



DOI:10.22144/ctujos.2026.089

DỰ ĐOÁN ĐỘ NGỌT DƯA LƯỚI SỬ DỤNG PHỔ KẾ MINI C11708MA

Phan Thị Hồng Châu^{1,2}, Võ Thị Tú Trinh², Lê Minh Hoàng², Lương Vinh Quốc Danh² và Nguyễn Chánh Nghiệm^{1*}

¹Khoa Tự động hóa, Trường Bách Khoa, Đại học Cần Thơ, Việt Nam

²Khoa Điện tử – Viễn Thông, Trường Bách Khoa, Đại học Cần Thơ, Việt Nam

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

Thông tin chung (Article Information)

Nhận bài (Received): 10/03/2026

Sửa bài (Revised): 31/03/2026

Duyệt đăng (Accepted): 13/05/2026

Title: Prediction of melon sweetness using a mini spectrometer C11708MA

Author(s): Phan Thi Hong Chau^{1,2}, Vo Thi Tu Trinh², Le Minh Hoang², Luong Vinh Quoc Danh² and Nguyen Chanh Nghiem^{1*}

Affiliation(s): ¹Faculty of Automation Engineering, College of Engineering, Can Tho University, Viet Nam; ²Faculty of Electronics and Telecommunications, College of Engineering, Can Tho University, Viet Nam

TÓM TẮT

Khả năng ứng dụng phổ kế mini C11708MA trong dự đoán không phá hủy độ ngọt ($^{\circ}$ Brix) của dưa lưới được đánh giá trong nghiên cứu. Phổ kế được tích hợp với vi điều khiển STM32 để thu nhận và truyền dữ liệu phổ về máy tính, nơi mô hình học máy được sử dụng để dự đoán $^{\circ}$ Brix. Phổ trong dải 640–1050 nm được thu thập từ 50 quả dưa lưới theo chiến lược lấy mẫu đa điểm. Các kỹ thuật tiền xử lý như lọc Savitzky–Golay, biến chuẩn hóa, hiệu chỉnh tán xạ nhân và các tổ hợp của chúng được áp dụng trước khi xây dựng mô hình hồi quy bình phương tối thiểu từng phần (PLS) và vector hỗ trợ (SVR). Kết quả cho thấy SVR thường cho hiệu suất tốt hơn PLS, với RPD khoảng 1,7–1,8, cho thấy khả năng dự đoán chấp nhận được cho các ứng dụng sàng lọc nhanh. Kết quả này tương đương một số nghiên cứu trước đây sử dụng phổ kế mini, cho thấy tiềm năng phát triển thiết bị quang phổ cầm tay chi phí thấp cho đánh giá nhanh chất lượng nông sản.

Từ khóa: Dưa lưới, độ ngọt, phổ kế mini, tiền xử lý

ABSTRACT

This study evaluates the feasibility of using a C11708MA mini-spectrometer for the non-destructive prediction of melon sweetness ($^{\circ}$ Brix). The spectrometer was integrated with an STM32 microcontroller to acquire and transmit spectral data to a computer for machine-learning-based prediction. Spectra in the range of 640–1050 nm were collected from 50 melons using a multi-point sampling strategy. Preprocessing techniques, including Savitzky–Golay filtering, standard normal variate, multiplicative scatter correction, and their combinations, were applied before developing partial least squares regression (PLS) and support vector regression (SVR) models. SVR generally outperformed PLS, achieving RPD values of about 1.7–1.8, indicating acceptable prediction capability for rapid screening. These results are comparable to previous studies using mini-spectrometers and highlight the potential for developing low-cost handheld spectroscopic devices for rapid fruit quality assessment.

Keywords: Melon, mini spectrometer, preprocessing, sweetness

1. GIỚI THIỆU

Dưa lưới (*Cucumis melo L.*) là loại trái cây có giá trị kinh tế cao không chỉ tại Việt Nam mà còn trên thế giới nhờ hương vị đặc trưng, giá trị dinh dưỡng và nhu cầu tiêu thụ ngày càng tăng (Gómez-García et al., 2020). Trong chuỗi sản xuất và xuất khẩu, độ ngọt ($^{\circ}\text{Brix}$) là một trong những chỉ tiêu quan trọng để đánh giá chất lượng và mức độ chín của quả (Erčan et al., 2024). Hiện nay, $^{\circ}\text{Brix}$ thường được xác định bằng khúc xạ kế, tuy nhiên phương pháp này mang tính phá hủy mẫu, tốn thời gian và khó áp dụng cho các hệ thống kiểm soát chất lượng quy mô lớn hoặc theo thời gian thực.

Trong những năm gần đây, các phương pháp đánh giá không phá hủy dựa trên công nghệ quang phổ vùng khả kiến – cận hồng ngoại (Vis–NIR) đã được nghiên cứu rộng rãi để dự đoán các chỉ tiêu chất lượng của nhiều loại trái cây như bưởi (C. Wang et al., 2025), xoài (Sohaib Ali Shah et al., 2021), táo (Pourdarbani et al., 2021), kiwi (Cevoli et al., 2024), chuối (Pahlawan et al., 2021), dưa hấu (Lazim et al., 2022) và dưa lưới (Zeb et al., 2021). Công nghệ này cho phép khai thác mối liên hệ giữa tín hiệu quang học và thành phần hóa học của quả để xây dựng các mô hình dự đoán nhanh các chỉ tiêu như độ ngọt, độ chín hoặc hàm lượng chất khô. Tuy nhiên, trong phần lớn các nghiên cứu, các hệ thống quang phổ để bàn có kích thước lớn và chi phí cao, phù hợp chủ yếu trong điều kiện phòng thí nghiệm và hạn chế khả năng ứng dụng thực địa vẫn được sử dụng (Wu et al., 2023).

Bên cạnh đó, nhiều nghiên cứu mới chỉ dừng lại ở việc chứng minh tiềm năng chung của công nghệ quang phổ Vis–NIR, trong khi việc đánh giá hiệu quả thực tế của các phổ kế mini, vốn có khả năng tích hợp và triển khai trực tiếp tại hiện trường, vẫn còn hạn chế. Tuy nhiên, một số nghiên cứu gần đây cho thấy các phổ kế mini chi phí thấp vẫn có thể đạt độ chính xác cao khi được tích hợp vào thiết bị cầm tay và kết hợp với các thuật toán học máy phù hợp. Chẳng hạn, Liu et al. (2024) đã phát triển một thiết bị di động sử dụng phổ kế C12880MA kết hợp với vi điều khiển STM32 để phát hiện lỗi mốc ở táo với độ chính xác trên 97%. Tương tự, Hotra et al. (2022) đã phân tích khả năng ứng dụng của các phổ kế C12880MA và C11708MA trong các hệ thống quang phổ nhỏ gọn. Nhờ công nghệ vi cơ điện tử (MEMS) và bán dẫn oxit kim loại bổ sung (CMOS), các phổ kế này có khả năng đo phổ với độ phân giải cao và độ lặp lại tốt, phù hợp cho các ứng dụng phân tích trực tiếp tại hiện trường. Kết quả nghiên cứu này cho thấy việc tích hợp các phổ kế mini vào thiết bị

cầm tay là khả thi khi phần cứng, sơ đồ mạch và phần mềm được thiết kế và tối ưu phù hợp. Những kết quả này cho thấy tiềm năng thay thế các hệ thống quang phổ truyền thống bằng các thiết bị mini chi phí thấp trong giám sát và đánh giá chất lượng trái cây, đồng thời mở ra cơ hội ứng dụng trong các mô hình nông nghiệp thông minh yêu cầu kiểm soát chất lượng nhanh và hiệu quả. Tuy nhiên, các nghiên cứu này chủ yếu tập trung vào một số loại trái cây như táo hoặc xoài, trong khi việc đánh giá khả năng dự đoán độ ngọt của dưa lưới bằng phổ kế mini vẫn còn hạn chế (Li et al., 2022; Liu et al., 2024). Điều này đặt ra một khoảng trống quan trọng, bởi dưa lưới có cấu trúc mô và đặc tính quang học khác biệt so với các loại trái cây khác, do đó hiệu quả dự đoán từ các thiết bị mini cần được kiểm chứng riêng. Ngoài ra, nhiều nghiên cứu chỉ sử dụng dữ liệu phổ thu thập tại một vài vị trí trên quả hoặc dữ liệu trung bình đại diện, nên chưa phản ánh đầy đủ tính không đồng nhất của quả; thậm chí, một số nghiên cứu còn sử dụng các kỹ thuật tăng cường dữ liệu dựa trên thuật toán để mở rộng tập mẫu (Khumaidi & Raafi'udin, 2022; Tan et al., 2024).

Từ những hạn chế nêu trên, nghiên cứu này tập trung đánh giá khả năng ứng dụng của phổ kế mini C11708MA trong dự đoán không phá hủy độ ngọt của dưa lưới. Nghiên cứu áp dụng chiến lược lấy mẫu đa điểm nhằm phản ánh tốt hơn sự biến thiên của dữ liệu phổ trên bề mặt quả, qua đó tránh lãng phí mẫu và nâng cao độ tin cậy của dữ liệu đầu vào. Đồng thời, các mô hình học máy được xây dựng và so sánh để nâng cao độ chính xác dự đoán từ dữ liệu quang phổ thu được. Kết quả nghiên cứu cũng được đối chiếu với các công trình trước sử dụng phổ kế chi phí thấp nhằm làm rõ tiềm năng ứng dụng của thiết bị trong các hệ thống đánh giá chất lượng trái cây không phá hủy.

Cụ thể, nghiên cứu hướng tới các mục tiêu sau: (1) Thu thập dữ liệu quang phổ Vis–NIR của dưa lưới bằng phổ kế mini C11708MA theo chiến lược lấy mẫu đa điểm; (2) Khảo sát và so sánh các kỹ thuật tiền xử lý phổ nhằm cải thiện chất lượng tín hiệu; (3) Xây dựng và đánh giá các mô hình học máy để dự đoán giá trị $^{\circ}\text{Brix}$ từ dữ liệu quang phổ; (4) Đánh giá tiềm năng ứng dụng của phổ kế mini trong phát triển thiết bị quang phổ cầm tay phục vụ giám sát chất lượng trái cây theo phương pháp không phá hủy.

2. PHƯƠNG PHÁP THỰC HIỆN

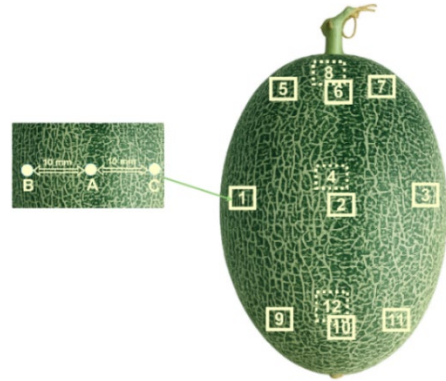
2.1. Chuẩn bị mẫu dưa lưới

Tổng cộng 50 quả dưa lưới giống Thanh Loan được thu hoạch trực tiếp tại RK Farm (Thành phố Cần Thơ) trong năm đợt khác nhau nhằm đảm bảo sự đa dạng về mức độ chín sinh lý của mẫu. Cụ thể, các đợt thu hoạch được thực hiện vào các ngày sau khi trồng 67, 72, 73, 76 và 77 ngày. Sau thu hoạch, mẫu được vận chuyển đến phòng thí nghiệm của Đại học Cần Thơ để tiến hành làm sạch bề mặt, mã hóa nhận dạng và xác định vị trí đo, đảm bảo tính nhất quán trong quá trình thu nhận dữ liệu quang phổ. Các quả sau đó được bảo quản ở nhiệt độ phòng trong khoảng 24 giờ trước khi đo nhằm ổn định trạng thái nhiệt và hạn chế sai số do biến thiên nhiệt độ đến kết quả phổ (Cozzolino et al., 2007).

Khối lượng quả dao động từ 1,8 đến 3,2 kg, với độ dày vỏ khoảng 5–20 mm. Để giảm thiểu ảnh hưởng của tính không đồng nhất không gian của hàm lượng chất rắn hòa tan (SSC), dữ liệu quang phổ được thu tại 12 vị trí trên mỗi quả, phân bố ở ba vùng đặc trưng gồm cuống, xích đạo và đài; mỗi vùng ghi nhận bốn phép đo độc lập (xem Hình 1). Tại mỗi vị trí đo, ba điểm kề nhau (B–A–C) được bố trí thẳng hàng theo phương ngang với khoảng cách xấp xỉ 10 mm và phổ thu được được lấy giá trị trung bình nhằm hạn chế sai khác cục bộ.

Song song với quá trình thu phổ, giá trị độ ngọt (SSC, °Brix) tại từng vị trí được xác định bằng khúc xạ kế cầm tay và sử dụng làm giá trị tham chiếu cho mô hình dự đoán. Để đảm bảo sự tương thích giữa phép đo phổ không phá hủy và phân tích tham chiếu, tại mỗi vị trí đo, một lớp vỏ dày khoảng 2 mm được loại bỏ trước khi lấy mẫu thịt quả. Sau đó, một khối thịt dưa kích thước khoảng $15 \times 15 \times 15 \text{ mm}^3$ được cắt từ lớp mô ngay bên dưới vùng đo phổ. Độ sâu lấy mẫu được lựa chọn nhằm xấp xỉ độ thâm nhập hiệu dụng của tín hiệu Vis–NIR và đảm bảo đủ dịch ép để xác định °Brix, đồng thời giảm ảnh hưởng của sự không đồng nhất thành phần trong mô quả.

Tổng cộng 600 phổ cùng giá trị °Brix tương ứng được thu thập. Dữ liệu sau đó được chia ngẫu nhiên thành tập hiệu chuẩn và tập dự đoán theo tỷ lệ 80:20 để phục vụ quá trình xây dựng và đánh giá mô hình.

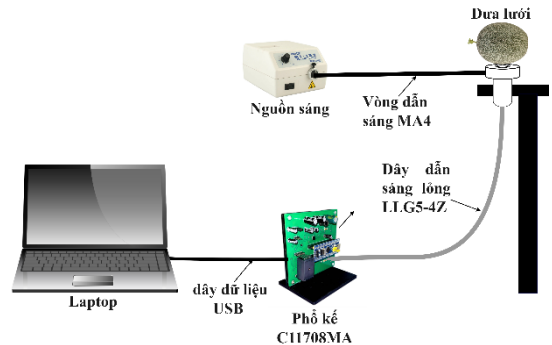


Hình 1. Vị trí đo phổ dưa lưới

2.2. Thiết kế hệ thống đo phổ Vis-NIR

Trong nghiên cứu này, một hệ thống đo quang phổ Vis–NIR được xây dựng nhằm khảo sát đặc tính quang học của dưa lưới theo phương pháp đo phổ tương tác. Hệ thống bao gồm nguồn sáng halogen công suất 150 W, vòng dẫn sáng MA4 (20 mm Mini Annular Ring Light), dây dẫn ánh sáng lỏng LIG5-4Z và phổ kế C11708MA (Hamamatsu Photonics K.K., Nhật Bản).

Nguồn sáng halogen được ghép với vòng chiếu sáng MA4 để tạo trường chiếu sáng đồng đều trên bề mặt mẫu. Tín hiệu phản xạ từ mẫu được thu nhận thông qua dây dẫn ánh sáng lỏng và truyền đến phổ kế C11708MA để ghi nhận phổ phản xạ (Hình 2).



Hình 2. Sơ đồ bố trí hệ thống đo phổ

2.2.1. Cấu trúc và nguyên lý hoạt động của phổ kế C11708MA

Phổ kế C11708MA có kích thước nhỏ gọn ($27,6 \times 16,8 \times 13 \text{ mm}^3$) và tích hợp hoàn chỉnh hệ quang học bên trong, bao gồm khe vào và cách tử nhiễu xạ phân xạ dạng lõm tạo trường ảnh phẳng tại vị trí ảnh của khe. Bức xạ đi qua khe được tán sắc bởi cách tử và ghi nhận bởi mảng cảm biến CMOS gồm 256 phần tử quang điện (Hamamatsu Photonics K.K., 2022).

Thiết bị hoạt động trong dải bước sóng 640–1050 nm với độ phân giải quang học khoảng 15 nm. Màng photodiode sử dụng cơ chế đọc tín hiệu rolling shutter, trong đó các phần tử cảm biến được kích hoạt tuần tự dưới sự điều khiển của xung nhịp CLK (tối đa 800 kHz) để tích phân dòng quang sinh ra khi bức xạ chiếu tới.

Sau quá trình tích phân, tín hiệu video tương tự được đưa tới bộ chuyển đổi tương tự–số (ADC) để chuyển thành dữ liệu số. Trong bước xử lý tiếp theo, tín hiệu này được chuẩn hóa theo thời gian tích phân τ bằng công thức $V = D_s/\tau$, trong đó D_s là giá trị tín hiệu số thu được từ ADC. Việc chuẩn hóa này giúp loại bỏ ảnh hưởng của thời gian tích phân, nhờ đó duy trì mối quan hệ tuyến tính giữa tín hiệu đo và cường độ bức xạ tới cảm biến (Hotra et al., 2022).

2.2.2. Thiết kế mạch kết nối và điều khiển

Trong hệ thống đo phổ được phát triển, phổ kế C11708MA được kết nối với vi điều khiển STM32F103C8T6 (bo mạch Blue Pill), đóng vai trò điều khiển và thu nhận dữ liệu phổ (xem Hình 3). Vi điều khiển này dựa trên kiến trúc ARM Cortex M3, hoạt động ở tần số 72 MHz và tích hợp bộ chuyển đổi ADC 12 bit cùng các bộ định thời, đáp ứng yêu cầu điều khiển và thu nhận dữ liệu thời gian thực của hệ thống.

Trong quá trình hoạt động, vi điều khiển tạo các tín hiệu điều khiển CLK và ST theo sơ đồ thời gian khuyến nghị để vận hành phổ kế. Do mức logic của vi điều khiển là từ 0 đến 3,3 V trong khi phổ kế sử dụng mức TTL từ 0 đến 5 V, một mạch chuyển mức logic được sử dụng để đảm bảo tương thích điện áp giữa hai thiết bị. Tín hiệu video analog từ phổ kế được đưa tới đầu vào ADC thông qua mạch khuếch đại đệm sử dụng op amp MCP6002 nhằm tăng trở kháng đầu vào và ổn định tín hiệu đo.

Hệ thống được cấp nguồn từ cổng USB của máy tính và sử dụng bộ lọc LC dạng π để giảm nhiễu trên đường nguồn. Tần số CLK được thiết lập khoảng 800 kHz và tần số lấy mẫu ADC khoảng 200 kHz để đảm bảo đồng bộ giữa quá trình tích phân và đọc dữ liệu. Cảm biến có thời gian phơi sáng tối thiểu khoảng 1,29 ms và hỗ trợ chân Gain cho phép tăng độ nhạy đo lên khoảng 3,5 lần khi cần thiết (Firago et al., 2022).



Hình 3. Thiết kế mạch thực tế cho C11708MA

2.2.3. Phần mềm điều khiển và xử lý dữ liệu

Hệ thống đo được thiết kế để kết nối trực tiếp với máy tính thông qua giao tiếp USB nhằm thuận tiện cho quá trình thu thập, lưu trữ và xử lý dữ liệu phổ. Phần mềm điều khiển được xây dựng theo kiến trúc hai lớp gồm firmware trên vi điều khiển và phần mềm xử lý trên máy tính.

Firmware trên vi điều khiển STM32F103C8T6 được lập trình bằng STM32CubeMX để tạo các tín hiệu điều khiển (ST, CLK, Gain) cho cảm biến C11708MA, thiết lập thời gian tích phân, thu nhận tín hiệu video và chuyển đổi sang dữ liệu số thông qua ADC. Dữ liệu phổ sau đó được tính trung bình và truyền tới máy tính qua cổng USB.

Phần mềm trên máy tính được phát triển bằng môi trường Processing, có chức năng hiển thị phổ theo thời gian thực và lưu trữ dữ liệu dưới dạng tệp .csv. Mỗi tệp chứa các thông tin như số thứ tự mẫu, vị trí đo, bước sóng và giá trị phổ tương ứng. Cách tiếp cận này giúp hệ thống vận hành ổn định và hỗ trợ hiệu quả cho quá trình xử lý dữ liệu và xây dựng mô hình dự đoán ở các bước nghiên cứu tiếp theo.

2.3. Tiền xử lý dữ liệu

Trước khi xây dựng mô hình dự đoán, dữ liệu phổ được tiền xử lý nhằm loại bỏ nhiễu ngẫu nhiên và giảm thiểu các sai lệch phát sinh trong quá trình thu nhận, chẳng hạn như sự thay đổi cường độ chiếu sáng, góc đo hoặc đặc tính bề mặt của mẫu. Những yếu tố này có thể gây ra hiện tượng tán xạ ánh sáng, dịch chuyển đường nền hoặc biến dạng tín hiệu phổ, từ đó làm giảm độ chính xác của các mô hình dự đoán. Vì vậy, việc áp dụng các phương pháp tiền xử lý phù hợp đóng vai trò quan trọng trong việc cải thiện chất lượng dữ liệu đầu vào trước khi xây dựng mô hình dự đoán.

Trong nghiên cứu này, một số phương pháp tiền xử lý phổ biến đã được áp dụng và so sánh, bao gồm biến chuẩn hóa (Standard Normal Variate – SNV), hiệu chỉnh phân tán nhân (Multiplicative Scatter Correction – MSC) và lọc Savitzky–Golay (SG). Phương pháp SNV được áp dụng cho từng phổ riêng lẻ theo công thức (Dhanoa et al., 2023):

$$y_i^{\text{SNV}} = \frac{y_i - \bar{y}_i}{\sigma_i} \quad (1)$$

Trong đó, y_i là giá trị phản xạ tại bước sóng i , \bar{y}_i và σ_i lần lượt là giá trị trung bình và độ lệch chuẩn của toàn bộ phổ mẫu cần chuẩn hóa. Phép chuẩn hóa này giúp khử ảnh hưởng của sự tán xạ và sai khác về hình học giữa các mẫu, đồng thời tăng tính đồng nhất của tín hiệu.

Phương pháp MSC được sử dụng để hiệu chỉnh phổ đo dựa trên phổ tham chiếu trung bình thông qua hồi quy tuyến tính (Windig et al., 2008):

$$y_i^{\text{MSC}} = \frac{y_i - \hat{a}_i}{\hat{b}_i} \quad (2)$$

Trong đó, y_i là giá trị phản xạ tại bước sóng i của mẫu cần hiệu chỉnh; \hat{a}_i và \hat{b}_i lần lượt là hệ số chặn và hệ số góc thu được từ hồi quy tuyến tính giữa y_i và y_{ref} (phổ tham chiếu, thường được xem là phổ trung bình của toàn bộ tập dữ liệu hiệu chuẩn với giá định là chứa thông tin hóa học tương tự như y_i). Phương pháp này giúp loại bỏ ảnh hưởng của tán xạ ánh sáng và sự khác biệt về hình học giữa các mẫu, đồng thời chuẩn hóa các phổ về cùng một thang đo, tạo điều kiện thuận lợi cho việc xây dựng mô hình định lượng.

Ngoài các phương pháp hiệu chỉnh tán xạ như SNV và MSC, bộ lọc SG được sử dụng để làm trơn dữ liệu phổ và giảm nhiễu tần số cao trong khi vẫn bảo toàn dạng đặc trưng của tín hiệu. Đồng thời, đạo hàm bậc nhất (FD) của phổ sau khi lọc SG được tính toán nhằm làm nổi bật các đặc trưng hấp thụ, giảm ảnh hưởng của dịch chuyển đường nền và cải thiện khả năng phân tách giữa các dải phổ chồng lấp.

Bên cạnh việc áp dụng các phương pháp riêng lẻ, nghiên cứu cũng khảo sát một số tổ hợp tiền xử lý gồm SNV + SG, SNV + SG + FD, MSC + SG và MSC + SG + FD. Các phương pháp kết hợp này giúp đồng thời hiệu chỉnh tán xạ ánh sáng, giảm nhiễu và bù trừ biến động đường nền, từ đó nâng cao chất lượng dữ liệu phổ.

Các kỹ thuật tiền xử lý sau đó được so sánh nhằm lựa chọn phương án phù hợp nhất cho việc xây dựng các mô hình hồi quy dự đoán độ ngọt ($^{\circ}\text{Brix}$) của dưa lưới. Nhiều nghiên cứu gần đây cũng cho thấy việc lựa chọn phương pháp tiền xử lý thích hợp có ảnh hưởng đáng kể đến hiệu suất của các mô hình phân tích quang phổ (Kusumiyati et al., 2021; Yang et al., 2022; Han et al., 2023).

2.4. Xây dựng mô hình hồi quy

Trong phân tích phổ Vis–NIR, việc xây dựng mô hình dự đoán chính xác đóng vai trò quan trọng trong khai thác thông tin từ dữ liệu phổ. Trong nghiên cứu này, hai phương pháp hồi quy phổ biến được sử dụng để dự đoán độ ngọt ($^{\circ}\text{Brix}$) của dưa lưới, gồm hồi quy bình phương tối thiểu từng phần (Partial Least Squares – PLS) và hồi quy vector hỗ trợ (Support Vector Regression – SVR). Hai mô hình này đại diện cho hai hướng tiếp cận khác nhau, bao gồm mô hình hồi quy tuyến tính truyền thống và phương pháp học máy phi tuyến. Việc áp dụng và so sánh hai phương pháp giúp đánh giá khả năng mô hình hóa dữ liệu phổ, đồng thời xác định phương pháp phù hợp cho bài toán dự đoán $^{\circ}\text{Brix}$ theo hướng không phá hủy.

PLS là phương pháp hồi quy tuyến tính được sử dụng rộng rãi trong phân tích dữ liệu quang phổ do khả năng xử lý hiệu quả các tập dữ liệu có số lượng biến lớn và tồn tại đa cộng tuyến. Phương pháp này xây dựng các biến tiềm ẩn (latent variables) từ ma trận dữ liệu phổ nhằm tối đa hóa mối liên hệ với biến mục tiêu ($^{\circ}\text{Brix}$). Nhờ quá trình trích xuất các thành phần thông tin quan trọng, PLS giúp giảm chiều dữ liệu, hạn chế nhiễu và cải thiện độ ổn định của mô hình (Mekonnen et al., 2020).

Bên cạnh đó, SVR được áp dụng để khai thác các quan hệ phi tuyến tiềm ẩn giữa dữ liệu phổ và giá trị $^{\circ}\text{Brix}$. Thuật toán này xây dựng hàm hồi quy trong đó sai số dự đoán được kiểm soát trong một khoảng dung sai xác định, đồng thời tối ưu hóa khả năng tổng quát hóa đối với dữ liệu chưa quan sát. Nhờ cơ chế này, SVR có khả năng mô tả tốt các mối quan hệ phức tạp trong dữ liệu phổ và thường cho hiệu suất dự đoán cao trong các bài toán nông nghiệp dựa trên quang phổ (Yang et al., 2025).

Trong nghiên cứu này, các mô hình được xây dựng trên cùng tập dữ liệu và siêu tham số được tối ưu bằng grid-search kết hợp với 10-fold cross-validation. Phương pháp này giúp giảm ảnh hưởng của việc phân chia dữ liệu ngẫu nhiên và cung cấp đánh giá đáng tin cậy hơn về khả năng dự đoán của mô hình đối với dữ liệu mới. Hiệu suất của các mô

hình được so sánh thông qua các chỉ số đánh giá thống nhất nhằm lựa chọn mô hình tối ưu cho dự đoán °Brix không phá hủy của dưa lưới.

2.5. Đánh giá mô hình hồi quy

Hiệu quả dự đoán của các mô hình hồi quy được đánh giá thông qua một số chỉ số thống kê thường dùng, bao gồm hệ số tương quan (R), sai số căn bậc hai trung bình (RMSE), sai số tuyệt đối trung bình (MAE) và tỷ lệ hiệu suất độ lệch (RPD) (Miloš & Bensa, 2017; Zhang et al., 2024).

$$R = \frac{\sum_{i=1}^n (y_i - y_m)(\hat{y}_i - \hat{y}_m)}{\sqrt{\sum_{i=1}^n (y_i - y_m)^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \hat{y}_m)^2}} \quad (3)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4)$$

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (5)$$

$$RPD = \frac{SD}{RMSE_p} \quad (6)$$

Trong đó, y_i và \hat{y}_i lần lượt là giá trị đo thực tế, giá trị dự đoán của mẫu thứ i ; y_m và \hat{y}_m là trị trung bình của các giá trị đo thực tế và giá trị dự đoán trong tập hiệu chuẩn hoặc tập dự đoán; n là số lượng mẫu trong tập dữ liệu và SD là độ lệch chuẩn của các giá trị tham chiếu, phản ánh mức độ phân tán của các giá trị đo thực tế.

Hệ số R phản ánh mức độ tương quan tuyến tính giữa giá trị dự đoán và giá trị đo thực tế; giá trị R càng tiến gần đến 1 thì mức độ phù hợp của mô hình càng cao. Hai chỉ số RMSE và MAE được sử dụng để đánh giá sai số giữa giá trị dự đoán và giá trị đo thực tế. Trong đó, RMSE nhạy cảm hơn với các sai số lớn, còn MAE thể hiện độ lệch trung bình tuyệt đối của toàn bộ tập dữ liệu. Giá trị của hai chỉ số này càng nhỏ thì độ chính xác dự đoán của mô hình càng tốt.

Chỉ số RPD được sử dụng để đánh giá khả năng dự đoán của mô hình. Cụ thể, $RPD < 1,4$ cho thấy mô hình có hiệu suất dự đoán kém và chỉ có thể phân biệt các giá trị cao và thấp của biến nghiên cứu. Mô hình đạt mức dự đoán chấp nhận được khi $1,4 \leq RPD < 1,8$. Khi $1,8 \leq RPD < 2,0$, mô hình có khả năng dự đoán tốt và có thể sử dụng cho dự đoán định lượng. Trong khi đó, $2,0 \leq RPD < 2,5$ phản ánh hiệu suất dự đoán rất tốt, và $RPD \geq 2,5$ cho thấy mô hình có

độ chính xác cao và phù hợp cho các ứng dụng thực tiễn (Zheng et al., 2019).

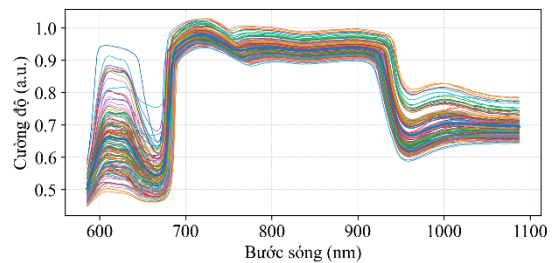
Việc kết hợp các chỉ số đánh giá này giúp đưa ra cái nhìn toàn diện về hiệu suất mô hình, bao gồm mức độ tương quan, khả năng mô tả dữ liệu và độ chính xác trong dự đoán độ ngọt (°Brix) của dưa lưới.

3. KẾT QUẢ VÀ THẢO LUẬN

3.1. Đặc trưng quang phổ Vis-NIR

Kết quả được trình bày ở Hình 4 cho thấy phổ phản xạ của các mẫu dưa lưới được đo trong dải bước sóng từ 640 đến 1050 nm. Có thể thấy rằng các đường phổ thu được có hình dạng tương đối tương đồng giữa các mẫu, cho thấy phép đo đạt độ lặp lại cao và dữ liệu có tính ổn định tốt. Phổ quan sát thể hiện ba vùng đặc trưng rõ rệt, phản ánh sự khác biệt về cơ chế tương tác giữa ánh sáng và vật chất trong từng khoảng bước sóng.

Thứ nhất, vùng 600–700 nm xuất hiện sự biến động mạnh về cường độ, chủ yếu liên quan tới sự hấp thụ của các sắc tố quang hợp như chlorophyll và carotenoid (Pratiwi et al., 2023). Thứ hai, vùng 700–900 nm có dạng tương đối phẳng với ít dao động, được xem như một “cửa sổ quang học” qua đó ánh sáng có thể xuyên sâu hơn vào mô quả, phản ánh đặc điểm cấu trúc bên trong quả (Products & Inspection, 2019). Thứ ba, vùng 900–1100 nm ghi nhận sự suy giảm đáng kể cường độ tín hiệu, thường liên quan tới các dải hấp thụ đặc trưng của nước và các nhóm chức hydroxyl (–OH), đồng thời có liên quan mật thiết tới sự hiện diện của carbohydrate, đường hòa tan và các hợp chất hòa tan khác – vốn là những yếu tố quyết định chất lượng cảm quan và giá trị thương phẩm của quả (G. Wang et al., 2025).



Hình 4. Phổ của 600 mẫu dưa thu được từ phổ kế C11708MA

Sự phân chia rõ ràng này phản ánh cả ảnh hưởng của sắc tố bề mặt, cấu trúc nội tại và thành phần hóa học như nước và đường; nhờ đó, việc khai thác thông tin phổ trong các vùng này mang lại tiềm năng mạnh mẽ cho việc xây dựng mô hình dự đoán các

chỉ tiêu nội tại như độ ngọt ($^{\circ}$ Brix) bằng phương pháp phổ không phá hủy.

3.2. So sánh hiệu quả các mô hình hồi quy

Bảng 1 trình bày kết quả so sánh hiệu suất dự đoán của các mô hình hồi quy được xây dựng từ dữ liệu phổ thu thập bằng phổ kế C11708MA trong dải bước sóng 640–1050 nm. Trong nghiên cứu này, mô hình PLS, SVR được áp dụng kết hợp với nhiều chiến lược tiền xử lý phổ khác nhau nhằm đánh giá ảnh hưởng của từng phương pháp xử lý dữ liệu đến độ chính xác của mô hình. Cụ thể, các phương pháp

tiền xử lý được xem xét bao gồm SNV, MSC và SG cùng với các chiến lược kết hợp SNV + SG, SNV + SG + FD, MSC + SG và MSC + SG + FD.

Ngoài ra, dữ liệu phổ gốc không áp dụng tiền xử lý cũng được sử dụng làm cơ sở tham chiếu. Việc so sánh hiệu suất dự đoán giữa các mô hình PLS và SVR dưới các điều kiện tiền xử lý khác nhau cho phép đánh giá mức độ cải thiện của từng kỹ thuật xử lý phổ, đồng thời xác định tổ hợp tiền xử lý tối ưu nhằm nâng cao độ chính xác và độ ổn định của mô hình dự đoán độ ngọt ($^{\circ}$ Brix) của dưa lưới.

Bảng 1. Hiệu suất các mô hình được xây dựng dựa trên dữ liệu được tiền xử lý khác nhau

Mô hình	Tiền xử lý dữ liệu phổ	Tập dữ liệu hiệu chỉnh			Tập dữ liệu dự đoán			
		R_c	RMSE _c	MAE _c	R_p	RMSE _p	MAE _p	RPD
PLS	SNV	0,80	1,62	1,25	0,78	1,71	1,39	1,60
	MSC	0,80	1,64	1,26	0,78	1,71	1,39	1,60
	SG	0,81	1,61	1,23	0,81	1,61	1,29	1,69
	SG + FD	0,75	1,80	1,38	0,79	1,69	1,36	1,61
	SNV + SG	0,77	1,73	1,38	0,78	1,71	1,42	1,60
	SNV + SG + FD	0,73	1,87	1,44	0,77	1,75	1,41	1,55
	MSC + SG	0,77	1,72	1,36	0,78	1,72	1,43	1,58
	MSC + SG + FD	0,73	1,86	1,46	0,77	1,75	1,44	1,56
	Không tiền xử lý	0,82	1,54	1,16	0,81	1,61	1,26	1,69
	SVR	SNV	0,82	1,57	1,10	0,80	1,62	1,24
MSC		0,83	1,52	1,13	0,80	1,63	1,28	1,67
SG		0,84	1,49	1,08	0,83	1,52	1,22	1,79
SG + FD		0,82	1,56	1,10	0,82	1,57	1,23	1,73
SNV + SG		0,80	1,62	1,17	0,81	1,59	1,25	1,71
SNV + SG + FD		0,82	1,58	1,12	0,82	1,58	1,22	1,73
MSC + SG		0,81	1,59	1,19	0,81	1,61	1,27	1,70
MSC + SG + FD		0,82	1,58	1,11	0,81	1,59	1,23	1,71
Không tiền xử lý		0,85	1,44	1,01	0,83	1,54	1,22	1,77

Nhìn chung, các mô hình đạt mức tương quan dự đoán khá tốt với giá trị R_p dao động từ 0,77 đến 0,83, trong khi giá trị RMSE_p nằm trong khoảng 1,52–1,75 $^{\circ}$ Brix, cho thấy dữ liệu phổ trong vùng bước sóng này có khả năng phản ánh thông tin liên quan đến độ ngọt của dưa lưới.

Đối với mô hình PLS, kết quả cho thấy phương pháp SG mang lại hiệu suất dự đoán tốt nhất với $R_p = 0,81$, RMSE_p = 1,61 $^{\circ}$ Brix và RPD_p = 1,69, tương đương với dữ liệu không tiền xử lý. Trong khi đó, việc kết hợp thêm đạo hàm bậc nhất (SG + FD) hoặc các phương pháp hiệu chỉnh tán xạ như SNV và MSC không cải thiện đáng kể hiệu suất mô hình, thậm chí trong một số trường hợp còn làm giảm độ chính xác dự đoán. Điều này cho thấy dữ liệu phổ ban đầu đã có chất lượng tương đối ổn định và ít bị ảnh hưởng bởi nhiễu tán xạ hoặc trôi đường nền.

Đối với mô hình SVR, kết quả dự đoán nhìn chung tốt hơn so với PLS trong hầu hết các trường hợp tiền xử lý. Cụ thể, mô hình SVR với tiền xử lý SG đạt hiệu suất tốt nhất với $R_p = 0,83$, RMSE_p = 1,52 $^{\circ}$ Brix và RPD_p = 1,79. Ngoài ra, các chiến lược tiền xử lý như SG, SG + FD, và SNV + SG + FD cũng cho kết quả dự đoán tương đối cao với R_p khoảng 0,82–0,83 và RMSE_p khoảng 1,52–1,58 $^{\circ}$ Brix.

So sánh giữa hai mô hình cho thấy SVR thể hiện khả năng dự đoán tốt hơn PLS, với giá trị R_p cao hơn và sai số dự đoán thấp hơn trong hầu hết các trường hợp. Điều này có thể được giải thích bởi khả năng của SVR trong việc mô hình hóa các mối quan hệ phi tuyến giữa dữ liệu phổ và độ ngọt của dưa lưới, trong khi PLS chủ yếu dựa trên giả định tuyến tính.

Nhìn chung, các giá trị RPD dao động từ 1,55 đến 1,79, cho thấy các mô hình có khả năng dự đoán ở mức chấp nhận được đối với phân tích định lượng

sơ bộ. Kết quả này cũng chỉ ra rằng việc áp dụng các phương pháp tiền xử lý phức tạp không phải lúc nào cũng cải thiện hiệu suất mô hình và trong một số trường hợp, dữ liệu phổ gốc hoặc chỉ áp dụng lọc SG đã đủ để xây dựng mô hình dự đoán hiệu quả.

3.3. Phân tích phổ sau tiền xử lý Savitzky–Golay và cấu trúc dữ liệu phổ

Kết quả so sánh hiệu suất mô hình ở mục 3.2 cho thấy tiền xử lý SG mang lại hiệu quả tốt nhất trong các phương án khảo sát, do đó phương pháp này được lựa chọn để phân tích sâu hơn dữ liệu phổ. Phương pháp SG giúp giảm nhiễu ngẫu nhiên và làm mượt đường phổ mà vẫn bảo toàn hình dạng cũng như vị trí của các đỉnh hấp thụ đặc trưng, qua đó làm nổi bật các đặc trưng phổ liên quan đến thành phần hóa học của mẫu.

Sau khi áp dụng bộ lọc SG, các đường phổ trở nên mượt hơn và ít dao động ngẫu nhiên so với phổ gốc. Phương pháp này giúp giảm nhiễu tần số cao trong khi vẫn bảo toàn hình dạng tổng thể của phổ phản xạ. Quan sát Hình 5 cho thấy cấu trúc phổ chính trong toàn bộ dải bước sóng vẫn được giữ nguyên sau khi xử lý; tuy nhiên, các biến động nhỏ do nhiễu đã được làm trơn rõ rệt, đặc biệt tại vùng bước sóng ngắn nơi tín hiệu thường có mức dao động lớn hơn. Việc giảm nhiễu này giúp cải thiện độ ổn định của dữ liệu và tạo điều kiện thuận lợi cho quá trình xây dựng mô hình hồi quy.

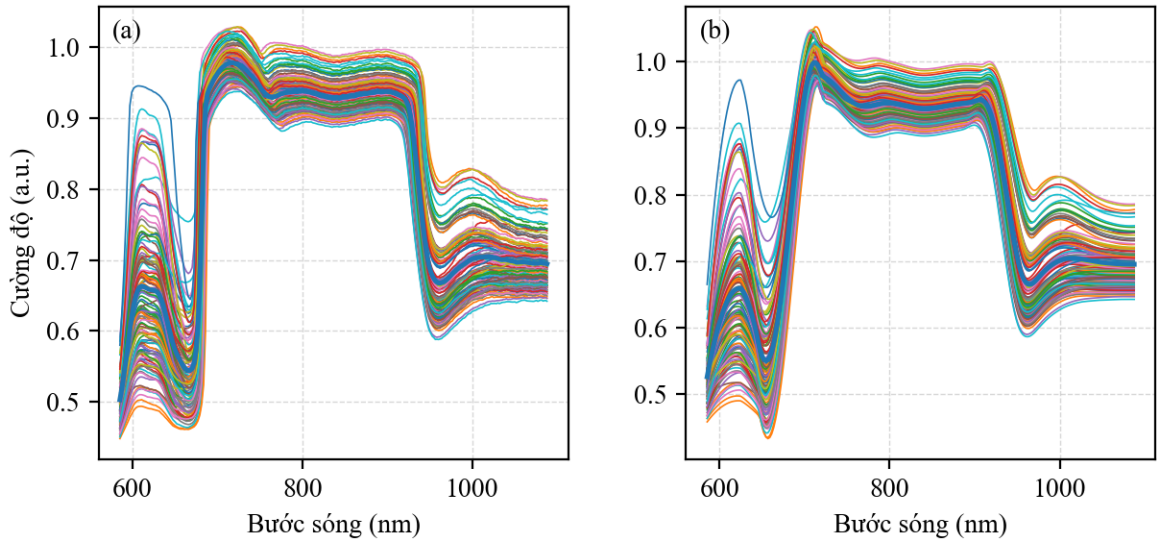
Phổ độ lệch chuẩn của dữ liệu sau tiền xử lý cũng được phân tích để đánh giá mức độ biến thiên của tín hiệu phổ giữa các mẫu theo từng bước sóng (Hình 6). Kết quả cho thấy độ lệch chuẩn đạt giá trị cao trong vùng khoảng 600–650 nm, với cực đại gần 620–630 nm, cho thấy sự khác biệt đáng kể giữa các mẫu tại vùng bước sóng này. Trong khoảng 700–900 nm, độ lệch chuẩn giảm xuống mức thấp hơn, phản ánh sự tương đồng tương đối của tín hiệu phổ giữa các mẫu. Tuy nhiên, từ khoảng 920 nm trở đi, độ lệch chuẩn tăng trở lại và xuất hiện một đỉnh khác

gần 940–950 nm trước khi giảm dần ở vùng bước sóng dài hơn. Các vùng bước sóng có độ lệch chuẩn cao thường chứa nhiều thông tin phân biệt giữa các mẫu, do đó có thể đóng vai trò quan trọng trong việc xây dựng mô hình dự đoán các chỉ tiêu chất lượng như °Brix. Đánh giá tiềm năng ứng dụng của phổ kế mini C11708MA trong dự đoán °Brix

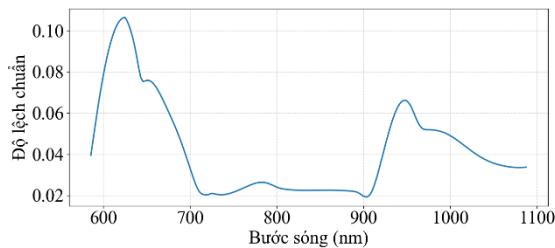
Kết quả nghiên cứu cho thấy dữ liệu phổ Vis–NIR trong dải bước sóng 640–1050 nm thu được từ phổ kế mini C11708MA có thể được sử dụng để xây dựng các mô hình dự đoán độ ngọt của dưa lưới với độ chính xác khả quan. Các mô hình hồi quy được phát triển đạt R_p trong khoảng 0,77–0,83 và sai số dự đoán $RMSEP$ từ 1,52 đến 1,75 °Brix. Theo các tiêu chí đánh giá thường được sử dụng trong phân tích phổ, các giá trị RPD từ 1,55 đến 1,79 cho thấy mô hình có khả năng dự đoán ở mức chấp nhận được đối với các ứng dụng định lượng sơ bộ hoặc phân loại chất lượng.

So sánh giữa các mô hình cho thấy phương pháp SVR đạt hiệu suất dự đoán tốt hơn so với PLS trong hầu hết các phương án tiền xử lý. Điều này có thể được giải thích bởi khả năng của SVR trong việc mô hình hóa các mối quan hệ phi tuyến giữa dữ liệu phổ và biến mục tiêu, trong khi PLS chủ yếu dựa trên giả định tuyến tính. Nhờ đó, SVR có thể khai thác hiệu quả hơn các đặc trưng phức tạp tiềm ẩn trong dữ liệu phổ, từ đó cải thiện độ chính xác của mô hình dự đoán.

Một điểm đáng chú ý là việc áp dụng các phương pháp tiền xử lý phổ phức tạp không mang lại sự cải thiện đáng kể về hiệu suất mô hình. Trong nhiều trường hợp, dữ liệu phổ gốc hoặc phổ được làm mượt bằng bộ lọc SG đã đủ để xây dựng các mô hình dự đoán hiệu quả. Điều này cho thấy tín hiệu phổ thu được từ phổ kế C11708MA có chất lượng tương đối ổn định và ít chịu ảnh hưởng bởi các yếu tố nhiễu như tán xạ ánh sáng hoặc trôi đường nền.



Hình 5. (a) Phổ thô và (b) phổ sau khi tiền xử lý bằng phương pháp Savitzky–Golay



Hình 6. Độ lệch chuẩn của dữ liệu phổ sau tiền xử lý

Bảng 2. So sánh hiệu suất các mô hình dự đoán độ ngọt của dưa lưới sử dụng các phương pháp thu nhận dữ liệu phổ khác nhau

Phổ kế	Mô hình	Hiệu suất	Tài liệu
C11708MA (Hamamatsu)	SVR	$R_p = 0,83$, $RMSE_p = 1,52$	Nghiên cứu này
F-750 (Felix Instruments, USA)	MLR	$R_p = 0,93$, $RMSE_p = 1,63$	Zeb et al. (2021)
USB-4000 (Ocean Optics)	PLS	$R_p = 0,78-0,80$, $RMSE_p = 0,83-0,92$	Yang et al. (2024)
PSR+ 3500 (Spectra Evolution)	PLS	$R_p = 0,80$	Kasampalis et al. (2025)

So với các nghiên cứu trước đây sử dụng các thiết bị thu nhận phổ khác nhau để dự đoán độ ngọt của dưa lưới (Bảng 2), các mô hình trong nghiên cứu này cho thấy hiệu suất dự đoán ở mức tương đương. Trong nghiên cứu này, phổ kế mini C11708MA kết hợp với mô hình SVR đạt $R_p = 0,83$ và $RMSE_p = 1,52$ °Brix. Kết quả này nằm trong khoảng hiệu suất tương tự với một số nghiên cứu trước đây. Chẳng hạn, Yang et al. (2024) sử dụng phổ kế USB-4000 để dự đoán độ ngọt của dưa lưới và báo cáo R_p trong khoảng 0,78–0,80 với $RMSE_p$ từ 0,83–0,92 °Brix. Trong khi đó, Zeb et al. (2021) sử dụng thiết bị thương mại F-750 để đánh giá độ ngọt của dưa lưới và đạt $R_p = 0,93$ với $RMSE_p = 1,63$ °Brix, cho thấy sai số dự đoán tương đương với nghiên cứu hiện tại.

Bên cạnh đó, Kasampalis et al. (2025) sử dụng hệ phổ kế PSR+ 3500 kết hợp với mô hình PLS cũng ghi nhận hiệu suất dự đoán SSC với R_p khoảng 0,80. Kết quả này cho thấy rằng, khi được kết hợp với các phương pháp xử lý dữ liệu và mô hình hóa phù hợp, các hệ thống quang phổ khác nhau, bao gồm cả thiết bị thương mại có độ phân giải cao và phổ kế mini, đều có thể đạt hiệu suất dự đoán đáng tin cậy.

Cần lưu ý rằng hiệu suất của mô hình dự đoán có thể bị ảnh hưởng bởi các yếu tố liên quan đến điều kiện đo, trong đó độ sạch của bề mặt dưa lưới là một yếu tố tiềm năng. Sự hiện diện của bụi bẩn, tạp chất hoặc độ ẩm trên bề mặt có thể làm thay đổi đặc tính phản xạ và gây nhiễu tín hiệu phổ, từ đó ảnh hưởng đến độ chính xác của mô hình. Trong phạm vi nghiên cứu này, định lượng ảnh hưởng của các mức độ sạch bề mặt khác nhau đến kết quả dự đoán chưa

được khảo sát. Nội dung này được xem xét trong các nghiên cứu tiếp theo nhằm đánh giá đầy đủ hơn tính ổn định và khả năng ứng dụng thực tế của phương pháp đề xuất.

Nhìn chung, kết quả nghiên cứu cho thấy phổ kế mini C11708MA, một dạng phổ kế nhỏ gọn và có chi phí tương đối thấp, có tiềm năng được sử dụng trong các hệ thống đánh giá chất lượng trái cây theo thời gian thực. Khi kết hợp với các kỹ thuật tiền xử lý phổ và lựa chọn bước sóng phù hợp, các mô hình dự đoán độ ngọt có thể đạt độ chính xác tương đương với các hệ thống phổ kế cầm tay hoặc thiết bị quang phổ có kích thước lớn hơn.

4. KẾT LUẬN

Khả năng ứng dụng của phổ phản xạ Vis-NIR trong dải bước sóng 640–1050 nm kết hợp với các mô hình học máy nhằm dự đoán độ ngọt ($^{\circ}$ Brix) của dưa lưới mà không phá hủy mẫu được đánh giá trong nghiên cứu. Kết quả phân tích phổ cho thấy dữ liệu thu được có tính ổn định cao và thể hiện các đặc

trung hấp thụ liên quan đến sắc tố bề mặt, cấu trúc mô quả cũng như các thành phần hóa học. Điều này cho thấy phổ Vis-NIR trong khoảng bước sóng nghiên cứu có tiềm năng cung cấp thông tin hữu ích cho việc dự đoán các chỉ tiêu chất lượng nội tại của quả.

Các mô hình PLS và SVR được xây dựng từ dữ liệu phổ thu được bằng phổ kế mini C11708MA với nhiều phương pháp tiền xử lý khác nhau. Kết quả cho thấy SVR đạt hiệu suất dự đoán tốt hơn PLS. Mô hình SVR đạt kết quả cao nhất với tiền xử lý SG, đạt $R_p = 0,83$, $RMSE_p = 1,52$ $^{\circ}$ Brix và $RPD_p = 1,79$. Các giá trị RPD cho thấy mô hình có khả năng dự đoán ở mức chấp nhận được cho đánh giá định lượng sơ bộ độ ngọt của dưa lưới.

Trong các nghiên cứu tiếp theo, việc mở rộng số lượng mẫu, đánh giá trên nhiều điều kiện canh tác khác nhau cũng như áp dụng các phương pháp lựa chọn bước sóng hoặc tối ưu mô hình có thể giúp cải thiện hơn nữa hiệu suất và khả năng ứng dụng thực tiễn của các mô hình dự đoán dựa trên dữ liệu phổ.

TÀI LIỆU THAM KHẢO

- Cevoli, C., Iaccheri, E., Fabbri, A., & Ragni, L. (2024). Data fusion of FT-NIR spectroscopy and Vis/NIR hyperspectral imaging to predict quality parameters of yellow flesh “Jintao” kiwifruit. *Biosystems Engineering*, 237, 157–169. <https://doi.org/10.1016/j.biosystemseng.2023.12.011>
- Cozzolino, D., Liu, L., Cynkar, W. U., Dambergs, R. G., Janik, L., Colby, C. B., & Gishen, M. (2007). Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition. *Analytica Chimica Acta*, 588(2), 224–230. <https://doi.org/10.1016/j.aca.2007.01.079>
- Dhanoa, M. S., López, S., Sanderson, R., Lister, S. J., Barnes, R. J., Ellis, J. L., & France, J. (2023). Methodology adjusting for least squares regression slope in the application of multiplicative scatter correction to near-infrared spectra of forage feed samples. *Journal of Chemometrics*, 37(11), 3511. <https://doi.org/10.1002/cem.3511>
- Ercan, U., Sonmez, I., Kabaş, A., Kabas, O., Calık Zyambo, B., Gölükçü, M., & Paraschiv, G. (2024). Quantitative assessment of brix in grafted melon cultivars: a machine learning and regression-based approach. *Foods*, 13, 3858. <https://doi.org/10.3390/foods13233858>
- Firago, V. A., Levkovich, N. V., & Shuliko, K. I. (2022). Diffuse Reflectance Spectrophotometers Based on C12880MA and C11708MA Mini-Spectrometers Hamamatsu. *Devices and Methods of Measurements*, 13(2), 93–104. <https://doi.org/10.21122/2220-9506-2022-13-2-93-104>
- Gómez-García, R., Campos, D. A., Aguilar, C. N., & Madureira, A. R. (2020). Valorization of melon fruit (*Cucumis melo* L.) by-products: Phytochemical and Biofunctional properties with Emphasis on Recent Trends and Advances. *Trends in Food Science & Technology*, 99, 507–519. <https://doi.org/10.1016/j.tifs.2020.03.033>
- Hamamatsu Photonics K.K. (2022). *Mini-spectrometers (MS series): C10988MA-01 and C11708MA*. https://www.hamamatsu.com/content/dam/hamamatsu-photonics/sites/documents/99_SALES_LIBRAR_Y/ssd/c10988ma-01_etc_kacc1169e.pdf
- Han, C., Jifan, Y., Hao, T., Jinshan, Y., & Huirong, X. (2023). Evaluation of the optical layout and sample size on online detection of apple watercore and SSC using Vis/NIR spectroscopy. *Journal of Food Composition and Analysis*, 123, 105528. <https://doi.org/10.1016/j.jfca.2023.105528>
- Hotra, O., Firago, V., Levkovich, N., & Shuliko, K. (2022). Investigation of the Possibility of Using Microspectrometers Based on CMOS Photodiode

- Arrays in Small-Sized Devices for Optical Diagnostics. *Sensors*, 22(11).
<https://doi.org/10.3390/s22114195>
- Kasampalis, D. S., Tsouvaltzis, P., & Siomos, A. S. (2025). Assessment of melon fruit nutritional composition using VIS / NIR / SWIR spectroscopy coupled with chemometrics. *Horticulturae*, 11(6), 658.
<https://doi.org/10.3390/horticulturae11060658>
- Khumaidi, A., & Raafi'udin, R. (2022). Effects of oversampling smote and spectral transformations in the classification of mango cultivars using near-infrared spectroscopy. *International Journal on Advanced Science, Engineering and Information Technology*, 12(3), 1047–1053.
<https://doi.org/10.18517/ijaseit.12.3.16001>
- Kusumiyati, Mubarak, S., Sutari, W., & Hadiwijaya, Y. (2021). Application of spectra pre-treatments on firmness assessment of intact sapodilla using vis-nir spectroscopy. *IOP Conference Series: Earth and Environmental Science*, 644, 012001.
<https://doi.org/10.1088/1755-1315/644/1/012001>
- Lazim, S. S. R. M., Nawi, M. N., Bejo, S. K., Shariff, A. R. M., & Abdullah, N. (2022). Prediction and classification of soluble solid contents to determine the maturity level of watermelon using visible and shortwave near infrared spectroscopy. *International Food Research Journal*, 29(6), 1372–1379.
<https://doi.org/10.47836/ifrj.29.6.13>
- Li, H., Zhang, M., Shen, M., Zhang, Z., Zhang, B., Zhang, H., Hu, J., Ren, X., Xing, L., & Zhao, J. (2022). Effect of ambient temperature on the model stability of handheld devices for predicting the apple soluble solids content. *European Journal of Agronomy*, 133, 126430.
<https://doi.org/10.1016/j.eja.2021.126430>
- Liu, H., Wei, Z., Lu, M., Gao, P., Li, J., Zhao, J., & Hu, J. (2024). A Vis/NIR device for detecting moldy apple cores using spectral shape features. *Computers and Electronics in Agriculture*, 220, 108898.
<https://doi.org/10.1016/j.compag.2024.108898>
- Mekonnen, B. K., Yang, W., Hsieh, T.-H., Liaw, S.-K., & Yang, F.-L. (2020). Accurate prediction of glucose concentration and identification of major contributing features from hardly distinguishable near-infrared spectroscopy. *Biomedical Signal Processing and Control*, 59, 101923.
<https://doi.org/10.1016/j.bspc.2020.101923>
- Miloš, B., & Bensa, A. (2017). Prediction of soil organic carbon using VIS-NIR spectroscopy: Application to Red Mediterranean soils from Croatia. *Eurasian Journal of Soil Science*, 6(4), 365–373. <https://doi.org/10.3390/s25165045>
- Pahlawan, M. F. R., Wati, R. K., & Masithoh, R. E. (2021). Development of a low-cost modular VIS/NIR spectroscopy for predicting soluble solid content of banana. *IOP Conference Series: Earth and Environmental Science*, 644(1).
<https://doi.org/10.1088/1755-1315/644/1/012047>
- Pourdarbani, R., Sabzi, S., & Arribas, J. I. (2021). Nondestructive estimation of three apple fruit properties at various ripening levels with optimal Vis-NIR spectral wavelength regression data. *Heliyon*, 7(9).
<https://doi.org/10.1016/j.heliyon.2021.e07942>
- Pratiwi, E. Z. D., Pahlawan, M. F. R., Rahmi, D. N., Amanah, H. Z., & Masithoh, R. E. (2023). Non-destructive evaluation of soluble solid content in fruits with various skin thicknesses using visible-shortwave near-infrared spectroscopy. *Open Agriculture*, 8(1), 1–12.
<https://doi.org/10.1515/opag-2022-0183>
- Products, F., & Inspection, Q. (2019). Applications of Non-destructive Technologies for Agricultural and Food Products Quality Inspection. *Sensors*, 19, 846. <https://doi.org/10.3390/s19040846>
- Sohaib Ali Shah, S., Zeb, A., Qureshi, W. S., Malik, A. U., Tiwana, M., Walsh, K., Amin, M., Alasmay, W., & Alanazi, E. (2021). Mango maturity classification instead of maturity index estimation: A new approach towards handheld NIR spectroscopy. *Infrared Physics and Technology*, 115, 103639.
<https://doi.org/10.1016/j.infrared.2021.103639>
- Tan, H., Hu, Y., Ma, B., Yu, G., & Li, Y. (2024). An improved DCGAN model : Data augmentation of hyperspectral image for identification pesticide residues of Hami melon. *Food Control*, 157, 110168.
<https://doi.org/10.1016/j.foodcont.2023.110168>
- Wang, C., Luo, X., Guo, Z., Wang, A., Zhou, R., & Cai, J. (2025). Influence of the peel on online detecting soluble solids content of pomelo using Vis-NIR spectroscopy coupled with chemometric analysis. *Food Control*, 167, 110777.
<https://doi.org/10.1016/j.foodcont.2024.110777>
- Wang, G., Luo, Y., Liu, Y., Huo, Y., Ouyang, A., & Zhu, D. (2025). Relationship between optical properties and internal quality of Orah Mandarins during storage. *Scientific Reports*, 15, 12208.
<https://doi.org/10.1038/s41598-025-95261-x>
- Windig, W., Shaver, J., & Bro, R. (2008). Loopy MSC: A simple way to improve multiplicative scatter correction. *Applied Spectroscopy*, 62(10), 1153–1159.
<https://doi.org/10.1366/000370208786049097>
- Wu, G., Abid, M., Zerara, M., Coileain, C. O., Chang, C. R., & Wu, H. C. (2023). Miniaturized Computational Spectrometer. *IEEE Nanotechnology Magazine*, 17(6), 36–42.
<https://doi.org/10.1109/MNANO.2023.3316870>

- Yang, S., Guo, W., Wang, J., Li, S., & Zeng, S. (2024). Optical properties of melon tissues and Monte Carlo simulation of light propagation through melon fruit. *Postharvest Biology and Technology*, 207, 112616. <https://doi.org/10.1016/j.postharvbio.2023.112616>
- Yang, Y., Ma, S., Qi, F., Wang, F., & Xu, H. (2025). Application of vis/nir spectroscopy in the rapid and non-destructive prediction of soluble solid content in milk jujubes. *Agriculture*, 15, 1382. <https://doi.org/10.3390/agriculture15131382>
- Yang, Y., Zhao, C., Huang, W., Tian, X., Fan, S., Wang, Q., & Li, J. (2022). Optimization and compensation of models on tomato soluble solids content assessment with online Vis/NIRS diffuse transmission system. *Infrared Physics and Technology*, 121, 104050. <https://doi.org/10.1016/j.infrared.2022.104050>
- Zeb, A., Qureshi, W. S., Ghafoor, A., Malik, A., Imran, M., Iqbal, J., & Alanazi, E. (2021). Is this melon sweet? A quantitative classification for near-infrared spectroscopy. *Infrared Physics & Technology*, 114, 103645. <https://doi.org/10.1016/J.INFRARED.2021.103645>
- Zhang, T., Li, Y., & Wang, M. (2024). Prediction of soil organic carbon and total nitrogen affected by mine using Vis – NIR spectroscopy coupled with machine learning algorithms in calcareous soils. *Scientific Reports*, 14(1), 28014. <https://doi.org/10.1038/s41598-024-73761-6>
- Zheng, G., Ryu, D., Jiao, C., Xie, X., & Cui, X. (2019). Visible and Near-Infrared Reflectance Spectroscopy Analysis of a Coastal Soil Chronosequence. *Remote Sensing*, 11, 2336. <https://doi.org/10.3390/rs11202336>