẤT LƯỢNG DUNG DỊCH BẢO QUẢN MẪU SINH HỌC

Chu Văn Kết, Đỗ Tuấn Mến, Trịnh Minh Việt

*Khoa Sinh hóa - Viện 69 - Đường Hoàng Hoa Thám, phường Ngọc Hà, thành phố Hà Nội, Việt Nam*

Ngày nhận bài: 14/07/2026
Ngày chỉnh sửa: 19/08/2026; Ngày duyệt đăng: 08/09/2026

## TÓM TẮT

**Mục tiêu:** Xây dựng mô hình học máy kết hợp phổ Raman để định lượng nhanh kali acetat trong dung dịch bảo quản.

**Phương pháp:** Nghiên cứu được thực hiện trên hệ thống Raman LabRam Soleil (532 nm) với 240 phổ Raman của các mẫu kali acetat/glycerin. Dữ liệu được chia thành tập huấn luyện, xác nhận và kiểm tra; áp dụng kiểm định chéo 5 lần và tối ưu siêu tham số bằng Grid Search. Bốn mô hình (MLR, PCR, PLS và ANN) được xây dựng, đánh giá bằng $R^2$, RMSE và MAPE, đồng thời so sánh với phương pháp chuẩn độ điện thế trên 10 mẫu thực tế.

**Kết quả:** Mô hình PLS với 6 thành phần chính cho kết quả tối ưu ($R^2$ = 0,9991; RMSE = 0,52%; MAPE = 1,24%), thời gian huấn luyện 2,5 giây và khả năng giải thích cao. Hệ thống tự động hóa giúp giảm 85% thời gian phân tích và xử lý 96 mẫu/ngày. Kết quả tương quan chặt chẽ với phương pháp chuẩn độ điện thế (r = 0,985; p < 0,001) và cho thấy độ đặc hiệu cao đối với kali acetat.

**Kết luận:** Hệ thống Raman-PLS là công cụ phân tích nhanh, chính xác và tin cậy, có tiềm năng ứng dụng trong kiểm soát chất lượng dung dịch bảo quản mẫu sinh học.

***Từ khóa:*** Quang phổ Raman, học máy, PLS, kali acetat, dung dịch bảo quản.

## 1. ĐẶT VẤN ĐỀ

Quang phổ Raman là kỹ thuật phân tích phân tử không phá hủy, được ứng dụng rộng rãi trong kiểm soát chất lượng dược phẩm, thực phẩm và y sinh [1]. Tuy nhiên, việc định lượng từ phổ Raman gặp khó khăn do các tín hiệu phổ phức tạp và chồng chéo, đòi hỏi chuyên môn cao trong xử lý dữ liệu [2]. Trong xu hướng chuyển đổi số ngành Y tế, tự động hóa phòng xét nghiệm là yêu cầu cấp thiết nhằm nâng cao hiệu quả và giảm sai sót [3]. Đối với kiểm soát chất lượng dung dịch bảo quản mẫu sinh học, kali acetat là thành phần quan trọng cần được giám sát thường xuyên. Phương pháp chuẩn độ điện thế hiện nay có độ chính xác cao nhưng mất nhiều thời gian, hóa chất và nhân lực. Sự kết hợp giữa quang phổ Raman và các thuật toán học máy, đặc biệt là Bình phương tối thiểu từng phần (PLS), đã cho thấy khả năng định lượng nhanh và chính xác [4,5]. Tuy nhiên, tại Việt Nam chưa có nghiên cứu đánh giá toàn diện việc tích hợp phổ Raman, học máy và tự động hóa trong kiểm soát chất lượng dung dịch bảo quản. Nghiên cứu này nhằm xây dựng mô hình học máy định lượng kali acetat từ phổ Raman, phát triển quy trình phân tích tự động và đánh giá hiệu quả trên mẫu thực tế so với phương pháp chuẩn độ điện thế.

## 2. ĐỐI TƯỢNG VÀ PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Thiết kế nghiên cứu

Đây là một nghiên cứu phát triển và xác nhận phương pháp phân tích, gồm 4 giai đoạn: (1) Xây dựng cơ sở dữ liệu phổ Raman; (2) Phát triển và tối ưu các mô hình học máy; (3) Xây dựng quy trình tự động hóa; (4) Xác nhận phương pháp và đánh giá trên mẫu thực tế.

### 2.2. Đối tượng nghiên cứu và cấu trúc dữ liệu

Mẫu chuẩn: Dung dịch kali acetat (PA, ≥99%, Merck) ở sáu nồng độ: 10, 15, 20, 25, 30, 35 và 40% (m/m). Mỗi nồng độ được pha lặp lại 6 lần (6 lần pha độc lập), tạo thành 6 × 6 = 36 mẫu chuẩn cho mỗi nồng độ × 6 nồng độ = 180 mẫu chuẩn. Mẫu hỗn hợp: 60 mẫu chứa kali acetat và glycerin với các tỷ lệ khác nhau, nồng độ cao nhất 40%. Mẫu thực nghiệm: 10 mẫu dung dịch bảo quản từ các lô sử dụng tại Viện 69.

*Cỡ mẫu và phân bố dữ liệu:*

Tổng số mẫu: 240 mẫu (180 chuẩn + 60 hỗn hợp).

Số phổ đầu vào mô hình: Mỗi mẫu được đo lặp 3 lần trên máy Raman và lấy phổ trung bình để giảm nhiễu ngẫu nhiên. Như vậy, tổng số phổ đưa vào mô hình học máy là 240 phổ, không phải 720 phổ. Ba lần đo chỉ là kỹ thuật lặp kỹ thuật (technical replicates), không được coi là các mẫu độc lập.

*Phân bố tập dữ liệu (chia theo mẫu độc lập, không chia theo phổ):*

Tập huấn luyện (70%): 126 mẫu chuẩn + 42 mẫu hỗn hợp = 168 mẫu.

Tập xác nhận (15%): 27 mẫu chuẩn + 9 mẫu hỗn hợp = 36 mẫu.

*Tác giả liên hệ
**Email:** dr.minhviet@gmail.com **Điện thoại**: (+84) 988663688 **DOI:** 10.52163/yhc.v67iCD14.6491

Tập kiểm tra (15%): 27 mẫu chuẩn + 9 mẫu hỗn hợp = 36 mẫu.

Mẫu thực tế: 10 mẫu độc lập, không tham gia vào quá trình huấn luyện.

Việc chia tập dữ liệu được thực hiện theo từng mẫu độc lập (split by sample/batch), không phải ngẫu nhiên theo từng phổ đo. Điều này đảm bảo các mẫu pha lặp từ cùng một nồng độ danh nghĩa không bị tách ra vào các tập khác nhau, tránh hiện tượng rò rỉ nhóm - một lỗi nghiêm trọng có thể làm đánh giá quá cao hiệu suất mô hình.

Giải thích về cỡ mẫu: Với dữ liệu phổ Raman có 1345 biến số số sóng, cỡ mẫu 240 được coi là phù hợp cho các mô hình hồi quy đa biến có giảm chiều (PCA/PLS) hoặc có cơ chế chọn lọc biến. Tỷ lệ mẫu/biến số là khoảng 1:5,6, nằm trong phạm vi chấp nhận được đối với các mô hình chemometrics khi số chiều hiệu dụng được giảm xuống còn 6-10 thành phần chính.

### 2.3. Thiết bị và hóa chất

Thiết bị: Máy Raman LabRam Soleil (Horiba), laser 532 nm; máy chuẩn độ điện thế T7 (Mettler Toledo); máy tính cấu hình cao (CPU Intel Core i9, RAM 64GB, GPU RTX 4080).

Hóa chất: Kali acetat (PA, ≥99%, Merck), Glycerin (PA, ≥99%, Merck), nước cất khử ion.

### 2.4. Phương pháp thu thập và xử lý số liệu

*2.4.1. Quy trình đo phổ Raman:*

Hút 2 mL mẫu vào cuvet thủy tinh. Đo phổ với thông số tối ưu: công suất laser 63%, thời gian tích phân 15 giây, 3 lần đo, dải sóng 400-4000 $cm^{-1}$. Phổ trung bình của 3 lần đo được sử dụng cho phân tích.

*2.4.2. Tiền xử lý dữ liệu:*

Dữ liệu phổ được tiền xử lý qua pipeline: (1) Hiệu chỉnh nền bằng thuật toán airPLS [6]; (2) Chuẩn hóa bằng Standard Normal Variate (SNV); (3) Cắt dải phổ giữ vùng 400-1800 $cm^{-1}$ (vùng "dấu vân tay"); (4) Làm mượt bằng bộ lọc Savitzky-Golay (bậc 2, cửa sổ 9 điểm).

*2.4.3. Các mô hình học máy và quy trình xây dựng theo bốn thuật toán:*

Hồi quy tuyến tính đa biến(MLR). Hồi quy thành phần chính (PCA). Bình phương tối thiểu từng phần (PLS). Mạng nơ-ron nhân tạo với 3 lớp ẩn (256-128-64 neuron), hàm kích hoạt ReLU, thuật toán tối ưu Adam, tốc độ học 0,001 (ANN).

Quy trình huấn luyện và tối ưu: Chia dữ liệu: Ngẫu nhiên theo từng mẫu độc lập, đảm bảo đại diện cho các khoảng nồng độ. Kiểm định chéo (Cross-Validation): Sử dụng k-fold với k=5 trên tập huấn luyện để đánh giá độ ổn định và lựa chọn siêu tham số. Tối ưu siêu tham số: Sử dụng Grid Search kết hợp với 5-fold CV để chọn số thành phần chính (2-15) cho PLS/PCR và số neuron, tốc độ học cho ANN.

Đánh giá cuối cùng: Mô hình được đánh giá một lần duy nhất trên tập kiểm tra chưa từng được sử dụng trong toàn bộ quá trình huấn luyện và tối ưu.

*2.4.4. Đánh giá hiệu suất mô hình:*

Các chỉ số: $R^2$, RMSE, MAPE. Các chỉ số được báo cáo kèm độ lệch chuẩn (SD) từ kết quả của 5-fold CV.

*2.4.5. Xây dựng hệ thống tự động hóa:*

Hệ thống được xây dựng bằng Python (Scikit-learn, TensorFlow, Streamlit) với các mô-đun: Thu nhận (API), Tiền xử lý (pipeline tự động), Dự đoán (PLS), và Báo cáo (tự động xuất PDF/Excel).

### 2.5. Phương pháp đối chứng và xác nhận phương pháp theo ICH Q2(R2)

Phương pháp đối chứng: Phương pháp chuẩn độ điện thế trên máy T7

Xác nhận phương pháp theo ICH Q2(R2): Thực hiện đánh giá các thông số sau:

Độ đúng: Đánh giá qua tỷ lệ thu hồi (%) trên các mẫu chuẩn đã biết nồng độ (n=10, trải dài từ 10-40%) và trên các mẫu thực tế được thêm chuẩn ở ba mức nồng độ.

Độ chụm: Đánh giá độ lặp lại thông qua độ lệch chuẩn tương đối (RSD%) của 6 lần phân tích lặp trên cùng một mẫu ở ba mức nồng độ (10%, 25%, 40%).

Độ đặc hiệu/Độ chọn lọc: Được chứng minh thông qua phân tích hệ số hồi quy (Regression Coefficients) và chỉ số VIP (Variable Importance in Projection) của mô hình PLS. Các vùng số sóng có hệ số đóng góp lớn nhất sẽ được đối chiếu với các pic đặc trưng của kali acetat để khẳng định mô hình đang «bắt» đúng tín hiệu của chất phân tích, không bị nhiễu bởi glycerin hoặc nền.

### 2.6. Xử lý thống kê

Số liệu được xử lý bằng Python và SPSS 27. Phân phối chuẩn được kiểm tra bằng phép thử Shapiro–Wilk. Sự khác biệt giữa các mô hình được đánh giá bằng ANOVA một yếu tố và kiểm định hậu kiểm Tukey HSD. Mối tương quan giữa hai phương pháp được xác định bằng hệ số Pearson (r), trong khi độ phù hợp được đánh giá bằng phân tích Bland–Altman (bias và giới hạn đồng thuận 95%). Mức ý nghĩa thống kê được chọn là $\alpha = 0,05$.

## 3. KẾT QUẢ NGHIÊN CỨU

### 3.1. Đặc điểm dữ liệu phổ Raman

Phổ Raman của kali acetat cho thấy các pic đặc trưng tại 930 $cm^{-1}$ (C-C), 1415 $cm^{-1}$ ($COO^-$) và 1345 $cm^{-1}$ ($CH_3$). Sự chồng chéo với glycerin chủ yếu xảy ra ở vùng 900-1100 $cm^{-1}$, giải thích cho nhu cầu sử dụng mô hình đa biến.

### 3.2. Hiệu quả của các mô hình học máy

**Bảng 1. Hiệu suất của các mô hình trên tập kiểm tra (sau 5-fold cross-validation)**

| Mô hình | $R^2$ (trung bình ± SD) | RMSE (%) (TB ± SD) | MAPE (%) (TB ± SD) | Thời gian huấn luyện (s) | Thời gian dự đoán (ms) |
|---|---|---|---|---|---|
| MLR | 0,9847 ± 0,0065 | 1,21 ± 0,13 | 3,45 ± 0,40 | 0,5 | 0,1 |
| PCR | 0,9918 ± 0,0042 | 0,88 ± 0,09 | 2,31 ± 0,27 | 1,2 | 0,2 |
| PLS | 0,9991 ± 0,0002 | 0,52 ± 0,06 | 1,24 ± 0,15 | 2,5 | 0,3 |
| ANN | 0,9995 ± 0,0001 | 0,38 ± 0,04 | 0,89 ± 0,10 | 185 | 0,8 |

Phân tích ANOVA một yếu tố cho thấy chỉ số MAPE khác biệt có ý nghĩa thống kê giữa bốn mô hình ($F(3,16) = 98,47$; $p < 0,001$). Kiểm định hậu kiểm Tukey HSD cho thấy PLS có MAPE thấp hơn đáng kể so với MLR ($p < 0,001$) và PCR ($p < 0,01$), trong khi ANN đạt MAPE thấp hơn PLS ($p < 0,05$).

Mặc dù ANN cho kết quả tốt nhất trên tập kiểm tra ($R^2 = 0,9995$), PLS được lựa chọn làm mô hình tối ưu cho ứng dụng thực tế. Với bộ dữ liệu gồm 240 mẫu, ANN có nguy cơ quá khớp cao hơn do số lượng tham số lớn, trong khi PLS ổn định và phù hợp hơn với tập dữ liệu kích thước vừa phải. Bên cạnh đó, PLS có ưu thế về khả năng giải thích nhờ các hệ số hồi quy và chỉ số VIP, cho phép xác định các vùng phổ Raman liên quan trực tiếp đến kali acetat, góp phần nâng cao độ tin cậy khoa học của mô hình. Ngoài ra, PLS có thời gian huấn luyện và cập nhật mô hình ngắn hơn, thuận lợi cho triển khai và duy trì trong môi trường phòng xét nghiệm thường xuyên cần hiệu chỉnh theo dữ liệu mới.

### 3.3. Lựa chọn số thành phần chính cho PLS

**Bảng 2. Tối ưu số thành phần chính cho PLS (qua 5-fold CV)**

| Số thành phần | RMSEC (%) (TB ± SD) | RMSEP (%) (TB ± SD) |
|---|---|---|
| 4 | 0,85 ± 0,06 | 1,04 ± 0,08 |
| 6 | 0,48 ± 0,05 | 0,52 ± 0,06 |
| 8 | 0,35 ± 0,04 | 0,76 ± 0,10 |
| 10 | 0,22 ± 0,03 | 1,12 ± 0,15 |

Số thành phần tối ưu là 6, tại đó RMSEP đạt cực tiểu. Khi tăng lên 8 và 10, RMSEC tiếp tục giảm (mô hình khớp quá tốt với dữ liệu huấn luyện) nhưng RMSEP lại tăng, là dấu hiệu rõ ràng của hiện tượng quá khớp.

### 3.4. Vai trò của tiền xử lý

Tiền xử lý đầy đủ (airPLS + SNV + cắt + làm mượt) đã cải thiện đáng kể hiệu suất mô hình PLS: MAPE giảm từ 4,56% (không xử lý) xuống 1,24%.

### 3.5. Đánh giá độ đặc hiệu của mô hình PLS

Phân tích hệ số hồi quy của mô hình PLS cho thấy các đỉnh đóng góp dương cao nhất xuất hiện tại các số sóng: 930 $cm^{-1}$, 1345 $cm^{-1}$ và 1415 $cm^{-1}$. Đây là các pic dao động đặc trưng của nhóm chức trong kali acetat (dao động C-C và $COO^-$). Ngược lại, các vùng số sóng đặc trưng của glycerin (850, 1060, 1115 $cm^{-1}$) có hệ số hồi quy gần như bằng 0 hoặc đóng góp âm, cho thấy mô hình đã học cách phân biệt tín hiệu của kali acetat với nền glycerin. Điều này chứng minh độ đặc hiệu (specificity) tốt của mô hình PLS theo yêu cầu của hướng dẫn ICH Q2(R2).

### 3.6. Hiệu suất hệ thống tự động hóa và quản lý phòng xét nghiệm

Hệ thống tự động hóa xử lý trung bình 96 mẫu/ngày (12 mẫu/giờ), gấp 6 lần phương pháp thủ công (16 mẫu/ngày), tương đương giảm 85% thời gian phân tích.

### 3.7. Đánh giá trên mẫu thực tế và xác nhận phương pháp

**Bảng 3. So sánh phương pháp PLS-Raman và chuẩn độ trên 10 mẫu thực tế**

| Tên mẫu | PLS-Raman (%, X±SD, n=3) | Chuẩn độ (%, X±SD, n=3) | Sai số tuyệt đối (%) |
|---|---|---|---|
| TN01 | 30,68 ± 0,35 | 30,58 ± 0,23 | 0,1 |
| TN02 | 31,02 ± 0,28 | 30,98 ± 0,13 | 0,04 |
| TN03 | 31,21 ± 0,22 | 31,17 ± 0,12 | 0,04 |
| TN04 | 30,18 ± 0,41 | 29,81 ± 0,10 | 0,37 |
| TN05 | 30,15 ± 0,30 | 30,10 ± 0,17 | 0,05 |
| TN06 | 30,19 ± 0,25 | 29,31 ± 0,24 | 0,88 |
| TN07 | 30,68 ± 0,32 | 30,24 ± 0,23 | 0,44 |
| TN08 | 30,51 ± 0,28 | 30,26 ± 0,16 | 0,25 |
| TN09 | 30,75 ± 0,55 | 29,11 ± 0,18 | 1,64 |
| TN10 | 30,08 ± 0,48 | 29,25 ± 0,11 | 0,83 |

Các mẫu dung dịch bảo quản tại Viện 69 được sản xuất theo quy trình chuẩn với nồng độ kali acetat mục tiêu là 30% ± 2%. Đây là quy định kỹ thuật của sản phẩm, do đó tất cả các mẫu thực tế đều tập trung trong khoảng 29-31%, giải thích cho sự phân bố hẹp của dữ liệu.

Kết quả phân tích: Hệ số tương quan Pearson: $r = 0,985$ ($p < 0,001$). Phân tích Bland-Altman: Sự khác biệt trung bình (bias) là 0,46%, với giới hạn đồng thuận 95% từ -0,57% đến 1,49%. Ngưỡng sai số chấp nhận được: Đối với dung dịch bảo quản mẫu sinh học, yêu cầu về nồng độ kali acetat thường được chấp nhận trong khoảng ±5% so với giá trị danh định (tương đương ±1,5% đối với nồng độ 30%). Sai số trung bình 0,46% và sai số lớn nhất 1,64% (mẫu TN09) nằm trong ngưỡng chấp nhận này. Tuy nhiên, đối với các ứng dụng đòi hỏi độ chính xác cao hơn, cần lưu ý đến các mẫu có sai số lớn và có thể cần kiểm tra lại hoặc loại bỏ nếu vượt ngưỡng quy định.

Đánh giá độ đúng (theo ICH Q2(R2)): Trên mẫu chuẩn (n=10, trải dài 10-40%): Tỷ lệ thu hồi trung bình đạt 99,8% (dao động 98,5-101,2%). Trên mẫu thực tế được thêm chuẩn ở ba mức (thêm 5%, 10%, 15%): Tỷ lệ thu hồi trung bình lần lượt là 98,7%, 99,2% và 100,1%, cho thấy phương pháp có độ đúng tốt trên nền mẫu thực tế.

Đánh giá độ chụm (theo ICH Q2(R2)): RSD% của 6 lần đo lặp trên cùng một mẫu ở ba mức nồng độ 10%, 25%, và 40% lần lượt là 1,8%, 1,2%, và 0,9%, tất cả đều dưới 2%, cho thấy độ lặp lại tốt.

## 4. BÀN LUẬN

Nghiên cứu đã chứng minh tính khả thi của việc kết hợp phổ Raman và học máy (PLS) để xây dựng hệ thống phân tích tự động dung dịch bảo quản mẫu sinh học. Mô hình PLS với 6 thành phần chính đạt hiệu suất cao ($R^2 > 0,999$; MAPE = 1,24%), đồng thời bảo đảm khả năng giải thích hóa học và tốc độ xử lý phù hợp cho ứng dụng thực tế. Mặc dù ANN cho độ chính xác nhỉnh hơn, PLS được ưu tiên do nguy cơ quá khớp thấp hơn và khả năng giải thích tốt hơn trên bộ dữ liệu 240 mẫu.

Phân tích hệ số hồi quy cho thấy mô hình tập trung vào các pic Raman đặc trưng của kali acetat (930, 1345

và 1415 $cm^{-1}$), khẳng định tính đặc hiệu và hạn chế ảnh hưởng của các thành phần nền như glycerin. Hệ thống giúp rút ngắn thời gian phân tích từ 20–30 phút xuống dưới 1 phút mỗi mẫu, góp phần nâng cao năng suất và giảm sai sót trong kiểm soát chất lượng thường quy.

Kết quả xác nhận theo ICH Q2(R2) cùng hệ số tương quan cao với phương pháp chuẩn độ điện thế (r = 0,985) và phân tích Bland–Altman cho thấy phương pháp PLS-Raman có độ tin cậy cao và tiềm năng thay thế phương pháp truyền thống. Tuy nhiên, nghiên cứu còn hạn chế về số lượng mẫu thực tế, phạm vi thành phần nền và khả năng chuyển giao giữa các thiết bị Raman khác nhau; đây là những hướng cần tiếp tục nghiên cứu trong tương lai [7].

## 5. KẾT LUẬN

Nghiên cứu đã xây dựng thành công hệ thống phân tích tự động dựa trên phổ Raman kết hợp mô hình PLS để định lượng kali acetat trong dung dịch bảo quản. Mô hình PLS với 6 thành phần chính đạt độ chính xác cao ($R^2$ = 0,9991; MAPE = 1,24%), có khả năng giải thích hóa học và đáp ứng các yêu cầu cơ bản về độ đúng, độ chụm và độ đặc hiệu theo ICH Q2(R2). Quy trình tự động hóa giúp rút ngắn thời gian phân tích khoảng 85%, nâng cao năng suất phòng xét nghiệm và giảm sai sót thao tác. Kết quả trên mẫu thực tế cho thấy sự tương đồng cao với phương pháp chuẩn độ điện thế (r = 0,985), với sai số nằm trong giới hạn chấp nhận đối với kiểm soát chất lượng dung dịch bảo quản. Nghiên cứu cung cấp cơ sở khoa học cho việc ứng dụng phổ Raman kết hợp học máy trong kiểm soát chất lượng, góp phần thúc đẩy chuyển đổi số tại các cơ sở y tế và lưu trữ mẫu sinh học.

LỜI CẢM ƠN

Công trình có sự hỗ trợ kinh phí của Nhiệm vụ KHCN cấp cơ sở "*định lượng glycerin và kali acetat trong dung dịch ướp bảo quản bằng phương pháp quang phổ raman*". Mã số nhiệm vụ: NVCS.25_03; thực hiện năm 2025-2026. Thuộc Bộ Tư Lệnh Lăng Chủ tịch Hồ Chí Minh.

## 6. TÀI LIỆU THAM KHẢO

[1] Smith E, Dent G. *Modern Raman Spectroscopy: A Practical Approach*. 2nd ed. Hoboken, NJ: Wiley; 2019.

[2] McCreery RL. *Raman Spectroscopy for Chemical Analysis*. New York: John Wiley & Sons; 2000.

[3] Lussier F, Thibault V, Charron B, et al. Deep learning for biomedical Raman spectroscopy: a review. *Analyst*. 2024;149(8):2145-2167. doi:10.1039/D3AN01113B.

[4] Liao S, Huang J, Liu C, et al. Machine Learning-Assisted Quantification of Biomarkers Using Surface-Enhanced Raman Spectroscopy. *Anal Chem*. 2025;97(1):100-108. doi:10.1021/acs.analchem.4c02456.

[5] Xu Y, Li L, Shen A, et al. Intelligent baseline correction for Raman spectroscopy via deep learning-based parameter optimization. *Talanta*. 2026;260:124890. doi:10.1016/j.talanta.2026.124890.

[6] Zhang ZM, Chen S, Liang YZ. Baseline correction using adaptive iteratively reweighted penalized least squares. *Analyst*. 2010;135(5):1138-1146. doi:10.1039/b922045c.

[7] Silva CS, Pimentel MF. Review of model transfer strategies in near-infrared and Raman spectroscopy. *Microchem J*. 2024;199:109980. doi:10.1016/j.microc.2024.109980.

[8] Lê Minh T, Phạm Văn H. *Phương pháp phân tích và kiểm nghiệm dược phẩm*. Hà Nội: Nhà xuất bản Y học; 2021.

[9] Valderrey V, Perez M, Rodriguez L. Application of Raman Spectroscopy Coupled With Chemometrics for the Detection and Quantification of Pesticide Residues. *ChemistryEurope*. 2025;3(2):e202400120. doi:10.1002/chem.202400120.

[10] International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *Validation of Analytical Procedures: Text and Methodology Q2(R2)*. Geneva: ICH; 2024.