



DOI:10.22144/ctujos.2026.088

## THIẾT KẾ VÀ ĐÁNH GIÁ TÍNH KHẢ THI CỦA MÔ ĐUN THU PHỔ TƯƠNG TÁC TỰ ĐỘNG CHO ĐÁNH GIÁ NHANH ĐỘ NGỌT CỦA QUẢ QUÝT

Lâm Văn Linh<sup>1,2</sup>, Nguyễn Đình Trí<sup>1</sup>, Lê Thành Trung<sup>1</sup>, Nguyễn Hoàng Tiên<sup>1</sup>, Phan Thanh Lương<sup>1</sup>, Trần Nhựt Thanh<sup>1</sup> và Nguyễn Chánh Nghiệm<sup>1,\*</sup>

<sup>1</sup>Trường Bách Khoa, Đại học Cần Thơ, Việt Nam

<sup>2</sup>Khoa Điện – Điện tử, Trường Cao đẳng Kiên Giang, Việt Nam

\*Tác giả liên hệ (Corresponding author): [ncngkiem@ctu.edu.vn](mailto:ncngkiem@ctu.edu.vn)

### Thông tin chung (Article Information)

Nhận bài (Received): 11/03/2026

Sửa bài (Revised): 18/03/2026

Duyệt đăng (Accepted): 12/05/2026

**Title:** Design and feasibility evaluation of an automated interactance spectral acquisition module for rapid mandarin sweetness assessment

**Author(s):** Lam Van Linh<sup>1,2</sup>, Nguyen Dinh Tri<sup>1</sup>, Le Thanh Trung<sup>1</sup>, Nguyen Hoang Tien<sup>1</sup>, Phan Thanh Luong<sup>1</sup>, Tran Nhut Thanh<sup>1</sup> and Nguyen Chanh Nghiem<sup>1,\*</sup>

**Affiliation(s):** <sup>1</sup>College of Engineering, Can Tho University, Viet Nam; <sup>2</sup>Faculty of Electric-Electronics, Kien Giang College, Viet Nam

### TÓM TẮT

Nghiên cứu trình bày thiết kế và đánh giá tính khả thi của mô đun thu phổ tương tác tự động dùng quang phổ cận hồng ngoại (NIR) để đánh giá độ ngọt quả quýt. Hệ thống tích hợp cơ cấu kẹp–xoay mẫu, đầu thu tịnh tiến với nguồn sáng vòng và buồng tối, cho phép thu phổ tự động tại nhiều vùng đo. So với thu phổ thủ công, thời gian đo giảm từ ~40 s xuống 17,7 s mỗi quả, tương ứng giảm 55,75%, đồng thời cải thiện độ ổn định phổ. Mô hình dự đoán độ ngọt (°Brix) từ 300 phổ trung bình của 75 quả quýt đạt RPD cao nhất 1,65, cho thấy tiềm năng ứng dụng của hệ thống trong đánh giá nhanh, không phá hủy chất lượng trái cây.

**Từ khóa:** Đánh giá chất lượng không phá hủy, đo phổ tự động, độ ngọt của quýt, quang phổ cận hồng ngoại, quang phổ tương tác

### ABSTRACT

This study presents the design and feasibility evaluation of an automated NIR interactance spectral acquisition module for mandarin sweetness assessment. The system integrates a clamping–rotation mechanism, a translational probe with a ring light source, and a dark chamber to enable automated measurements at multiple fruit regions. Compared with manual acquisition, the average measurement time was reduced from ~40 s to 17.7 s per fruit, corresponding to a 55.75% reduction, while spectral stability was improved. Sweetness prediction models developed from 300 averaged spectra of 75 mandarins achieved a maximum RPD of 1.65, indicating the system's potential for rapid and non-destructive fruit quality evaluation.

**Keywords:** Automated spectral measurement, interactance spectroscopy, mandarin sweetness, near-infrared spectroscopy, non-destructive quality assessment

## 1. GIỚI THIỆU

Việt Nam là quốc gia có điều kiện khí hậu nhiệt đới thuận lợi cho sự phát triển của nhiều loại cây ăn quả, tại những vùng miền khác nhau lại có đa dạng

các loại trái cây đặc sản, đặc trưng riêng và quả quýt là một trong số đó. Tại Đồng bằng sông Cửu Long, quýt đường là một trong những giống cam quýt có chất lượng ngon, nổi tiếng của vùng. Việc đánh giá chất lượng quýt dựa vào độ ngọt thông qua hàm

lượng chất rắn hòa tan (Soluble Solids Content – SSC) có vai trò quan trọng trong phân loại, nâng cao chất lượng trong thương mại hóa sản phẩm. Phương pháp đánh giá độ ngọt bằng khúc xạ kế có kết quả đang được sử dụng rộng rãi. Tuy nhiên, phương pháp này tốn thời gian, phụ thuộc vào môi trường, không phù hợp cho kiểm tra số lượng lớn và cần phải ép lấy nước trái cây (Magwaza & Opara, 2015). Trong bối cảnh đó, các kỹ thuật đánh giá không phá hủy ngày càng được quan tâm (Nguyen et al., 2020; Nghiệm và ctv., 2021) đặc biệt là phương pháp quang phổ cận hồng ngoại (Near-Infrared – NIR) nhờ khả năng phản ánh chính xác đặc tính hóa học và cấu trúc bên trong của quả mà không cần phá hủy mẫu đo.

Kỹ thuật đánh giá không phá hủy dựa trên quang phổ cận hồng ngoại (750–2500 nm) là một phương pháp sử dụng quang phổ NIR tương tác với quả để phân tích và đo lường độ ngọt của quả quýt sau thu hoạch mà không cần phá hủy mẫu. Khi có một nguồn ánh sáng chiếu tới quả quýt, ánh sáng tác động lên các liên kết C–H, N–H, O–H có trong các chất hữu cơ liên quan tới độ ngọt. Tín hiệu phổ thu được mang các đặc trưng do có tương tác với các chất hữu cơ bên trong của quả (Ozaki et al., 2021). Những giá trị phổ thu được của từng mẫu được sử dụng để huấn luyện các mô hình học máy để phân tích các đặc trưng của phổ, sau đó cho ra một mô hình có khả năng dự đoán được giá trị độ ngọt của quả dựa trên dữ liệu phổ đầu vào (Mei & Li, 2023).

Trong kỹ thuật quang phổ, có ba chế độ thu quang phổ phổ biến là phản xạ (reflectance), tương tác (interactance) và xuyên thấu (transmittance). Ở chế độ đo phản xạ, ánh sáng chiếu từ bên ngoài lên mẫu, đầu thu phổ không cần cách ly và không tiếp xúc với mẫu. Chế độ xuyên thấu yêu cầu một nguồn ánh sáng mạnh hơn để ánh sáng có thể xuyên qua mẫu và đi vào đầu thu phổ ở hướng đối diện, đầu thu phổ có thể không cần tiếp xúc với mẫu. Trong chế độ phản xạ, ánh sáng chiếu vào mẫu, đầu thu phổ được đặt cạnh bên nguồn sáng, cách ly với nguồn sáng và áp sát vào mẫu. Chế độ tương tác có sự kết hợp giữa phản xạ và xuyên (Schaare & Fraser, 2000; Pasquini, 2003). Chế độ tương tác có yêu cầu kỹ thuật cao hơn hai chế độ thu phổ còn lại. Tuy nhiên, trong nghiên cứu trên quả kiwi của Schaare and Fraser (2000) chế độ tương tác có hiệu quả tốt nhất trong mọi trường hợp dự đoán SSC, mật độ (density) và góc sắc thái thịt quả (Flesh Hue Angle) của quả kiwi vàng *Actinidia chinensis* tại thời điểm thu hoạch. Do đó, chế độ thu phổ tương tác được sử dụng trong nghiên cứu này để thu phổ và thiết kế mô đun thu phổ tương tác tự động.

Trong nghiên cứu này, một mô đun thu phổ tương tác tự động được thiết kế và phát triển nhằm chuẩn hóa điều kiện đo và cải thiện tính lặp lại của dữ liệu phổ thu được từ quả quýt sau thu hoạch. Nghiên cứu tập trung đánh giá tính khả thi bước đầu của mô đun thông qua ba tiêu chí chính: (i) hiệu suất cải thiện của quy trình thu phổ tự động so với phương pháp thủ công, (ii) độ ổn định và độ lặp lại của phổ thu được và (iii) khả năng hỗ trợ xây dựng các mô hình dự đoán độ ngọt từ dữ liệu phổ (Sun et al., 2014; Santos et al., 2021; Luo et al., 2024; Xiao et al., 2025). Với mục tiêu của nghiên cứu là đánh giá khả năng hoạt động của mô đun thu phổ tự động, phạm vi khảo sát mô hình dự đoán được giới hạn ở một số phương pháp tiền xử lý phổ phổ biến và hai mô hình học máy cơ bản nhằm kiểm chứng khả năng khai thác dữ liệu phổ của mô đun đề xuất trong các hệ thống phân loại chất lượng trái cây tự động dựa trên phân tích phổ cận hồng ngoại.

## 2. PHƯƠNG PHÁP NGHIÊN CỨU

### 2.1. Mô đun thu phổ tương tác

Để đánh giá tính khả thi của mô đun thu phổ tương tác tự động được đề xuất, các phép thu phổ được thực hiện bằng hai cấu hình: mô đun thu phổ vận hành thủ công và mô đun thu phổ tự động. Mô đun đo thủ công được sử dụng làm cơ sở so sánh, trong khi mô đun tự động được thiết kế nhằm chuẩn hóa điều kiện đo và cải thiện độ lặp lại của quá trình thu phổ.

#### 2.1.1. Mô đun thu phổ thủ công

Việc thu phổ tương tác thủ công được thực hiện bằng mô-đun thu phổ thủ công, được bố trí như thể hiện trong Hình 1. Mô đun gồm bốn thành phần chính: phổ kế NIRScan Nano EVM (1), nguồn sáng halogen MI-152 (2), đầu thu quang phổ (3) và đầu ra nguồn sáng dạng vòng (ring light) (4).

Nguồn sáng halogen MI-152 (Dolan-Jenner) có công suất 150 W được sử dụng để cung cấp nguồn chiếu sáng ổn định với phổ nằm trong dải bước sóng phù hợp với phổ kế được sử dụng. Công suất của nguồn sáng đủ lớn để ánh sáng có thể thâm nhập một phần vào mô quả, đáp ứng yêu cầu kỹ thuật của phép đo phổ theo chế độ tương tác. Ánh sáng từ nguồn được dẫn tới mẫu thông qua dây dẫn quang MA4 Ring light, trong đó đầu ra dạng vòng cho phép chiếu sáng đồng đều lên vùng đo. Tín hiệu phổ tương tác từ mẫu được thu nhận bởi đầu thu quang phổ và truyền về phổ kế thông qua dây dẫn quang LIG5-4Z (Thorlabs). Vì kích thước đầu thu sợi quang tương đối nhỏ nên bề mặt đầu thu có thể áp sát lên bề mặt quả quýt với kích thước khác nhau.

Phần mềm sử dụng trong nghiên cứu là DLP NIRScan Nano EVM (Texas Instruments), có khả năng thu phổ trong dải bước sóng 900–1700 nm.

Quy trình thu thập phổ bắt đầu bằng việc xác định bốn vùng đo cách đều nhau trên đường xích đạo của quả. Vùng đo đầu tiên được chọn ngẫu nhiên bởi một điểm trên đường xích đạo, từ đó hai điểm đo kế tiếp cho vùng đo được xác định trong đó hai điểm đo liền kề cách nhau khoảng 1 cm (Hình 2a). Phần trung bình từ 3 điểm đo này được sử dụng để xây dựng mô hình dự đoán độ ngọt vì phổ trung bình có thể giúp giảm nhiễu đo và tăng cường tính đại diện của phổ tại vùng đo (Manley, 2014; Bec et al., 2025; Xiao et al., 2025).

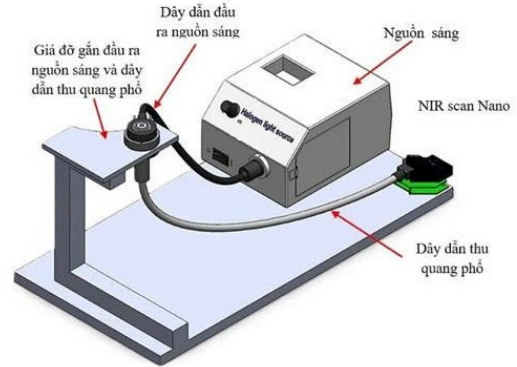
Trong quá trình thu phổ, quả được giữ bằng tay và áp sát bề mặt vào đầu thu quang phổ để đảm bảo thu tín hiệu theo chế độ tương tác và hạn chế ánh sáng môi trường xâm nhập vào đầu thu. Quá trình thu phổ được điều khiển bằng phần mềm NIRScan NanoEVM GUI và mỗi phổ sau khi thu được lưu dưới dạng tệp CSV để phục vụ cho quá trình tổng hợp dữ liệu. Quy trình này được thực hiện cho ba điểm đo tại vùng đo đầu tiên và được lặp lại cho các vùng đo tiếp theo trên đường xích đạo của quả.

Mô đun thu phổ thủ công có cấu trúc đơn giản và dễ triển khai trong các thí nghiệm ban đầu. Tuy nhiên, độ chính xác và tính đồng nhất của dữ liệu phổ vẫn phụ thuộc đáng kể vào thao tác của người vận hành. Thời gian trung bình để thu phổ tại ba điểm trên một quả khoảng 40 s, trong đó phần lớn thời gian được sử dụng để xác định vị trí điểm cần đo và nhập thông tin tệp dữ liệu. Do đó, khi cần thu phổ tại nhiều điểm lân cận để tính giá trị trung bình, quá trình thu mẫu trở nên tốn thời gian và làm giảm hiệu suất thu thập dữ liệu. Bên cạnh đó, trong quá trình đo thủ công, người vận hành phải thao tác lặp lại nhiều lần với nguồn sáng có cường độ cao, có thể gây khó chịu cho mắt khi làm việc trong thời gian dài. Sau một thời gian thao tác liên tục, các yếu tố chủ quan như lực tiếp xúc giữa đầu dò và mẫu, góc đặt mẫu và độ ổn định của thao tác có thể thay đổi, dẫn đến sự sai lệch trong dữ liệu phổ thu được. Vì vậy, việc phát triển mô đun thu phổ tương tác tự động là cần thiết nhằm chuẩn hóa điều kiện đo, giảm ảnh hưởng của các yếu tố chủ quan từ người vận hành và nâng cao hiệu suất thu thập dữ liệu phổ.

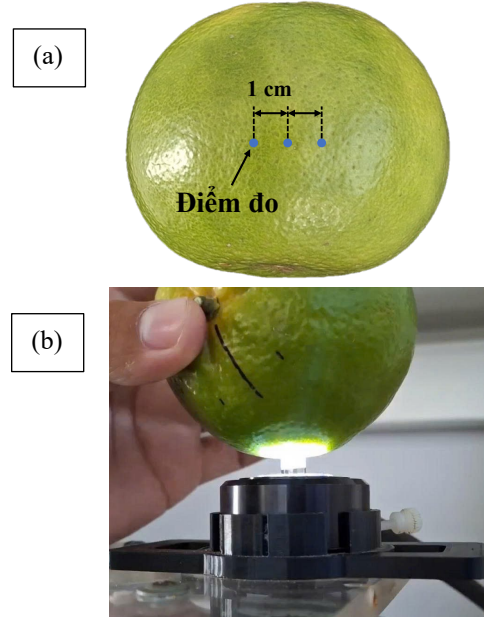
### 2.1.2. Mô đun thu phổ tự động

Mô hình thiết kế 3D của mô đun thu phổ tương tác tự động được trình bày trong Hình 3. Mô đun được bố trí trong buồng tối (1) nhằm cách ly ánh sáng môi trường và bảo đảm điều kiện chiếu sáng ổn

định trong quá trình đo. Bên trong buồng tối, mô đun thu phổ gồm hai cụm chức năng chính: mô đun kẹp và xoay mẫu và mô đun thu phổ tích hợp nguồn sáng.



**Hình 1. Mô đun thu phổ theo chế độ tương tác bằng phương pháp thủ công**

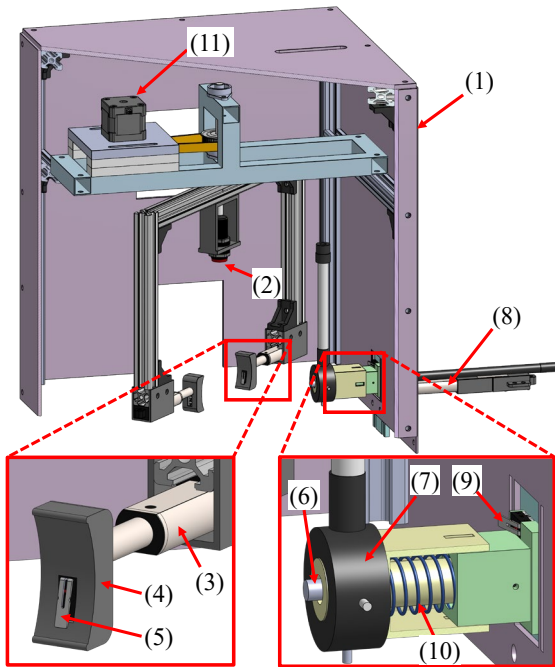


**Hình 2. (a) Đánh dấu điểm thu phổ (b) Thu phổ theo chế độ tương tác trên mô đun thủ công**

Trong mô đun kẹp và xoay, cảm biến tiệm cận (2) được sử dụng để phát hiện sự hiện diện của quả trong vùng đo. Hai xi lanh điện (5) điều khiển các đầu kẹp cong (3) áp sát vào quả để cố định mẫu trong quá trình thu phổ. Công tắc hành trình tích hợp trên đầu kẹp (4) xác định thời điểm tiếp xúc với bề mặt quả và gửi tín hiệu dừng cơ cấu chấp hành nhằm đảm bảo lực kẹp ổn định và tránh gây hư hỏng mẫu.

Mô đun thu phổ gồm đầu thu ánh sáng tương tác (6) kết nối với phổ kế NIRScan Nano EVM thông

qua dây dẫn quang lỏng (liquid light guide – LLG5-4Z) và nguồn sáng dạng vòng (7), được lắp trên cơ cấu tịnh tiến dẫn động bởi xi lanh điện (8). Cơ cấu này cho phép đầu thu áp sát bề mặt quả để thực hiện phép đo phổ theo chế độ tương tác. Công tác hành trình (9) kết hợp với lò xo giảm chấn (10) được sử dụng để giới hạn hành trình tịnh tiến và giảm lực ép lên bề mặt quả, đảm bảo điều kiện tiếp xúc ổn định trong quá trình thu phổ.



**Hình 3. Thiết kế 3D của mô đun thu phổ tương tác tự động**

Ghi chú: (1) buồng tối, (2) cảm biến tiệm cận, (3) đầu tay kẹp, (4) xi lanh tay kẹp, (5) công tắc hành trình đầu tay kẹp, (6) đầu thu phổ tương tác, (7) nguồn sáng vòng, (8) xi lanh đầu thu phổ, (9) công tắc hành trình cụm thu phổ, (10) lò xo, (11) động cơ xoay tay kẹp.

Kết quả được thể hiện ở Hình 4 cho thấy lưu đồ hoạt động của hệ thống thu phổ tương tác tự động. Quả quýt được đặt vào giữa mô đun kẹp cố định và xoay. Khi cảm biến tiệm cận trên mô đun kẹp và xoay phát hiện có quả quýt, tín hiệu kích hoạt được gửi về vi điều khiển, hai xi lanh điện đồng thời áp sát hai đầu kẹp vào quả quýt. Khi cả hai đầu kẹp tiếp

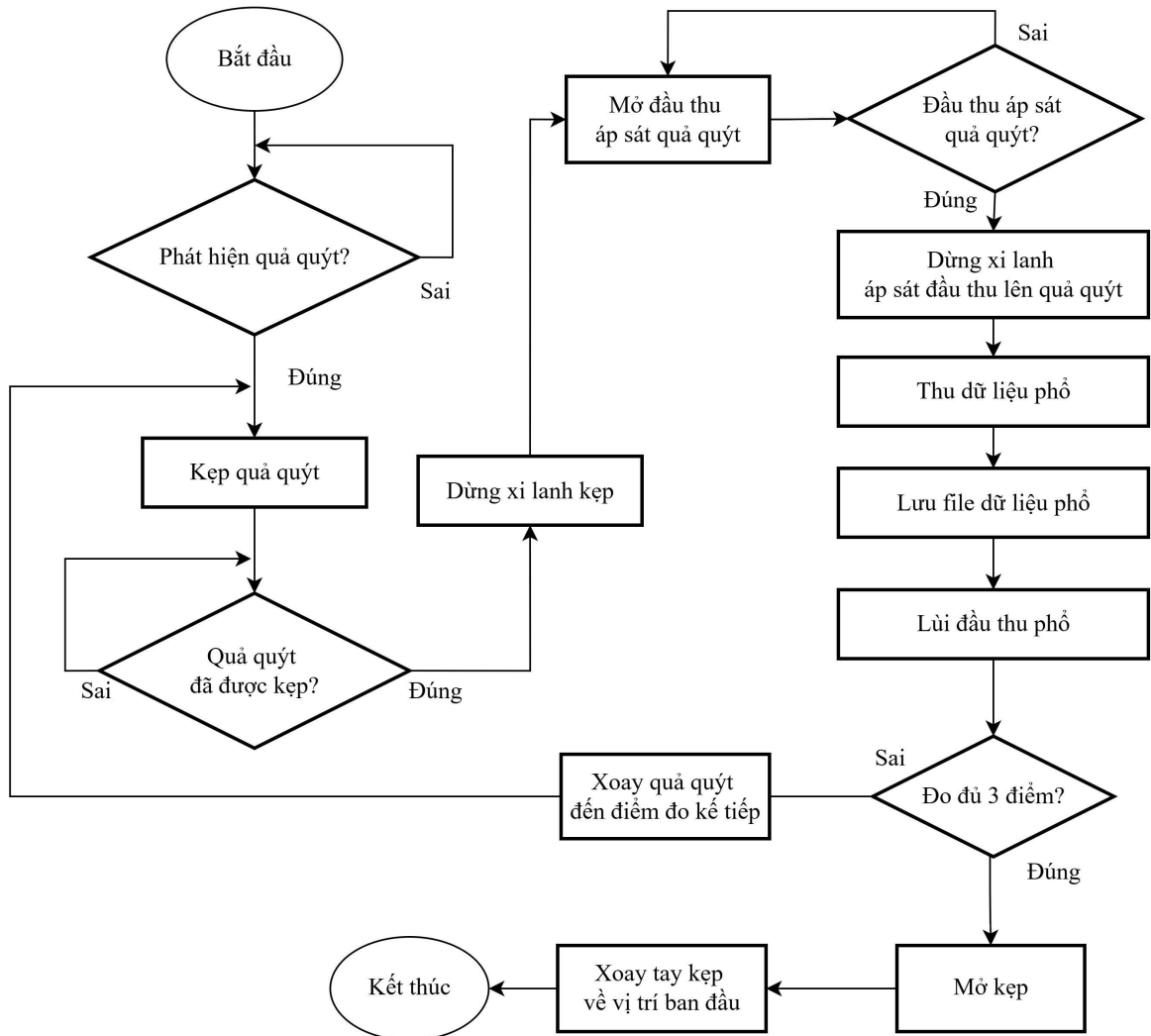
xúc với quả quýt, hai công tắc hành trình ngay lập tức gửi tín hiệu dừng hai xi lanh, giữ lực kẹp ổn định, tránh gây hư hỏng quả quýt. Sau khi quả quýt được giữ cố định, xi lanh điện của mô đun tích hợp đầu thu phổ và đầu ra nguồn sáng được kích hoạt. Đầu thu được áp sát lên bề mặt quả quýt để thu phổ tương tác. Sau khi hoàn thành thu phổ tại điểm đo đầu tiên, xi lanh điện này lại được kích hoạt để kéo đầu thu phổ ra khỏi bề mặt quả quýt một đoạn đủ ngắn (tương ứng khoảng thời gian 0,5 s) để tay kẹp có thể xoay quả quýt tới điểm đo kế tiếp. Quá trình này được lặp lại để thu thêm phổ của 2 điểm liền kề kế tiếp, trong đó 2 điểm đo liền kề cách nhau 1 cm. Phổ trung bình xác định từ phổ của 3 điểm đo liền kề này được sử dụng để dự đoán độ ngọt quả quýt trong thực tế ứng dụng.

Tuy nhiên, để có nhiều mẫu dữ liệu phổ phục vụ cho việc xây dựng mô hình dự đoán độ ngọt phổ trung bình được thu thập trên bốn vùng đo cách đều nhau trên đường xích đạo của quả quýt. Vì thế, sau khi thu đủ phổ tại 3 điểm của một vùng đo, động cơ xoay tay kẹp được kích hoạt để xoay quả quýt đến điểm đo phổ của vùng đo phổ tiếp theo. Do bốn vùng đo cách đều nhau trên đường xích đạo, vùng đo tiếp theo được xác định bằng cách điều khiển động cơ xoay tay kẹp một góc 90 độ. Quá trình thu phổ lặp lại cho đến khi thu đủ phổ của bốn vùng đo. Khi phổ của bốn vùng đo được thu thập, đầu thu phổ được kéo ra khỏi bề mặt quả quýt. Đồng thời, tay kẹp được xoay về vị trí ban đầu.

## 2.2. Thu thập dữ liệu phổ

Để làm rõ đặc điểm của bộ dữ liệu phục vụ cho quá trình phân tích, thông tin chi tiết về dữ liệu phổ thu thập bằng phương pháp đo thủ công và bằng mô đun thu phổ tự động được trình bày trong Bảng 1. Tổng cộng 75 quả quýt được sử dụng trong thí nghiệm thu phổ. Đối với mỗi quả, phổ được thu tại bốn vùng đo khác nhau, do đó mỗi phương pháp thu phổ tạo ra 300 mẫu phổ trung bình. Hai bộ dữ liệu được ký hiệu lần lượt là bộ dữ liệu A (đo thủ công) và bộ dữ liệu B (đo tự động).

Khi xây dựng mô hình học máy, mỗi bộ dữ liệu được chia thành hai phần theo tỉ lệ 8:2, bao gồm tập hiệu chuẩn (calibration set) dùng để xây dựng mô hình và tập dự đoán (prediction set) dùng để đánh giá khả năng dự đoán của mô hình.



Hình 4. Lưu đồ hoạt động của mô đun thu phổ tương tác tự động

Bảng 1. Thông tin bộ dữ liệu sử dụng trong thí nghiệm

Bộ dữ liệu	Ngày đo	Phương pháp đo	Số quả	Số mẫu
A	20/07/2025	Đo thủ công	25	100
	21/07/2025	Đo thủ công	25	100
	22/07/2025	Đo thủ công	25	100
B	20/07/2025	Đo tự động	25	100
	21/07/2025	Đo tự động	25	100
	22/07/2025	Đo tự động	25	100

Khi thu thập phổ tương tác bằng mô đun thu phổ tự động, điểm thu phổ ban đầu của vùng thu phổ đầu tiên cũng được xác định ngẫu nhiên. Dựa vào khảo sát, đường kính trung bình của quả quýt đường được sử dụng để thu phổ xấp xỉ 6,7 cm. Đặt  $l = 1$  cm là

khoảng cách giữa điểm đầu tiên và một điểm liền kề,  $R \approx 3,35$  cm là bán kính trung bình của quả quýt và  $\alpha$  là góc quay tính từ tâm của quả quýt để khoảng cách hai điểm đo là  $l = 1$  cm. Góc quay giữa hai điểm liền kề nhau được tính theo công thức sau:

$$\alpha = \frac{180 \times l}{r \times \pi} = \frac{180 \times 1}{3,35\pi} \approx 17,1^\circ \quad (1)$$

Khung kẹp được quay nhờ bộ truyền động đai với tỉ số truyền  $i = 1/2$  và sử dụng động cơ bước với số bước mỗi vòng là  $N = 800$  bước. Để quay quả quýt một góc  $\alpha$  thì số bước động cơ cần quay được tính như sau:

$$n = \frac{N}{360i} \alpha = \frac{800 \times 17,1}{360 \times 2} = 19 \quad (2)$$

### 2.3. Đánh giá độ tương đồng của phổ

Để đánh giá bước đầu độ ổn định của quá trình thu phổ, một khảo sát sơ bộ được thực hiện dựa trên các phép đo lặp liên tiếp tại cùng một vùng đo. Tại một điểm đo, phổ tương tác được thu bốn lần liên tiếp trong cùng một phiên đo, sử dụng cùng phổ kế NIRScan Nano EVM. Do mục tiêu của thí nghiệm này chỉ nhằm quan sát mức độ dao động ngắn hạn giữa các lần thu phổ, số lần đo lặp được giới hạn ở bốn lần. Việc phân tích này không nhằm đánh giá đầy đủ độ lặp lại của hệ thống mà chỉ cung cấp một chỉ báo ban đầu về tính ổn định của phép đo.

Độ tương đồng giữa các phổ đo lặp được đánh giá thông qua độ lệch chuẩn của biên độ tại đỉnh phổ có cường độ lớn nhất trong dải bước sóng khảo sát. Trước hết, bước sóng tương ứng với đỉnh phổ cao nhất được xác định trên phổ trung bình của các lần đo lặp. Sau đó, giá trị biên độ tại bước sóng này được trích xuất từ từng phổ đo lặp và dùng để tính độ lệch chuẩn. Giá trị độ lệch chuẩn càng nhỏ cho thấy biên độ phổ giữa các lần đo có sự dao động càng thấp, từ đó phản ánh mức độ ổn định cao hơn của quá trình thu phổ.

Độ lệch chuẩn được tính theo công thức:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (3)$$

Trong đó,  $x_i$  là biên độ phổ tại bước sóng đỉnh của lần đo thứ  $i$ ,  $\bar{x}$  là giá trị trung bình của các biên độ và  $n$  là số lần đo lặp. Giá trị  $\sigma$  càng nhỏ cho thấy các phổ đo lặp có mức độ tương đồng cao hơn.

### 2.4. Xây dựng mô hình dự đoán độ ngọt

Trong nghiên cứu này, việc xây dựng mô hình dự đoán chỉ nhằm đánh giá bước đầu khả năng sử dụng dữ liệu phổ thu được từ mô đun thu phổ tự động. Do đó, phạm vi khảo sát được giới hạn ở một số phương pháp tiền xử lý phổ phổ biến và hai loại mô hình hồi quy cơ bản là PLS (Partial Least Squares) và SVR (Support Vector Regression). Các mô hình được xây dựng lần lượt trên hai bộ dữ liệu: phổ đo thủ công (bộ dữ liệu A) và phổ đo tự động (bộ dữ liệu B). Hiệu quả của mô đun đề xuất được đánh giá gián tiếp thông qua khả năng cải thiện hiệu suất dự đoán độ °Brix của các mô hình này.

Do hiệu suất dự đoán của các mô hình phụ thuộc đáng kể vào việc lựa chọn phương pháp tiền xử lý phổ, một số kỹ thuật tiền xử lý phổ phổ biến đã được áp dụng, bao gồm: biến chuẩn hóa (standard normal variate – SNV), hiệu chỉnh tán xạ nhân

(multiplicative scatter correction – MSC), lọc Savitzky–Golay (SG), đạo hàm bậc nhất sau khi lọc SG (first derivative – FD) và chuẩn hóa min–max (MM). Các công thức toán học của một số phương pháp tiền xử lý chính được sử dụng trong nghiên cứu này được trình bày như sau.

Giải thuật SNV được sử dụng để giảm ảnh hưởng của tán xạ và sự khác biệt về cường độ tín hiệu trong dữ liệu phổ (Mishra et al., 2020). Phương pháp này chuẩn hóa mỗi phổ bằng cách trừ giá trị trung bình và chia cho độ lệch chuẩn của phổ theo công thức (Lu et al., 2020):

$$x_{ij}(\text{SNV}) = \frac{x_{ij} - \bar{x}_i}{\sqrt{\frac{\sum_{i=1}^m (x_{ij} - \bar{x}_i)^2}{m-1}}} \quad (4)$$

trong đó  $x_{ij}$ ,  $x_{ij}(\text{SNV})$  lần lượt là biên độ tín hiệu tại bước sóng thứ  $j$  của phổ thứ  $i$  trước và sau khi xử lý SNV,  $m$  là số bước sóng và  $\bar{x}_i$  là giá trị trung bình của phổ thứ  $i$ , được xác định bởi:

$$\bar{x}_i = \frac{1}{m} \sum_{j=1}^m x_{ij} \quad (5)$$

Giải thuật MSC xem mỗi phổ là kết quả của một phép biến đổi tuyến tính của phổ tham chiếu, thường là phổ trung bình của tập dữ liệu. Các tham số dịch chuyển và hệ số tỉ lệ được ước lượng bằng hồi quy bình phương tối thiểu và được sử dụng để hiệu chỉnh phổ theo phương trình:

$$x_{corr} = \frac{1}{b}(x - a) \quad (6)$$

trong đó  $x$  là phổ ban đầu,  $x_{corr}$  là phổ sau khi hiệu chỉnh,  $a$  và  $b$  lần lượt là tham số dịch chuyển và hệ số tỉ lệ được xác định bởi thuật toán MSC (Mishra et al., 2020).

Trong dữ liệu phổ NIR, tín hiệu đo thường chứa nhiều nhiễu ngẫu nhiên và các dao động tần số cao do đặc tính của đầu dò và điều kiện đo. Do đó, các kỹ thuật làm trơn phổ (spectral smoothing) thường được áp dụng nhằm giảm nhiễu và cải thiện chất lượng tín hiệu (Beć et al., 2025). Lọc SG là phương pháp làm trơn dữ liệu phổ dựa trên việc xấp xỉ cục bộ các điểm dữ liệu bằng đa thức bậc thấp trong một cửa sổ trượt theo phương pháp bình phương tối thiểu (Zhang & Mouazen, 2023). Giá trị phổ sau khi lọc được tính theo công thức:

$$y_i = \sum_{k=-m}^m c_k y_{i+k}, \quad (7)$$

Trong đó:  $y_{i+k}$  và  $y_i$  là giá trị phổ ban đầu và sau khi lọc;  $c_k$  là hệ số lọc và  $(2m + 1)$  là kích thước của số lọc.

### 2.5. Chỉ tiêu đánh giá hiệu suất mô hình

Hiệu suất của các mô hình dự đoán độ ngọt được sử dụng để đánh giá hiệu quả của việc sử dụng mô đun thu phổ tự động và đo thủ công. Hiệu suất mô hình được đánh giá thông qua các chỉ số thống kê phổ biến, bao gồm hệ số tương quan ( $R$ ), căn bậc hai sai số trung bình bình phương (Root Mean Square Error – RMSE), sai số tuyệt đối trung bình (Mean Absolute Error – MAE) và tỉ số dự đoán đối với độ lệch (Ratio of Prediction to Deviation – RPD). Các chỉ số này được sử dụng rộng rãi trong đánh giá các mô hình dự đoán dựa trên dữ liệu quang phổ và được tính cho tập hiệu chuẩn (c) và tập dự đoán (p) (Lộc và ctv., 2024).

## 3. KẾT QUẢ VÀ THẢO LUẬN

### 3.1. Đánh giá hiệu suất cải thiện khi tự động hóa quy trình đo phổ tự động cho quả quýt sau thu hoạch

Để đánh giá hiệu suất hoạt động của mô đun thu phổ tự động, thời gian của từng bước trong chu trình đo được xác định thông qua các thí nghiệm lặp lại trong điều kiện vận hành thực tế của hệ thống. Thời gian của mỗi giai đoạn (kẹp mẫu, áp sát đầu thu phổ, thu phổ và xoay mẫu) được ghi nhận tự động trong quá trình điều khiển của hệ thống và tính trung bình từ 80 lần đo lặp. Do cơ cấu kẹp được điều khiển theo hành trình của xi lanh điện kết hợp với công tắc hành trình, thời gian kẹp mẫu chủ yếu phụ thuộc vào tốc độ tịnh tiến của cơ cấu chấp hành và gần như không

thay đổi đáng kể đối với các quả có kích thước khác nhau trong phạm vi thí nghiệm.

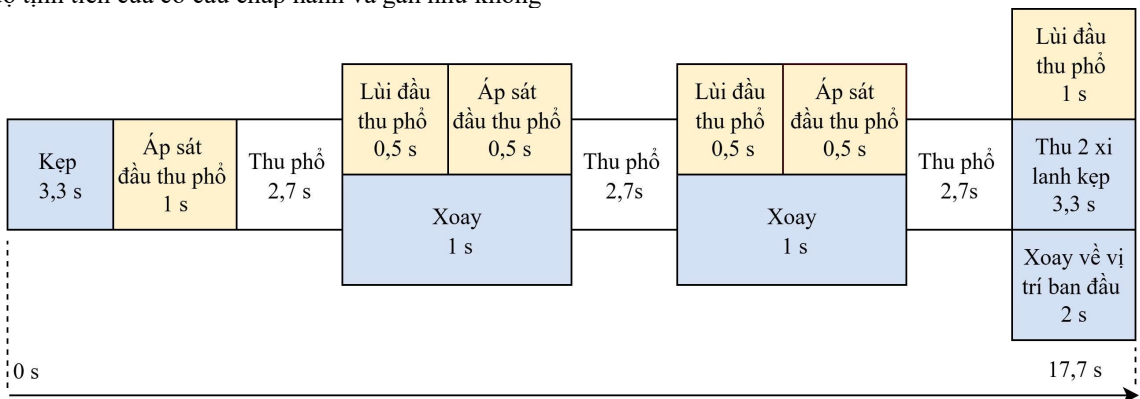
Trục thời gian hoạt động của mô đun thu phổ tương tác tự động được trình bày trong Hình 5. Tổng thời gian trung bình cần thiết để thu phổ tại ba điểm trên một quả quýt là 17,7 s, giảm 55,75% so với phương pháp thu phổ thủ công ( $\approx 40$  s). Chu trình đo bao gồm các bước chính: kẹp cố định mẫu ( $\approx 3,3$  s), áp sát đầu thu phổ ( $\approx 1$  s), thu phổ bằng phổ kế NIRScan Nano ( $\approx 2,7$  s) và xoay quả đến điểm đo tiếp theo. Các bước này được thực hiện tuần tự và tự động hóa hoàn toàn.

Trong phương pháp thủ công, phần lớn thời gian bị tiêu tốn vào việc xác định điểm thu phổ, thao tác áp sát mẫu vào đầu thu và nhập dữ liệu. Ngược lại, trong mô đun tự động, các bước này được chuẩn hóa và thực hiện theo một trình tự cố định, giúp rút ngắn đáng kể thời gian đo và loại bỏ sự phụ thuộc vào thao tác của người vận hành.

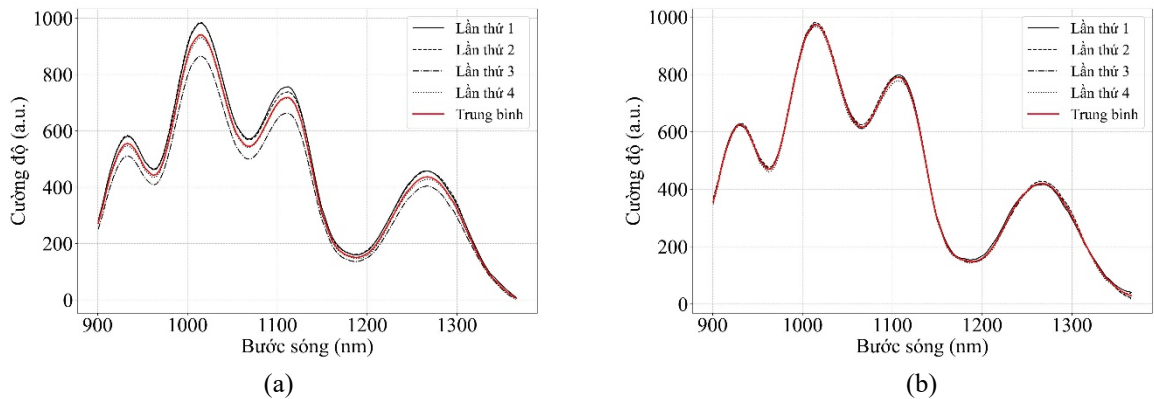
Kết quả cho thấy việc tự động hóa không chỉ cải thiện hiệu suất thu thập dữ liệu mà còn giúp chuẩn hóa quy trình đo, tạo điều kiện thuận lợi cho các thí nghiệm thu thập dữ liệu phổ với quy mô lớn và trong thực tế ứng dụng.

### 3.2. Kết quả đánh giá độ tương đồng của phổ

Để đánh giá mức độ ổn định của tín hiệu phổ trong điều kiện đo lặp lại, các phổ tương tác thu bằng phương pháp thủ công tại cùng một điểm đo, sử dụng cùng phổ kế NIRScan Nano EVM, được trình bày trong Hình 6a. Có thể quan sát thấy sự dao động đáng kể giữa các lần đo. Sự khác biệt này chủ yếu xuất phát từ các yếu tố chủ quan trong quá trình thao tác như lực tiếp xúc giữa mẫu và đầu thu, góc đặt mẫu cũng như sự ổn định của đầu thu phổ trong quá trình thu phổ.



Hình 5. Trục thời gian hoạt động của mô đun thu phổ tương tác tự động



**Hình 6. (a) Biểu đồ phổ tương tác thu bằng phương pháp thủ công, (b) Biểu đồ phổ tương tác thu bằng mô đun thu phổ tự động**

Biên độ tại đỉnh phổ trong các lần đo thủ công có giá trị trung bình  $940 \pm 56$  a.u., cho thấy mức độ dao động tương đối lớn giữa các lần đo. Điều này phản ánh sự thiếu ổn định của quá trình thu phổ khi thao tác thủ công, mặc dù hình dạng tổng thể của các phổ vẫn tương đối tương đồng.

Mặc dù các sai lệch này có thể được hiệu chỉnh bằng các thuật toán xử lý đường nền, chẳng hạn như baseline correction, bản chất của vấn đề là phổ thu được đã bị biến dạng do sự không ổn định của thao tác thu phổ thủ công. Do đó, ngay cả sau khi xử lý, các phổ này vẫn khó phản ánh chính xác các đặc trưng phổ nội tại của mẫu so với các phổ được thu thập trong điều kiện đo đã được chuẩn hóa.

Ngược lại, kết quả được thể hiện tại Hình 6b cho thấy các phổ thu được bằng mô đun tự động có độ trùng khớp cao giữa các lần đo. Biên độ tại đỉnh phổ có giá trị trung bình  $973 \pm 6$  a.u., với độ lệch chuẩn nhỏ hơn khoảng 9 lần so với trường hợp đo thủ công. Kết quả này cho thấy mức độ dao động của phổ giữa các lần đo đã giảm đáng kể khi sử dụng mô đun thu phổ tự động.

Kết quả trên cho thấy mô đun thu phổ tự động có khả năng đảm bảo điều kiện đo lặp lại nhờ cơ cấu kẹp cố định mẫu và cơ cấu tịnh tiến của đầu thu phổ được điều khiển với lực và góc tiếp xúc ổn định. Sự ổn định của phổ đo tại cùng một điểm cho thấy mô đun có khả năng đáp ứng các yêu cầu kỹ thuật của phép đo phổ tương tác trên quả ớt.

### 3.3. Mô hình dự đoán độ ngọt

Kết quả được trình bày ở Bảng 2 trình bày mô hình PLS khi áp dụng trên hai nguồn dữ liệu với các phương pháp tiền xử lý khác nhau. Nhìn chung, hiệu suất dự đoán phụ thuộc mạnh vào cả chất lượng dữ liệu phổ và phương pháp tiền xử lý được sử dụng.

Đối với dữ liệu đo thủ công, mô hình PLS xây dựng từ dữ liệu thô cho độ chính xác dự đoán trung bình ( $R_p = 0,581$ ,  $RPD = 1,23$ ). Việc áp dụng các phương pháp tiền xử lý giúp cải thiện nhẹ hiệu suất dự đoán, trong đó MSC cho kết quả tốt nhất với  $RPD = 1,30$  đối với dữ liệu đo thủ công. Trong khi đó, đối với dữ liệu đo tự động, mô hình PLS cho kết quả ổn định hơn. Mô hình xây dựng từ dữ liệu thô đạt  $R_p = 0,753$  và  $RPD = 1,46$ , cho thấy dữ liệu phổ thu bằng mô đun tự động có chất lượng cao hơn và ít nhiễu hơn so với dữ liệu đo thủ công. Điều này cho thấy việc chuẩn hóa điều kiện đo thông qua mô đun tự động giúp cải thiện tính ổn định của dữ liệu phổ đầu vào.

Kết quả của mô hình SVR được trình bày ở Bảng 3. Tương tự như mô hình PLS, hiệu suất dự đoán của SVR cũng phụ thuộc vào chất lượng dữ liệu phổ và phương pháp tiền xử lý. Đối với dữ liệu đo thủ công, mô hình SVR sử dụng dữ liệu thô đạt  $R_p = 0,704$  và  $RPD = 1,395$ . Khi áp dụng bộ lọc Savitzky–Golay, hiệu suất dự đoán được cải thiện với  $RPD$  đạt 1,628, là giá trị cao nhất trong nhóm dữ liệu đo thủ công.

Đối với dữ liệu đo tự động, mô hình SVR tiếp tục cho kết quả tốt hơn so với dữ liệu đo thủ công. Mô hình sử dụng dữ liệu thô đạt  $R_p = 0,735$  và  $RPD = 1,468$ , trong khi áp dụng tiền xử lý SG, giá trị  $RPD$  đạt 1,648, cao nhất trong toàn bộ các trường hợp khảo sát.

Nhìn chung, các mô hình xây dựng từ dữ liệu phổ đo tự động cho kết quả ổn định hơn so với dữ liệu đo thủ công. Điều này cho thấy mô đun thu phổ tự động giúp cải thiện chất lượng dữ liệu đầu vào thông qua việc chuẩn hóa điều kiện đo và giảm ảnh hưởng của các yếu tố chủ quan trong quá trình thao tác.

**Bảng 2. Kết quả tiên khả thi ứng dụng thuật toán PLS cho chế độ đo thủ công và đo tự động**

Phương pháp đo	Phương pháp tiên xử lý phổ	Bộ dữ liệu hiệu chuẩn			Bộ dữ liệu dự đoán			
		$R_c$	RMSE <sub>c</sub>	MAE <sub>c</sub>	$R_p$	RMSE <sub>p</sub>	MAE <sub>p</sub>	RPD
Đo thủ công	Không áp dụng	0,615	0,960	0,776	0,581	0,915	0,764	1,23
	SNV	0,698	0,883	0,689	0,564	0,732	0,595	1,20
	MSC	0,670	0,927	0,732	0,647	0,669	0,53	1,30
	SG	0,547	1,010	0,806	0,612	0,950	0,812	1,25
	FD	0,586	0,966	0,768	0,605	1,019	0,888	1,23
	MM	0,611	0,960	0,779	0,595	0,920	0,746	1,24
Đo tự động	Không áp dụng	0,661	0,924	0,713	0,753	0,693	0,565	1,46
	SNV	0,614	0,940	0,742	0,757	0,838	0,641	1,45
	MSC	0,647	0,938	0,728	0,651	0,791	0,631	1,30
	SG	0,643	0,923	0,733	0,679	0,883	0,733	1,34
	FD	0,663	0,923	0,718	0,433	0,886	0,704	1,10
	MM	0,504	1,026	0,816	0,698	0,927	0,760	1,32

**Bảng 3. Kết quả tiên khả thi ứng dụng thuật toán SVR cho chế độ đo thủ công và đo tự động**

Phương pháp đo	Phương pháp tiên xử lý phổ	Bộ dữ liệu hiệu chuẩn			Bộ dữ liệu dự đoán			
		$R_c$	RMSE <sub>c</sub>	MAE <sub>c</sub>	$R_p$	RMSE <sub>p</sub>	MAE <sub>p</sub>	RPD
Đo thủ công	Không áp dụng	0,751	0,805	0,599	0,704	0,796	0,597	1,395
	SNV	0,711	0,877	0,670	0,545	0,727	0,596	1,188
	MSC	0,706	0,895	0,686	0,636	0,680	0,548	1,265
	SG	0,693	0,877	0,691	0,801	0,715	0,571	1,628
	FD	0,687	0,890	0,681	0,725	0,874	0,727	1,404
	MM	0,795	0,740	0,538	0,763	0,730	0,549	1,546
Đo tự động	Không áp dụng	0,808	0,730	0,506	0,735	0,758	0,559	1,468
	SNV	0,741	0,838	0,619	0,770	0,660	0,532	1,476
	MSC	0,748	0,832	0,616	0,57	0,789	0,610	1,220
	SG	0,725	0,827	0,614	0,803	0,735	0,547	1,648
	FD	0,775	0,784	0,661	0,716	0,866	0,716	1,405
	MM	0,838	0,674	0,427	0,750	0,772	0,590	1,512

Giá trị RPD trong nghiên cứu này dao động trong khoảng 1,10 – 1,65, nằm trong mức “khả thi” đối với các nghiên cứu tiên khả thi về đánh giá chất lượng nông sản. Để đạt được mức RPD > 2,0, thường được xem là ngưỡng cho các ứng dụng dự đoán đáng tin cậy, cần xem xét các hướng cải tiến như lựa chọn bước sóng tối ưu, các phương pháp hiệu chỉnh tán xạ nâng cao hoặc các mô hình học máy phi tuyến phức tạp hơn.

#### 4. KẾT LUẬN

Mô đun thu phổ tương tác tự động cho quả quýt sau thu hoạch được thiết kế và phát triển thành công trong nghiên cứu. Kết quả thực nghiệm cho thấy mô đun có khả năng chuẩn hóa điều kiện đo, cải thiện đáng kể độ lặp lại của phổ thu được và rút ngắn thời gian thu phổ so với phương pháp thủ công. Các thử nghiệm bước đầu với mô hình dự đoán độ ngọt cho thấy dữ liệu phổ thu bằng mô đun tự động có chất lượng tốt hơn và cho kết quả dự đoán ổn định hơn so với dữ liệu đo thủ công.

Tuy nhiên, nghiên cứu hiện tại mới dừng ở mức đánh giá tính khả thi bước đầu với quy mô dữ liệu còn hạn chế và số lượng phương pháp tiền xử lý phổ cũng như mô hình học máy được khảo sát chưa nhiều. Các nghiên cứu tiếp theo có thể mở rộng thí nghiệm trên nhiều mẫu và nhiều phiên đo khác nhau để đánh giá đầy đủ độ lặp lại của hệ thống đo và khảo sát toàn diện hơn các phương pháp tiền xử lý phổ, bao gồm các tổ hợp tiền xử lý, các phương pháp

lựa chọn bước sóng, cũng như các mô hình học máy và học sâu tiên tiến nhằm tối ưu hóa hiệu suất dự đoán độ ngọt của quả quýt.

## LỜI CẢM ƠN

Nghiên cứu được thực hiện dưới sự tài trợ kinh phí của Bộ Giáo dục và Đào tạo, mã đề tài: B2024-TCT-20.

## TÀI LIỆU THAM KHẢO

- Beć, K. B., Grabska, J., & Huck, C. W. (2025). Interpretability in near-infrared (NIR) spectroscopy: Current pathways to the long-standing challenge. *TrAC Trends in Analytical Chemistry*, 189, 118254. <https://doi.org/10.1016/j.trac.2025.118254>
- Lu, R., Van Beers, R., Saeys, W., Li, C., & Cen, H. (2020). Measurement of optical properties of fruits and vegetables: A review. *Postharvest Biology and Technology*, 159, 111003. <https://doi.org/10.1016/j.postharvbio.2019.111003>
- Luo, W., Zhang, J., Liu, S., Huang, H., Zhan, B., Fan, G., & Zhang, H. (2024). Prediction of soluble solid content in Nanfeng mandarin by combining hyperspectral imaging and effective wavelength selection. *Journal of Food Composition and Analysis*, 126, 105939. <https://doi.org/10.1016/j.jfca.2023.105939>
- Magwaza, L. S., & Opara, U. L. (2015). Analytical methods for determination of sugars and sweetness of horticultural products—A review. *Scientia Horticulturae*, 184, 179–192. <https://doi.org/10.1016/j.scienta.2015.01.001>
- Manley, M. (2014). Near-infrared spectroscopy and hyperspectral imaging: non-destructive analysis of biological materials. *Chem. Soc. Rev.*, 43(24), 8200–8214. <https://doi.org/10.1039/C4CS00062E>
- Mei, M., & Li, J. (2023). An overview on optical non-destructive detection of bruises in fruit: Technology, method, application, challenge and trend. *Computers and Electronics in Agriculture*, 213, 108195. <https://doi.org/10.1016/j.compag.2023.108195>
- Mishra, P., Roger, J. M., Rutledge, D. N., & Woltering, E. (2020). SPORT pre-processing can improve near-infrared quality prediction models for fresh fruits and agro-materials. *Postharvest Biology and Technology*, 168, 111271. <https://doi.org/10.1016/j.postharvbio.2020.111271>
- Nguyen, C.-N. C.-N., Phan, Q.-T., Tran, N.-T., Fukuzawa, M., Nguyen, P.-L., & Nguyen, C.-N. C.-N. (2020). Precise sweetness grading of Mangoes (*Mangifera indica* L.) based on random forest technique with low-cost multispectral sensors. *IEEE Access*, 8, 212371–212382. <https://doi.org/10.1109/ACCESS.2020.3040062>
- Nguyễn, C. N., Nguyễn, P. L., Nguyễn, H. D., & Nguyễn, C. N. (2021). Tổng quan về đánh giá chất lượng trái cây bằng phương pháp không phá hủy. *TNU Journal of Science and Technology*, 226(11), 158–167. <https://doi.org/10.34238/tnu-jst.4673>
- Nguyễn, P. L., Dương, V. S., Trần, N. T., Nguyễn, C. N., & Nguyễn, C. N. (2024). Dự đoán độ ngọt của xoài trên cơ sở dữ liệu phổ thu thập từ cảm biến đa phổ giá thành thấp. *CTU Journal of Science*, 60(4), 11–19. <https://doi.org/10.22144/ctujos.2024.385>
- Ozaki, Y., Huck, C., Tsuchikawa, S., & Engelsen, S. B. (Eds.). (2021). *Near-Infrared Spectroscopy*. Springer Singapore. <https://doi.org/10.1007/978-981-15-8648-4>
- Pasquini, C. (2003). Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>
- Santos, C. S. P., Cruz, R., Gonçalves, D. B., Queirós, R., Bloore, M., Kovács, Z., Hoffmann, I., & Casal, S. (2021). Non-Destructive Measurement of the Internal Quality of Citrus Fruits Using a Portable NIR Device. *Journal of AOAC INTERNATIONAL*, 104(1), 61–67. <https://doi.org/10.1093/jaoacint/qsaa115>
- Schaare, P., & Fraser, D. (2000). Comparison of reflectance, interactance and transmission modes of visible-near infrared spectroscopy for measuring internal properties of kiwifruit (*Actinidia chinensis*). *Postharvest Biology and Technology*, 20(2), 175–184. [https://doi.org/10.1016/S0925-5214\(00\)00130-7](https://doi.org/10.1016/S0925-5214(00)00130-7)
- Sun, T., Xu, W., Hu, T., & Liu, M. (2014). *Application of LS-SVM and Variable Selection Methods on Predicting SSC of Nanfeng Mandarin Fruit* (pp. 249–262). [https://doi.org/10.1007/978-3-642-54344-9\\_31](https://doi.org/10.1007/978-3-642-54344-9_31)
- Xiao, Y., Li, C., Jin, C., Luo, J., Qi, H., & Zhang, C. (2025). Detection of soluble solid content in

citrus fruit using near-infrared spectroscopy with machine learning regression: An exploration of the influence of sampling positions. *Journal of Food Composition and Analysis*, 142(March), 107554.  
<https://doi.org/10.1016/j.jfca.2025.107554>

Zhang, J., & Mouazen, A. M. (2023). Fractional-order Savitzky–Golay filter for pre-treatment of on-line vis–NIR spectra to predict phosphorus in soil. *Infrared Physics & Technology*, 131, 104720.  
<https://doi.org/10.1016/j.infrared.2023.104720>