**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KỸ THUẬT MÁY TÍNH**

**VÕ HUY HOÀNG**

**KHÓA LUẬN TỐT NGHIỆP**

**ĐO ĐƯỜNG HUYẾT KHÔNG XÂM LẤN SỬ DỤNG QUANG PHỔ CẬN HỒNG NGOẠI VÀ TRÍ TUỆ NHÂN TẠO**

**NON-INVASIVE BLOOD GLUCOSE MEASUREMENT USING NEAR-INFRAME SPECTROSCOPY AND ARTIFICIAL INTELLIGENCE**

**CỬ NHÂN NGÀNH KỸ THUẬT MÁY TÍNH**

**TP. HỒ CHÍ MINH, 2024**

**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA KỸ THUẬT MÁY TÍNH**

**VÕ HUY HOÀNG – 21522103**

**KHÓA LUẬN TỐT NGHIỆP**

**ĐO ĐƯỜNG HUYẾT KHÔNG XÂM LẤN SỬ DỤNG QUANG PHỔ CẬN HỒNG NGOẠI VÀ TRÍ TUỆ NHÂN TẠO**

**NON-INVASIVE BLOOD GLUCOSE MEASUREMENT USING NEAR-INFRAME SPECTROSCOPY AND ARTIFICIAL INTELLIGENCE**

**CỬ NHÂN NGÀNH KỸ THUẬT MÁY TÍNH**

**GIẢNG VIÊN HƯỚNG DẪN**

**TS. PHẠM QUỐC HÙNG**

**TP. HỒ CHÍ MINH, 2024**

THÔNG TIN HỘI ĐỒNG CHẤM KHÓA LUẬN TỐT NGHIỆP

Hội đồng chấm khóa luận tốt nghiệp, thành lập theo Quyết định số …………………… ngày ………………….. của Hiệu trưởng Trường Đại học Công nghệ Thông tin.

LỜI CẢM ƠN

**MỤC LỤC**

[TÓM TẮT KHÓA LUẬN 2](#_Toc186018460)

[CHƯƠNG 1. GIỚI THIỆU VỀ ĐỀ TÀI 3](#_Toc186018461)

[CHƯƠNG 2: TỔNG QUAN TÀI LIỆU 6](#_Toc186018462)

[2.1. Các phương pháp theo dõi đường huyết 7](#_Toc186018463)

[2.2. Các nghiên cứu về phương pháp đo Glucose không xâm lấn sử dụng NIR 10](#_Toc186018464)

[2.2.1. Tình hình nghiên cứu trong nước 10](#_Toc186018465)

[2.2.2. Tình hình nghiên cứu trên thế giới 11](#_Toc186018466)

[2.3. Lý thuyết về phương pháp đo glucose không xâm lấn VIS - NIR 12](#_Toc186018467)

[2.3.1. Quang phổ VIS – NIR 12](#_Toc186018468)

[2.3.2. Ảnh hưởng quang phổ đến Glucose 13](#_Toc186018469)

[2.3.4. Độ hấp thụ của ánh sáng trong môi trường đo 14](#_Toc186018470)

[2.3.5. Ảnh hưởng của ánh sáng 15](#_Toc186018471)

[2.3.6. Lựa chọn khoảng cách đo 16](#_Toc186018472)

[CHƯƠNG 3: MỤC TIÊU VÀ PHƯƠNG PHÁP 19](#_Toc186018473)

[3.1. Mục tiêu đặt ra 19](#_Toc186018474)

[3.2. Triển khai phần cứng 19](#_Toc186018475)

[3.2.1. Tổng quan sơ đồ kết nối phần cứng 19](#_Toc186018476)

[3.2.2. Chi tiết các linh kiện phần cứng 20](#_Toc186018477)

[3.2.2.1. Máy tính nhúng 20](#_Toc186018478)

[3.2.2.2. Cảm biến quang phổ 21](#_Toc186018479)

[3.2.2.3. Nguồn sáng 23](#_Toc186018480)

[3.2.2.4. Màn hình OLED 24](#_Toc186018481)

[3.2.2.5. USB Dongle 4G 25](#_Toc186018482)

[3.2.2.6. Mạch tăng áp 25](#_Toc186018483)

[3.2.2.7. Mạch giảm áp 26](#_Toc186018484)

[3.3. Quy trình lấy mẫu 27](#_Toc186018485)

[3.4. Kịch bản lấy mẫu 28](#_Toc186018486)

[3.5. Triển khai phần mềm 28](#_Toc186018487)

[3.5.1. Tiền xử lý dữ liệu thô 29](#_Toc186018488)

[3.5.1.1. Multiplicative Scater Correction 30](#_Toc186018489)

[3.5.1.2. Standard Scaler 31](#_Toc186018490)

[3.5.2. Mô hình huấn luyện 32](#_Toc186018491)

[3.5.2.1 Support Vector Regression 32](#_Toc186018492)

[3.5.2.2. Gaussian Process Regression 34](#_Toc186018493)

[3.5.2.3. Random Forest Regression 35](#_Toc186018494)

[3.5.3. Đánh giá mô hình 36](#_Toc186018495)

[3.5.3.1. Độ chính xác phân tích 37](#_Toc186018496)

[3.5.3.2. Độ chính xác lâm sàn 38](#_Toc186018497)

[3.5.4. Lựa chọn đặc trưng F-test 39](#_Toc186018498)

[3.6. IoMT 40](#_Toc186018499)

[3.6.1. Phương thức 4G 41](#_Toc186018500)

[3.6.2. Firebase 42](#_Toc186018501)

[3.7. Tổng quan sơ đồ giải thuật của thiết bị 42](#_Toc186018502)

[CHƯƠNG 4: KẾT QUẢ VÀ THẢO LUẬN 44](#_Toc186018503)

[4.1. Tập dữ liệu 44](#_Toc186018504)

[4.2. Tiền xử lý dữ liệu 47](#_Toc186018505)

[4.2.1. Standard Scaler 48](#_Toc186018506)

[4.2.2. Multiplicative Scatter Correction 51](#_Toc186018507)

[4.2.3. Kết hợp Multiplicative Scatter Correction và Standard Scaler 53](#_Toc186018508)

[4.3. Huấn luyện mô hình máy học hồi quy 57](#_Toc186018509)

[4.3.1. Độ chính xác phân tích 57](#_Toc186018510)

[4.3.2. Độ chính xác lâm sàn 61](#_Toc186018511)

[4.5. Mô hình phần cứng 68](#_Toc186018512)

[4.5.1. Thiết kế 3D đề tài 68](#_Toc186018513)

[4.5.2. Sơ đồ kết nối phần cứng 69](#_Toc186018514)

[4.6. Ứng dụng điện thoại 70](#_Toc186018515)

[4.7. Đánh giá mô hình 72](#_Toc186018516)

[CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN 74](#_Toc186018517)

[5.1. Kết luận 74](#_Toc186018518)

[5.1.1. Những điều đạt được 74](#_Toc186018519)

[5.2. Hướng phát triển đề tài 75](#_Toc186018520)

[TÀI LIỆU THAM KHẢO 77](#_Toc186018521)

DANH MỤC HÌNH

[Hình 1. 1: Xu hướng bệnh tiểu đường 3](#_Toc186019030)

[Hình 2. 1: Hai phương pháp phổ biến trong theo dõi đường huyết............................ 7](#_Toc186019034)

[Hình 2. 2:Ảnh hưởng nồng độ glucose đến ánh sáng 15](#_Toc186019035)

[Hình 2. 3: Đường đi ánh sáng 16](#_Toc186019036)

[Hình 2. 4: Khoảng cách được bố trí của cảm biến và đèn. 18](#_Toc186019037)

[Hình 3. 1: Sơ đồ kết nối phần cứng...........................................................................19](#_Toc186019038)

[Hình 3. 2: Cấu tạo của cảm biến AS7265x 22](#_Toc186019039)

[Hình 3. 3: Dãi quang phổ của bóng đèn Halogen 24](#_Toc186019040)

[Hình 3. 4: Quy trình lấy mẫu 27](#_Toc186019041)

[Hình 3. 5: Sơ đồ thể hiện quy trình dự đoán nồng độ glucose. 29](#_Toc186019042)

[Hình 3. 6: Support Vector Regression 33](#_Toc186019043)

[Hình 3. 7: Clarke Grid Error 39](#_Toc186019044)

[Hình 3. 8: Mô hình IoMT triển khai trong đề tài 41](#_Toc186019045)

[Hình 3. 9: Sơ đồ giải thuật của thiết bị. 42](#_Toc186019046)

[Hình 4. 1: Phân phối VIS - NIR với giá trị Glucose.................................................45](#_Toc186025040)

[Hình 4. 2: Hệ số tương quan Pearson 47](#_Toc186025041)

[Hình 4. 3: Giá trị khi sử dụng StandardScaler ở bước sóng 410nm và 435nm 48](#_Toc186025042)

[Hình 4. 4: Giá trị khi sử dụng StandardScaler ở bước sóng 460nm và 485nm 49](#_Toc186025043)

[Hình 4. 5: Giá trị khi sử dụng StandardScaler ở bước sóng 510nm và 535nm 49](#_Toc186025044)

[Hình 4. 6: Giá trị khi sử dụng StandardScaler ở bước sóng 560nm và 585nm 49](#_Toc186025045)

[Hình 4. 7: Giá trị khi sử dụng StandardScaler ở bước sóng 610nm và 645nm 50](#_Toc186025046)

[Hình 4. 8: Giá trị khi sử dụng StandardScaler ở bước sóng 680nm và 705nm 50](#_Toc186025047)

[Hình 4. 9: Giá trị khi sử dụng StandardScaler ở bước sóng 730nm và 760nm 50](#_Toc186025048)

[Hình 4. 10: Giá trị khi sử dụng StandardScaler ở bước sóng 810nm và 860nm 51](#_Toc186025049)

[Hình 4. 11: Giá trị khi sử dụng StandardScaler ở bước sóng 900nm và 940nm 51](#_Toc186025050)

[Hình 4. 12: Xu hướng giá trị VIS – NIR đến Glucose khi sử dụng MSC 52](#_Toc186025051)

[Hình 4. 13: Giá trị hệ số tương quan Pearson khi sử dụng MSC 52](#_Toc186025052)

[Hình 4. 14: Giá trị bước sóng 410nm – 435nm khi kết hợp phương pháp 54](#_Toc186025053)

[Hình 4. 15: Giá trị bước sóng 460nm – 485nm khi kết hợp phương pháp 55](#_Toc186025054)

[Hình 4. 16: Giá trị bước sóng 510nm – 535nm khi kết hợp phương pháp 55](#_Toc186025055)

[Hình 4. 17: Giá trị bước sóng 585nm – 560nm khi kết hợp phương pháp 55](#_Toc186025056)

[Hình 4. 18: Giá trị bước sóng 610nm – 645nm khi kết hợp phương pháp 56](#_Toc186025057)

[Hình 4. 19: Giá trị bước sóng 680nm – 705nm khi kết hợp phương pháp 56](#_Toc186025058)

[Hình 4. 20: Giá trị bước sóng 730nm – 760nm khi kết hợp phương pháp 56](#_Toc186025059)

[Hình 4. 21: Giá trị bước sóng 810nm – 860nm khi kết hợp phương pháp 57](#_Toc186025060)

[Hình 4. 22: Giá trị bước sóng 900nm – 940nm khi kết hợp phương pháp 57](#_Toc186025061)

[Hình 4. 23: Giá trị phân tán với Standard Scaler. (a) RF, (b) SVR, (c) GPR 59](#_Toc186025062)

[Hình 4. 24: Giá trị phân tán với MSC. (a) RF, (b) SVR, (c) GPR 60](#_Toc186025063)

[Hình 4. 25: Giá trị phân tán với MSC + SS. (a) RF, (b) SVR, (c) GPR 61](#_Toc186025064)

[Hình 4. 26: Phân phối trong CGE với phương pháp Standard Scaler 62](#_Toc186025065)

[Hình 4. 27: Phân phối trong CGE với phương pháp MSC 63](#_Toc186025066)

[Hình 4. 28: Phân phối trong CGE với phương pháp MSC và Standard Scaler 64](#_Toc186025067)

[Hình 4. 29: F-Score của 18 bước sóng 65](#_Toc186025068)

[Hình 4. 30: Phân phối sau khi chọn F - test trong CGE với phương pháp Standard Scaler 67](#_Toc186025069)

[Hình 4. 31: Phân phối sau khi chọn F - test trong CGE với phương pháp MSC 67](#_Toc186025070)

[Hình 4. 32: Phân phối sau khi chọn F - test trong CGE với phương pháp MSC và SS 67](#_Toc186025071)

[Hình 4. 33: Mô hình 3D mặt trước 68](#_Toc186025072)

[Hình 4. 34: Mô hình 3D bên trong 68](#_Toc186025073)

[Hình 4. 35: Ứng dụng ProSugr theo dõi nồng độ đường huyết. 70](#_Toc186025074)

[Hình 4. 36: Giao diện chính của ứng dụng 71](#_Toc186025075)

[Hình 4. 37: Ứng dụng ở mục “Detail” 72](#_Toc186025076)

[Hình 4. 38: Đóng gói thiết bị 72](#_Toc186025077)

DANH MỤC BẢNG

[Bảng 2. 1: So sánh ưu và nhược điểm các phương pháp. 9](#_Toc186025092)

[Bảng 3. 1: So sánh đèn LED và đèn Halogen...........................................................23](#_Toc186025097)

[Bảng 4. 1: Phạm vi giá trị các bước sóng..................................................................45](#_Toc186025103)

[Bảng 4. 2: Hệ số tương quan Pearrson với giá trị Glucose của từng bước sóng 47](#_Toc186025104)

[Bảng 4. 3: Chi tiết hệ số tương quan Pearson với Glucose 53](#_Toc186025105)

[Bảng 4. 4: Kết quả mô hình học máy 58](#_Toc186025106)

[Bảng 4. 5: Giá trị Clarke Grid Error của các mô hình với Standard Scaler 61](#_Toc186025107)

[Bảng 4. 6: Giá trị Clarke Grid Error của các mô hình với MSC 62](#_Toc186025108)

[Bảng 4. 7: Giá trị Clarke Grid Error của các mô hình khi kết hợp 2 phương pháp tiền xử lý 63](#_Toc186025109)

[Bảng 4. 8: Giá trị R2 và RMSE khi áp dụng F-Test 66](#_Toc186025110)

[Bảng 4. 9: Sơ đồ kết nối phần cứng 69](#_Toc186025111)

[Bảng 4. 10 – Độ chênh lệch giá trị đo từ thiết bị và ACCU CHEK 73](#_Toc186025112)

DANH MỤC TỪ VIẾT TẮT

TÓM TẮT KHÓA LUẬN

Bệnh tiểu đường đang ngày càng gia tăng với nhiều diễn biến phức tạp và để lại các di chứng nghiêm trọng. Đối với người bệnh, việc theo dõi thường xuyên nồng độ glucose trong máu là yếu tố thiết yếu nhằm duy trì mức đường huyết trong giới hạn an toàn, từ đó giảm nguy cơ biến chứng.

Hiện nay, có nhiều phương pháp kiểm tra nồng độ glucose trong máu, trong đó phương pháp đo xâm lấn là phổ biến nhất. Tuy nhiên, phương pháp này tồn tại nhiều hạn chế như gây tổn thương da, chảy máu, đau đớn và nguy cơ nhiễm trùng, đặc biệt đối với người chưa quen thao tác. Để khắc phục những nhược điểm này, nghiên cứu tập trung phát triển thiết bị đo đường huyết không xâm lấn, dựa trên nguyên lý tác động của các bước sóng khả kiến (VIS) và cận hồng ngoại (NIR) lên nồng độ glucose.

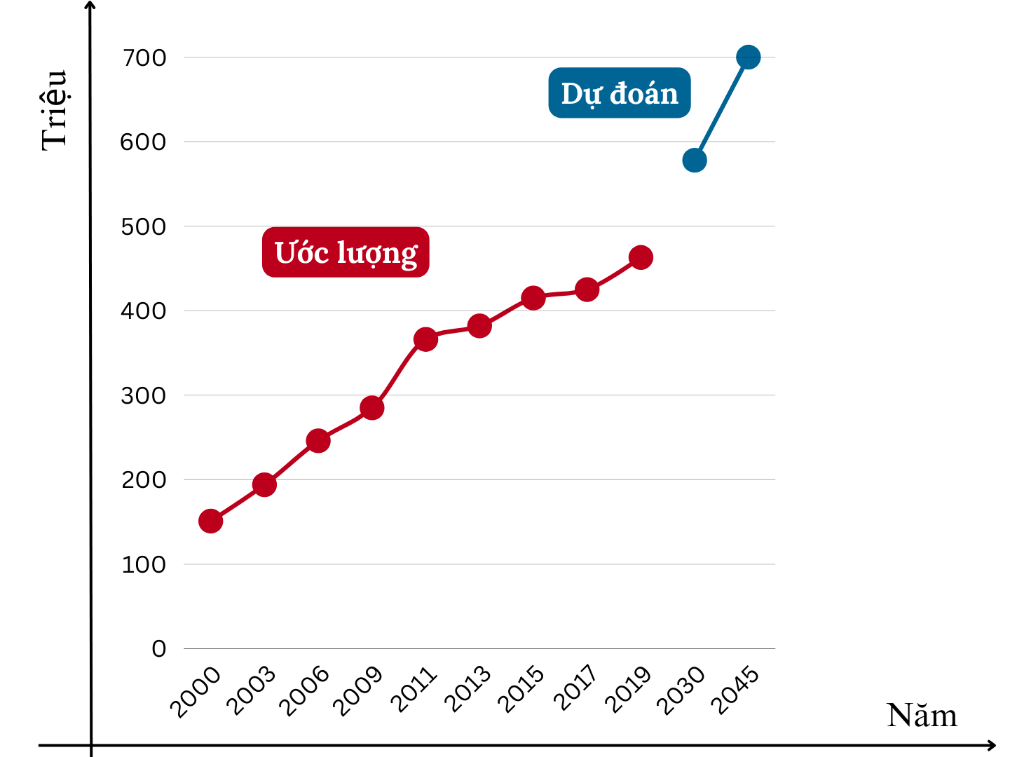
Trong quá trình thực hiện, đề tài đã thiết kế và xây dựng thành công mô hình đo nồng độ glucose không xâm lấn với độ chính xác trung bình trên 90% khi so sánh với máy đo glucose xâm lấn ACCU CHEK Instant. Kết quả cũng cho thấy một số bước sóng, bao gồm …, …, …, …, …, …, …, …, có tác động đáng kể đến nồng độ glucose.

Ngoài ra, thiết bị còn được tích hợp công nghệ IoT, cho phép người dùng theo dõi dữ liệu nồng độ glucose ở thời điểm đo và lưu trữ thông tin trong một khoảng thời gian nhất định. Tính năng này hỗ trợ người dùng nhận biết sự biến đổi nồng độ glucose, từ đó đưa ra các điều chỉnh phù hợp và kịp thời để cải thiện sức khỏe.

CHƯƠNG 1. GIỚI THIỆU VỀ ĐỀ TÀI

Bệnh tiểu đường đã trở thành một trong những căn bệnh mãn tính nghiêm trọng và phổ biển trong những năm trở lại đây, gây ra các biến chứng nguy hiểm đến tính mạng, khiến người bệnh tàn tật, giảm tuổi thọ và cực kỳ tốn kém [1].

Tỷ lệ mắc bệnh tiểu đường toàn cầu ở người từ 20–79 tuổi vào năm 2021 được ước tính là 10,5% (536,6 triệu người), và sẽ tăng lên 12,2% (783,2 triệu người) vào năm 2045. Tỷ lệ mắc bệnh tiểu đường ở nam và nữ là tương đương và cao nhất ở nhóm tuổi 75–79. Tỷ lệ mắc bệnh (vào năm 2021) ước tính cao hơn ở khu vực thành thị (12,1%) so với khu vực nông thôn (8,3%), và ở các quốc gia có thu nhập cao (11,1%) so với các quốc gia có thu nhập thấp (5,5%). Sự gia tăng tương đối lớn nhất về tỷ lệ mắc bệnh tiểu đường giữa năm 2021 và 2045 dự kiến sẽ xảy ra ở các quốc gia có thu nhập trung bình (21,1%) so với các quốc gia có thu nhập cao (12,2%) và thu nhập thấp (11,9%). Chi phí chăm sóc sức khỏe toàn cầu liên quan đến bệnh tiểu đường ước tính là 966 tỷ USD vào năm 2021 và dự báo sẽ đạt 1.054 tỷ USD vào năm 2045 [2].



Hình 1. 1: Xu hướng bệnh tiểu đường [1]

Duy trì mức glucose ổn định là yếu tố quan trọng đối với sức khỏe con người, vì glucose là nguồn năng lượng chính cho tế bào, đặc biệt là tế bào não và cơ. Khi lượng glucose trong máu vượt quá mức bình thường (tăng đường huyết), cơ thể đối mặt với nguy cơ tổn thương nhiều cơ quan như tim, thận, mắt và hệ thần kinh. Ngược lại, lượng glucose quá thấp (hạ đường huyết) có thể gây mệt mỏi, chóng mặt và trong trường hợp nghiêm trọng, dẫn đến mất ý thức [3].

Sự cân bằng nồng độ glucose được kiểm soát bởi insulin và glucagon - hai hormone chính chịu trách nhiệm điều chỉnh quá trình chuyển hóa glucose trong cơ thể. Bất kỳ sự rối loạn nào trong quá trình này đều có thể ảnh hưởng đến các hệ thống quan trọng như tim mạch, thần kinh và tiêu hóa. Do đó, việc kiểm soát lượng glucose trong máu không chỉ đảm bảo năng lượng cho cơ thể mà còn ngăn ngừa những vấn đề sức khỏe nghiêm trọng liên quan đến các cơ quan và hệ thống chức năng [3].

Theo [4], bệnh tiểu đường được chẩn đoán khi nồng độ glucose duy trì cao hơn các ngưỡng xác định. Cụ thể, nồng độ glucose khi nhịn ăn từ 126 mg/dL (7.0 mmol/L) trở lên hoặc nồng độ glucose 2 giờ sau ăn đạt từ 200 mg/dL (11.1 mmol/L) trở lên đều được xem là tiêu chí chẩn đoán bệnh. Ở người khỏe mạnh, mức nồng độ glucose thường dao động trong khoảng an toàn: từ 70–100 mg/dL (3.9–5.6 mmol/L) khi nhịn ăn và từ 70–140 mg/dL (3.9–7.8 mmol/L) khi đo ngẫu nhiên. Việc duy trì mức nồng độ glucose ổn định đóng vai trò quan trọng trong việc bảo vệ sức khỏe, bởi nếu không được kiểm soát, tăng đường huyết kéo dài có thể dẫn đến nhiều biến chứng nguy hiểm. Do đó, theo dõi thường xuyên để điều chỉnh lối sống phù hợp là biện pháp thiết yếu để phòng ngừa và quản lý bệnh tiểu đường hiệu quả..

Hiện nay, việc kiểm tra nồng độ glucose trong máu chủ yếu được thực hiện thông qua ba phương pháp: xâm lấn, ít xâm lấn và không xâm lấn. Trong đó, phương pháp xâm lấn vẫn phổ biến nhờ độ chính xác cao và tính tiện lợi, nhưng lại gây đau đớn và để lại tổn thương trên da do phải sử dụng kim lấy mẫu máu. Điều này đặc biệt gây khó chịu đối với bệnh nhân tiểu đường type 2, những người cần theo dõi nồng độ glucose máu thường xuyên để can thiệp kịp thời. Việc lấy mẫu máu nhiều lần trong ngày không chỉ tạo cảm giác đau đớn mà còn ảnh hưởng đến chất lượng cuộc sống của người bệnh.

Để khắc phục hạn chế này, các phương pháp ít xâm lấn như đo nồng độ glucose trong dịch mô kẽ thông qua cảm biến đã được phát triển. Tuy nhiên, các phương pháp này yêu cầu thay thế đầu dò định kỳ và có chi phí sử dụng cao, khiến chúng chưa thể trở thành giải pháp tối ưu cho nhiều bệnh nhân.

Do đó, nghiên cứu này tập trung vào phát triển một phương pháp không xâm lấn để đo nồng độ glucose trong máu, nhằm giảm thiểu đau đớn, đơn giản hóa quy trình đo, và hỗ trợ giám sát nồng độ glucose hiệu quả. Hệ thống được xây dựng sẽ ưu tiên chi phí thấp, tích hợp công nghệ IoT, cung cấp giao diện thân thiện với người dùng, đồng thời hỗ trợ lưu trữ lịch sử giám sát để bệnh nhân dễ dàng theo dõi và quản lý sức khỏe của mình.

# CHƯƠNG 2: TỔNG QUAN TÀI LIỆU

Tiểu đường xảy ra chủ yếu do sự thiếu hụt hoặc suy giảm chức năng của insulin, một hormone quan trọng được sản xuất bởi các tế bào β trong đảo Langerhans của tuyến tụy. Insulin là một hormone polypeptide có vai trò điều chỉnh mức glucose trong máu, giúp cung cấp năng lượng cho tế bào. Nó thúc đẩy quá trình lưu trữ glucose tại gan, cơ và mô mỡ. Khi lượng insulin không đủ hoặc không hoạt động hiệu quả, glucose không được chuyển hóa bình thường, dẫn đến tình trạng nồng độ glucose vượt quá mức cho phép. Hậu quả là các tế bào không nhận đủ năng lượng để hỗ trợ quá trình chuyển hóa, gây ra các rối loạn trong cơ thể [5].

Tiểu đường được chia thành hai loại chính, dựa trên cơ chế gây bệnh và đặc điểm của từng loại [1]:

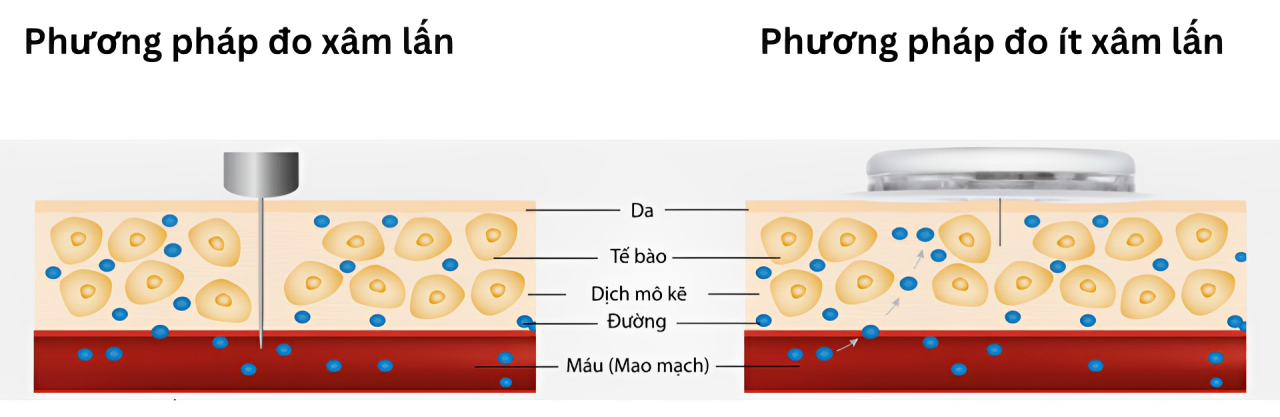
* Tiểu đường type 1: Tiểu đường có đặc trưng chính là bệnh nhân không còn insulin trong cơ thể hoặc còn rất ít khi hệ thống miễn dịch của cơ thể bị tấn công, tế bào Beta tuyến tụy bị phá hủy khiến insulin ngừng sản xuất và từ đó lượng đường trong máu tăng mạnh.
* Tiểu đường type 2: Đây là loại tiểu đường phổ biến nhất hiện nay khi nguyên nhân chính dẫn đến tình trạng này là việc dung nạp đường từ thức ăn, cơ thể ít vận động, không kiểm soát được cân nặng và không tiêu thụ hết lượng đường được hấp thụ. Tiểu đường type 2 xảy ra do sự cạn kiệt các tế bào Beta tuyến tụy dẫn tới giảm nồng độ Insulin được sản xuất.

Nếu không được phát hiện và điều trị kịp thời, tiểu đường có thể gây ra nhiều biến chứng trong cơ thể như [3]:

* Các biến chứng về tim mạch: Đường huyết không được kiểm soát khiến các mạch máu rất dễ bị tổn thương gây nên các bệnh như xơ vữa động mạch, tắc nghẽn mạch máu, tăng huyết áp, đột quỵ, nhồi máu cơ tim,...
* Biến chứng về hô hấp: Dễ mắc các chứng bệnh về viêm phổi, viêm phế quản hơn so với người bình thường.
* Biến chứng ở thận: Ở các bệnh nhân tiểu đường, thận sẽ phải hoạt động năng suất hơn nhằm thanh lọc và cân bằng máu, trong khi các biến chứng của tiểu đường có thể gây áp lực lên các mạch máu ở thận. Những tổn thương này có thể dẫn đến suy thận và giảm khả năng phục hồi ở thận.

## 2.1. Các phương pháp theo dõi đường huyết

Thay vì sử dụng phương pháp truyền thống, hiện nay đã có nhiều nghiên cứu và các đề xuất về các phương pháp theo dõi nhanh nồng độ glucose:



Hình 2. 1: Hai phương pháp phổ biến trong theo dõi nồng độ glucose hiện nay

Phương pháp đo xâm lấn sử dụng một kim nhỏ, chích trực tiếp vào ngón tay để có thể xâm nhập tới vùng mao mạch và lấy máu để đo nồng độ glucose. Máy đo sử dụng một điện cực có chứa enzyme glucose oxidase. Khi máu được đưa vào thiết bị, glucose trong mẫu máu sẽ phản ứng với enzyme glucose oxidase, tạo ra hydrogen peroxide (H₂O₂) và axit gluconic. Hydrogen peroxide (H₂O₂) sinh ra từ phản ứng trên sẽ phản ứng tiếp với một điện cực thứ hai, tạo ra dòng điện. Mức độ dòng điện này tỷ lệ thuận với lượng glucose có trong mẫu máu. Máy đo sẽ sử dụng giá trị dòng điện đo được để tính toán và hiển thị nồng độ glucose trong máu. Số liệu này sẽ được hiển thị trên màn hình thiết bị, cho phép người dùng biết được mức đường huyết hiện tại.

Tuy nhiên, phương pháp đo nồng độ glucse xâm lấn cũng tồn tại một số hạn chế nhất định. Đầu tiên, nếu lượng máu lấy được quá ít, việc thực hiện phản ứng hóa học có thể không chính xác. Thêm vào đó, nếu ngón tay không được vệ sinh kỹ máu có thể bị pha tạp, ảnh hưởng đến kết quả đo hoặc khi độ ẩm môi trường cao, các chất phản ứng hóa học trên đầu đo có thể bị ảnh hưởng, làm gián đoạn và sai lệch kết quả đo. Bên cạnh đó, việc sử dụng kim chích để lấy máu không chỉ gây đau mà còn có thể dẫn đến tình trạng bầm tím tại vị trí chích, nguy cơ nhiễm trùng nếu không tuân thủ quy trình vệ sinh đúng cách trước và sau lấy mẫu.

Bên cạnh phương pháp đo nồng độ glucose xâm lấn, đo ít xâm lấn đã ngày càng trở nên phổ biến nhờ vào việc khắc phục một phần những hạn chế của phương pháp truyền thống. Máy đo bao gồm 2 phần chính: Đầu đọc cầm tay và cảm biến đo. Một trong những phương pháp ít xâm lấn nổi bật là đo nồng độ glucose qua dịch kẽ. Máy đo thông qua dịch kẽ có nhiều ưu điểm, như việc kim chỉ đâm xuyên vào dịch mô kẽ, giảm đáng kể cảm giác đau đớn so với phương pháp truyền thống. Ngoài ra, các thiết bị này còn cho phép đo đường huyết liên tục thông qua các cảm biến dán lên da, giúp theo dõi mức glucose một cách liên tục và thuận tiện.

Tuy nhiên, phương pháp này cũng tồn tại một số nhược điểm, trong đó chi phí cao là một trong những nhược điểm lớn, khi cảm biến cần phải thay thế định kỳ do sự hao mòn của nó trong quá trình sử dụng. Ví dụ, sản phẩm đo đường huyết "FreeStyle Libre" của Abbott yêu cầu thay đầu kim sau mỗi 14 ngày sử dụng, điều này làm tăng chi phí và không phải lúc nào cũng phù hợp với người sử dụng có ngân sách hạn chế.

Bảng 2. 1: So sánh ưu và nhược điểm các phương pháp.

|  |  |  |
| --- | --- | --- |
| Phương pháp | Ưu điểm | Nhược điểm |
| Xâm lấn | * Độ chính xác cao, đáng tin cậy, có thể cung cấp kết quả nhanh và chính xác. * Chi phí thấp và dễ dàng tiếp cận. Phổ biến trong các cơ sở y tế. * Các thiết bị đo có sẵn và dễ dàng sử dụng trong nhiều trường hợp. | * Quá trình lấy mẫu gây khó chịu cho bệnh nhân, khiến việc thực hiện thường xuyên trở nên không thoải mái. * Có nguy cơ nhiễm trùng, bầm tím tại các điểm lấy máu. * Không phù hợp cho việc theo dõi đường huyết liên tục, đòi hỏi người bệnh phải lấy mẫu nhiều lần trong ngày. |
| Ít xâm lấn | * Không cần lấy máu, giúp giảm thiểu cảm giác đau đớn. * Phù hợp với việc theo dõi liên tục, đặc biệt trong các phương pháp đo đường huyết qua dịch kẽ. * Dễ sử dụng và thuận tiện cho bệnh nhân theo dõi tình trạng đường huyết hàng ngày. * Có thể theo dõi đường huyết trong thời gian dài mà không cần phải can thiệp nhiều. | * Đầu dò cảm biến cần thay thế thường xuyên do sự hao mòn của thiết bị, gây tăng chi phí dài hạn. * Giá thành thiết bị và cảm biến thường khá cao, gây khó khăn cho người sử dụng. * Cảm giác khó chịu khi cảm biến tiếp xúc trực tiếp với mô hoặc da, có thể gây kích ứng hoặc khó chịu với một số người. |
| Không xâm lấn | * Không gây đau đớn, mang lại sự thoải mái cho người sử dụng. * Không cần lấy máu, giúp giảm nguy cơ nhiễm trùng và khó chịu. * Tiện lợi và dễ sử dụng, phù hợp cho người muốn theo dõi đường huyết mà không cần can thiệp xâm lấn. | * Giá thành của các thiết bị đo không xâm lấn thường cao, khiến việc tiếp cận khó khăn đối với một số đối tượng bệnh nhân. * Độ chính xác của các phương pháp đo không xâm lấn vẫn chưa đạt đến mức độ cao như phương pháp xâm lấn, điều này có thể ảnh hưởng đến việc đưa ra kết luận chính xác về tình trạng đường huyết. * Các công nghệ hiện tại vẫn còn hạn chế về độ chính xác và ổn định so với các phương pháp khác. |

Trong những năm gần đây, phương pháp không xâm lấn đã được đưa vào nghiên cứu và phát triển trong việc dự đoán giá trị nồng độ glucose. Phương pháp này cung cấp một giải pháp chính xác, không gây đau đớn và quy trình đo dễ dàng. Đã có nhiều đề xuất được giới thiệu trong việc dự đoán nồng độ đường huyết, nổi bật trong số đó là việc đo đường huyết bằng phương pháp quang học đã được chứng minh là chính xác và tin cậy trong các nghiên cứu. Theo [1], bảng … so sánh những ưu và nhược điểm của các kỹ thuật quang phổ.

Bảng 1: So sánh ưu và nhược điểm của các kỹ thuật quang phổ

|  |  |  |
| --- | --- | --- |
| Kỹ thuật | Ưu điểm | Nhược điểm |
| Cận hồng ngoại | - Cường độ tín hiệu tỷ lệ với phân tử glucose.  - Có thể hoạt động với nhiều bề mặt. | - Tín hiệu glucose yêu cầu mô hình học máy phức tạp để giải thích.  - Mức độ tán xạ cao. |
| Quang phổ Raman | - Ít nhạy cảm với nhiệt độ và nước.  - Độ đặc hiệu cao. | - Dễ bị nhiễu, dẫn đến tỷ lệ tín hiệu/nhiễu (SNR) thấp.  - Chi phí đầu tư cao  - Yêu cầu nguồn laser mạnh, có thể gây nguy hiểm cho tế bào. |
| Huỳnh quang | - Độ nhạy cao với phân tử glucose do không bị tán xạ ánh sáng.  - Độ nhạy tốt nhờ tính chất quang học đặc biệt. | - Rất nhạy cảm với pH và/hoặc oxy cục bộ.  - Dễ bị nhiễu với các vật thể. |
| Cảm biến điện từ | - Chi phí thấp, kích thước nhỏ.  - Ít bị ion hóa. | - Độ chọn lọc kém do hằng số điện môi bị ảnh hưởng bởi các thành phần máu khác.  - Rất nhạy cảm với thay đổi nhỏ về nhiệt độ. |

Dựa trên các yếu tố như cường độ tín hiệu tỷ lệ thuận với nồng độ glucose, khả năng ứng dụng cao trên nhiều loại bề mặt, chi phí thấp, và đặc biệt là khả năng sử dụng mô hình học máy để giải thích giá trị các bước sóng trong dải quang phổ NIR nhằm dự đoán nồng độ glucose, đề tài đã lựa chọn tập trung vào khai thác và phát triển dựa trên quang phổ NIR.

## 2.2. Các nghiên cứu về phương pháp đo Glucose không xâm lấn sử dụng NIR

### 2.2.1. Tình hình nghiên cứu trong nước

Tại Việt Nam, cụ thể ở UIT, đã có hai khóa luận tốt nghiệp về phương pháp đo nồng độ glucose không xâm lấn sử dụng quang phổ VIS-NIR. Nghiên cứu tập trung vào vùng bước sóng từ 410nm đến 940nm, xác định tác động của chúng đến phân tử glucose và áp dụng các mô hình hồi quy máy học để dự đoán nồng độ glucose. Tuy nhiên, vẫn còn tồn tại những hạn chế, cụ thể như sau:

* **Khóa luận tốt nghiệp năm 2022:** Đề tài đã khảo sát vùng bước sóng từ 415nm đến 910nm, sử dụng hai cảm biến AS7341 và AS7263, qua đó xác định các bước sóng 760nm, 860nm, 810nm, và 730nm có ảnh hưởng đáng kể đến nồng độ glucose. Nhóm nghiên cứu đã áp dụng mô hình hồi quy tuyến tính đa biến (Multiple Linear Regression) để dự đoán, đạt độ chính xác 79,56% với hệ số R². Khi kiểm thử trên 10 tình nguyện viên, sai số trung bình đạt 7,41%. Tuy nhiên, hạn chế chính của nghiên cứu là mô hình máy học chưa được đa dạng và phần cứng thiết bị chưa được tối ưu.
* **Khóa luận tốt nghiệp năm 2023:** Nghiên cứu áp dụng nhiều phương pháp học máy, bao gồm Multiple Linear Regression, Support Vector Machine, Decision Tree, Random Forest, và Gradient Boosting. Kết quả cho thấy Gradient Boosting và Random Forest có hiệu suất tốt nhất với R² lần lượt là 0.4945 và 0.2813. Các bước sóng quan trọng được xác định gồm 460nm, 680nm, 730nm, 810nm, 860nm, 900nm, và 940nm. Hệ thống cũng tích hợp công nghệ IoMT để hỗ trợ theo dõi glucose từ xa. Tuy nhiên, hai mô hình máy học Multiple Linear Regression và Support Vector Machine cho kết quả thấp, với R² lần lượt là -0.7 và -0.24, không phù hợp để giải thích các đặc điểm phi tuyến của dữ liệu chứng tỏ trong tập dữ liệu xuất hiện các nhiều nhiễu ảnh hưởng trực tiếp đến kết quả dự đoán.

**2.2.2. Tình hình nghiên cứu trên thế giới**

Nghiên cứu [6] giới thiệu một cảm biến sinh học đeo dựa trên quang phổ VIS – NIR để giám sát nồng độ đường huyết không xâm lấn. Hệ thống này tập trung vào bốn bước sóng 950nm, 850nm, 660nm, 530nm và đạt hiệu suất cao với SEP < 6.16 mg/dl và Rp > 0.86.

Nghiên cứu [7] thực hiện đo ở 18 bước sóng từ 410nm đến 940nm, cho thấy mối tương quan cao (0.98) giữa nồng độ glucose và cường độ truyền qua tại 4 bước sóng (485nm, 645nm, 860nm, và 940nm). Các phương pháp máy học được áp dụng cho dự đoán glucose, trong đó các phương pháp hồi quy có tỷ lệ sai số 9%, trong khi phân tích lưới lỗi Clarke chỉ ra rằng 99.75% kết quả đo glucose nằm trong vùng chấp nhận lâm sàng. Dù nghiên cứu chỉ thực hiện trong môi trường nước pha glucose, nhưng đã chứng minh được sự ảnh hưởng của giá trị glucose đến các giá trị bước sóng.

Theo nghiên cứu [8], hệ thống đo nồng độ glucose không xâm lấn dựa trên kỹ thuật quang phổ NIR đã được triển khai thành công với mô hình tính toán xử lý dữ liệu hiệu quả, đạt hệ số tương quan 0.908.

Nghiên cứu [9], giới thiệu phương pháp phát hiện glucose bằng cách nhận ánh sáng hồng ngoại phản xạ sau khi glucose bị hấp thụ, với bước sóng cố định 940nm. Đèn LED hồng ngoại phát ánh sáng bước sóng này chiếu vào da tình nguyện viên, và cảm biến thu lại giá trị bước sóng phản xạ. Bốn thuật toán máy học (PCR, PLS, SVR, RFR) được sử dụng để xây dựng mô hình, và RFR có hiệu suất tốt nhất được chọn làm mô hình dự đoán glucose cuối cùng.

Dựa trên những nghiên cứu trước, đề tài tập trung vào dải bước sóng từ 410 *nm* đến 940 *nm* để kiểm tra khả năng xuyên qua ngón tay của ánh sáng VIS-NIR và xác định các bước sóng có ảnh hưởng đáng kể đến nồng độ glucose trong máu. Dữ liệu thu thập từ cảm biến, sau khi ánh sáng truyền qua ngón tay, sẽ được xử lý và phân tích bằng các mô hình học máy để dự đoán nồng độ glucose. Để đánh giá độ chính xác và khả năng ứng dụng trong lâm sàng của các giá trị dự đoán, đề tài sẽ sử dụng lưới lỗi Clarke (Clarke Error Grid) như một tiêu chuẩn đánh giá bên cạnh các phương pháp đánh giá mô hình học máy.

## 2.3. Lý thuyết về phương pháp đo glucose không xâm lấn VIS - NIR

### 2.3.1. Quang phổ VIS – NIR

Visible – Near Infrared (VIS – NIR) là một dãi quang phổ kết hợp giữa ánh sáng mà mắt người có thể nhìn thấy, với bước sóng từ 380 *nm* đến 760 *nm* và ánh sáng cận hồng ngoại nằm trong dải bước sóng từ 780 *nm* đến 2500 *nm* [6].

Visible – Near Infrared (VIS-NIR) là một dải quang phổ kết hợp giữa ánh sáng khả kiến (visible light) và ánh sáng cận hồng ngoại (near infrared).

* **Ánh sáng khả kiến (VIS)**: Có bước sóng từ 380 *nm* đến 720 *nm*, nằm trong vùng ánh sáng mà mắt người có thể nhìn thấy. Phổ khả kiến được sử dụng để xác định nồng độ dung dịch dựa trên tương tác giữa ánh sáng và các thành phần hóa học trong mẫu.
* **Ánh sáng cận hồng ngoại (NIR)**: Có bước sóng từ 700 nm đến 2500 nm, được dùng để phân tích dao động và chuyển động quay của các liên kết hóa học trong phân tử.

Phổ NIR tập trung vào các vùng hấp thụ liên quan đến dao động và bội số của các nhóm chức C-H, N-H, O-H, và S-H. Những đặc tính này giúp VIS-NIR được sử dụng phổ biến trong việc phân tích nồng độ các hợp chất hóa học cũng như thông tin cấu trúc phân tử. [6]

### 2.3.2. Ảnh hưởng quang phổ VIS – NIR đến nồng độ glucose

Theo [10], khi thực hiện thử nghiệm lựa chọn các bước sóng, Glucose có các đỉnh hấp thụ ánh sáng ở các bước sóng 940 *nm*, 970 *nm*, 1197 *nm*, 1408 *nm*, 1536 *nm*, 1688 *nm*, 1925 *nm*, 2100 *nm*, 2261 *nm* và 2326 *nm*. Tuy nhiên, ở bước sóng 940 *nm*, sự suy giảm của tín hiệu quang học do các thành phần khác trong máu như nước, tiểu cầu, hồng cầu là ít nhất, do đó có thể đạt được độ sâu thâm nhập mong muốn và nồng độ glucose thực tế có thể được dự đoán. Cũng theo [10] quang phổ NIR cho phép tín hiệu thâm nhập vào mô ở độ sâu trong khoảng từ 1 đến 100 *mm*. Độ sâu thâm nhập giảm dần khi giá trị bước sóng tín hiệu tăng lên.

Dựa vào các nghiên cứu [7] [6] [1], ta nhận thấy rằng các bước sóng trong vùng VIS – NIR có ảnh hưởng đến sự thay đổi nồng độ glucose trong máu. Theo [8], quá trình phản xạ và hấp thụ ánh sáng ở các bước sóng cụ thể có thể gây ra dao động trong các phân tử glucose. Một trong những phát hiện quan trọng là dao động của liên kết CH xuất hiện gần bước sóng 920 nm, và dao động này ít bị ảnh hưởng bởi sự dao động của nhóm -OH trong phân tử nước. Đây là một yếu tố quan trọng khi đo glucose trong cơ thể người, vì cơ thể con người chứa một lượng nước lớn, có thể làm ảnh hưởng đến kết quả đo nếu không lựa chọn bước sóng phù hợp.

Ngoài ra, nghiên cứu cũng chỉ ra rằng các bước sóng đặc trưng cho các dao động của các liên kết CH2 (930 nm), CH (960 nm), và OH (984 nm) trong phân tử glucose có thể giúp phát hiện glucose một cách chính xác, vì chúng phản ánh các chuyển động dao động đặc trưng trong cấu trúc phân tử glucose [8]. Đặc biệt, bước sóng 940 nm đã được chứng minh là không bị ảnh hưởng bởi sự thay đổi cấu trúc ngón tay, bao gồm cả độ rộng của ngón tay, giúp tăng độ chính xác khi đo nồng độ glucose không xâm lấn [8].

### 2.3.4. Độ hấp thụ của ánh sáng trong môi trường đo

Khi tia sáng tương tác với mô cơ thể người, cường độ của nó bị suy giảm do hiện tượng tán xạ và hấp thụ bởi các thành phần bên trong mô. Định lý Beer-Lambert đóng vai trò quan trọng trong việc đo lường sự hấp thụ này [1], định lý chỉ ra rằng mức độ hấp thụ ánh sáng tỷ lệ thuận với nồng độ của chất trong môi trường và chiều dài quang học mà ánh sáng đi qua.

Trong nghiên cứu này, nồng độ glucose trong mô ảnh hưởng đáng kể đến các đặc tính quang học của mô. Mô có nồng độ glucose thấp làm tăng mức độ tán xạ ánh sáng, dẫn đến chiều dài quang học lớn hơn và giảm sự hấp thụ ánh sáng. Ngược lại, ở mô có nồng độ glucose cao, mức độ tán xạ giảm, chiều dài quang học ngắn hơn, và mức độ hấp thụ ánh sáng tăng lên. Do đó, cường độ ánh sáng truyền qua mô chứa nhiều glucose sẽ giảm so với mô có ít glucose, vì ánh sáng bị hấp thụ nhiều hơn.

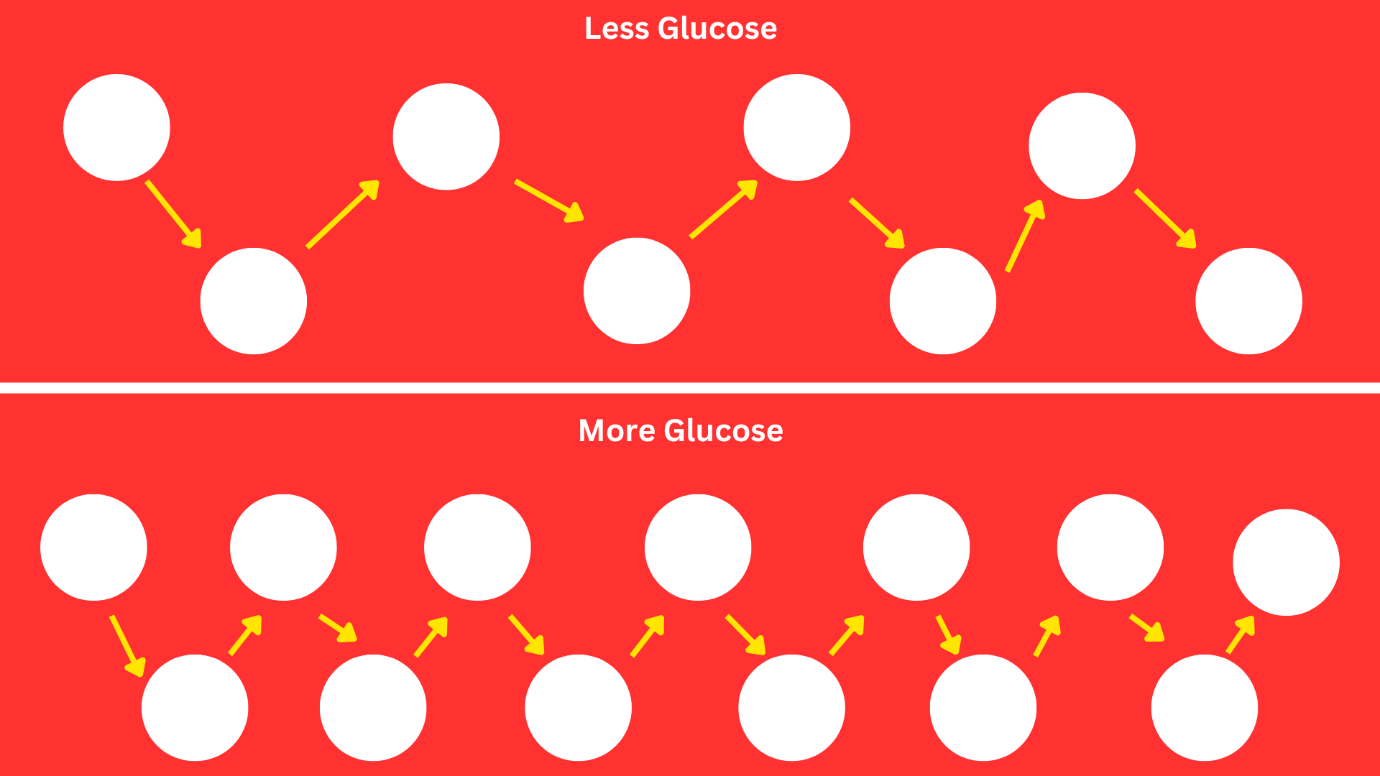
Phương pháp phân tích định lượng glucose dựa trên sự hấp thụ ánh sáng tuyến tính. Định lý Beer-Lambert được áp dụng để tính toán nồng độ hấp thụ của các phân tử trong mô, đặc biệt là khi ánh sáng truyền qua ngón tay. Trong nghiên cứu này, ánh sáng được sử dụng để đánh giá mức độ hấp thụ và mối tương quan với nồng độ glucose. Theo định lý Beer-Lambert, sự suy giảm ánh sáng tỷ lệ thuận với độ dày của mô và nồng độ của các thành phần bên trong nó. Định lý Beer – Lambert’s được thể hiện ở phương trình 2.1:

*(2.1)*

Với *A* (Absorbance) – đo lường ánh sáng bị hấp thụ bởi mẫu. *I0* biểu diễn cường độ ánh ban đầu trước khi được chiếu vào mẫu. *I* thể hiện cường độ ánh sáng sau khi đã truyền qua mẫu. là hệ số hấp thụ mol – đây là hệ số đặc trưng cho khả năng hấp thụ ánh sáng của một chất được gọi là hệ số hấp thụ mol với đơn vị của là *Lmol-1cm-1*. *C* là nồng độ mol được kí hiệu là *mol/L*. *d* là chiều dài đường truyền của ánh sáng.

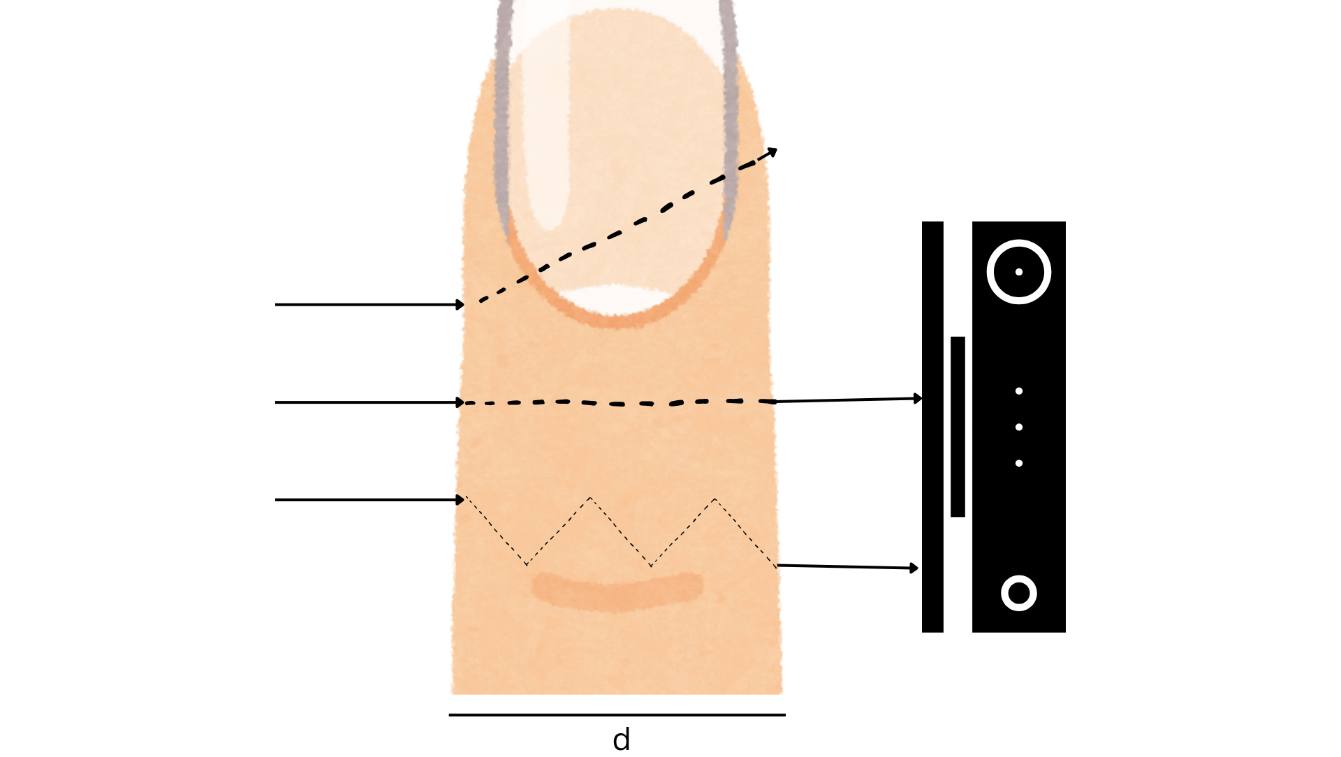
### 2.3.5. Ảnh hưởng của ánh sáng

Theo công thức 2.1, mức độ ánh sáng bị hấp thụ bởi mẫu tỷ lệ thuận với CCC - nồng độ mol. Điều này có nghĩa là khi nồng độ glucose tăng, giá trị ánh sáng hấp thụ cũng tăng lên, và ngược lại. Như thể hiện trong hình 2.2, khi nồng độ glucose trong máu cao, ánh sáng sẽ bị hấp thụ hoặc tán xạ nhiều hơn so với khi nồng độ glucose thấp.



Hình 2. 2:Ảnh hưởng nồng độ glucose đến ánh sáng

Theo bài [11], định luật Beer-Lambert giả định rằng bức xạ ánh sáng tới là đơn sắc và mẫu không gây ra hiện tượng tán xạ. Tuy nhiên, trong thực tế đo lường glucose trong máu, sai số xảy ra chủ yếu do đặc tính tán xạ của mô. Khoảng 80% sự suy giảm ánh sáng cận hồng ngoại trong mô là do tán xạ, trong khi 20% còn lại là do hấp thụ.



Hình 2. 3: Đường đi ánh sáng

Như minh họa trong hình 2.3, việc ánh sáng có được cảm biến thu nhận hay không phụ thuộc vào nhiều yếu tố như vị trí đặt tay, góc đặt tay so với bề mặt cảm biến, sắc tố melanin của mỗi người, độ dày ngón tay. Những yếu tố này làm cho quá trình đo lường phức tạp, đòi hỏi sự kết hợp giữa phân tích bước sóng và các mô hình học máy để tìm ra mối tương quan giữa nồng độ glucose và tín hiệu NIR. Việc áp dụng các phương pháp này là cần thiết để cải thiện độ chính xác của phép đo glucose không xâm lấn.

### 2.3.6. Lựa chọn khoảng cách đo

Dựa theo [12], Định luật Beer – Lambert cũng có thể được biểu diễn dưới dạng:

*(2.2)*

Công thức 2.2 liên quan đến cường độ ánh sáng *I* sau khi truyền qua khoảng cách *d*, với ​ là cường độ ánh sáng ban đầu. Trong đó, là hệ số suy giảm quang học của môi trường, được xác định bởi:

* Hệ số hấp thụ: biểu thị khả năng môi trường hấp thụ năng lượng ánh sáng.
* Hệ số tán xạ: mô tả mức độ ánh sáng bị lệch hướng khi truyền qua môi trường.

Khoảng cách *d* là quãng đường mà ánh sáng phải đi qua trong môi trường đo. Dựa trên công thức này, việc tối ưu hóa giá trị *d* là rất quan trọng với mục tiêu tín hiệu ánh sáng sau khi truyền qua môi trường đo sẽ rõ ràng và biểu thị các đặc trưng rõ ràng hơn.

Khoảng cách đo được thiết lập 1cm được biểu diễn trong **hình 5,** nhằm đảm bảo các yếu tố an toàn, hiệu quả và độ bền trong quá trình đo lường nồng độ glucose. Thứ nhất, khoảng cách này giúp đảm bảo an toàn cho người sử dụng, tránh nguy cơ gây bỏng hoặc khó chịu khi ánh sáng từ nguồn halogen có nhiệt độ cao chiếu trực tiếp vào ngón tay. Thứ hai, việc tối ưu hóa khoảng cách giúp cường độ ánh sáng truyền qua ngón tay đạt mức đủ lớn, từ đó cải thiện độ chính xác trong việc phân tích mối tương quan giữa bước sóng và nồng độ glucose. Cuối cùng, khoảng cách được thiết kế hợp lý để hạn chế khả năng ngón tay chạm vào các linh kiện điện tử trên cảm biến, giúp giảm thiểu nguy cơ hư hại cho thiết bị đo lường.



Hình 2. 4: Khoảng cách được bố trí của cảm biến và đèn.

## 2.4. Nguồn sáng LEDS

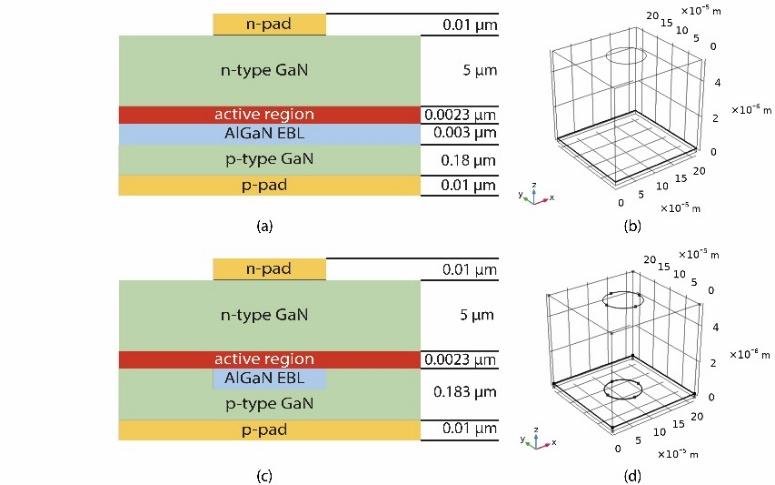
Điốt phát sáng (LEDs) đã thu hút được nhiều sự quan tâm cho mục đích áp dụng cho công nghệ chiếu sáng và hiển thị ở trạng thái rắn. LEDs có thể sử dụng để làm nguồn sáng tại các bước sóng cụ thể trong đề tài này. Tuy nhiên, hiệu suất phát sáng suy giảm ở đèn LEDs do giảm hiệu suất lượng tử bên trong (IQE) xảy ra ở mức tương đối cao đặc biệt khi dòng cung cấp dòng. Đây vẫn là một thách thức cần vượt qua và vẫn chưa có giải pháp cụ thể. Sự giảm hiệu suất lượng tử bên trong không phải do một cơ chế đơn lẻ nào gây ra mà liên quan đến nhiều yếu tố khác nhau bao gồm sự tập trung cục bộ, hiệu ứng Stark giới hạn lượng tử (QCSE), bão hòa kết hợp lại phát quang, sự kết hợp lại của dòng Auger, rò rỉ dòng trực tiếp ra khỏi các giếng lượng tử và rò rỉ các hạt điện tích. Sự suy giảm hiệu suất chủ yếu là do sự xuất hiện của sự kết hợp lại không phát quang và rò rỉ điện tích bên trong và bên ngoài vùng hoạt động hay các giếng lượng tử. Sự kết hợp lại của dòng Auger là một quá trình kết hợp lại không phát quang và chỉ sinh ra nhiệt của các dòng mang điện tích, và nó đóng vai trò chính trong việc giảm hiệu suất của đèn LEDs khi mật độ dòng cung cấp cao. Các hạt điện tích được tạo ra bên ngoài hoặc có thể được kết hợp lại bên trong các giếng lượng tử (QWs). Hiệu ứng hạt tải nóng, còn được gọi là rò rỉ điện tích trực tiếp, được coi là một trong những quá trình kết hợp lại không phát quang. Ở đây, các electron không tái kết hợp hoàn toàn trong các giếng lượng tử sẽ tràn từ vùng hoạt động (các giếng lượng tử) sang các lớp bán dẫn p. Quá trình này được gọi là sự rò rỉ điện tích do phát xạ nhiệt thay vì phát quang. Sự rò rỉ điện tích cũng là một trong những cơ chế gây tranh cãi nhất và hiện nay vẫn chưa có kết luận cụ thể. Sự tập trung cục bộ của dòng điện dẫn đến nồng độ hạt tải điện cao hơn trong các giếng lượng tử dưới vùng n-pad, sau đó gây ra sự gia tăng dòng Auger và rò rỉ electron trong diode phát quang, hiện tượng này xảy ra do tăng cường mật độ điện tích tại các khu vực cục bộ bên dưới các lớp điện cực. Đây cũng là một yếu tố quan trọng làm giảm hiệu quả phát quang của diode phát quang. Lớp chặn điện tích electron (EBL) là một lớp mỏng vật liệu có vùng cấm rộng được chèn vào giữa vùng hoạt động và các lớp loại p. Nó tạo ra một rào cản đối với dòng electron và dải hóa trị tại các bề mặt vùng hoạt động/EBL và EBL/loại p. Rào cản trong dải dẫn điện ở vùng hoạt động/giao diện EBL có khả năng làm giảm sự rò rỉ điện tích ra khỏi vùng hoạt động. Một số loại EBL đã được phát triển để cải thiện độ sụt giảm hiệu suất trong diode phát quang ví dụ: EBL AlGaN, EBL InAlN, EBL siêu mạng InAlN/GaN (SL) và EBL AlGaN/GaN SL. Tuy nhiên, rào cản trong dải hóa trị ở vùng hoạt động/giao diện EBL cũng chặn dòng lỗ trống đến vùng hoạt động, điều này có thể dẫn đến hiệu suất cung cấp lỗ trống thấp. Do sự phân cực nguyên nhân do sự không khớp giữa các vật liệu ở các lớp khác nhau, một tấm điện tích dương tồn tại ở bề mặt tiếp xúc giữa EBL và vùng hoạt động, kéo dải dẫn của EBL xuống. Điều này dẫn đến việc giảm chiều cao rào cản ở vùng hoạt động/giao diện EBL đối với các electron ở dải dẫn, trong khi chiều cao rào cản đối với các lỗ trống ở dải hóa trị tăng lên. Việc giảm chiều cao rào cản ở dải dẫn làm giảm hiệu ứng chặn của EBL đối với các electron và việc tăng chiều cao rào cản ở dải hóa trị gây ra hiệu suất cung cấp lỗ trống thấp vào lớp hoạt động, điều này có liên quan đến sự xuất hiện của các electron bị rò rỉ. Hiệu ứng chặn của AlxGa1-xN EBL đối với các electron phụ thuộc vào thành phần Al (xAl). Sự gia tăng thành phần Al của AlxGa1-xN dẫn đến dải tần rộng hơn, dẫn đến chiều cao rào cản cao hơn ở giao diện vùng hoạt động/EBL, do đó làm giảm rò rỉ điện tử dẫn đến tăng hiệu suất của đèn LED. Tuy nhiên, việc chèn AlGaN EBL vào diode phát quang sẽ dẫn đến tăng điện trở suất bên trong, đặc biệt đối với các thành phần Al cao hơn. Hiệu suất của đèn LED giảm khi thành phần Al trong EBL p-AlGaN tăng.

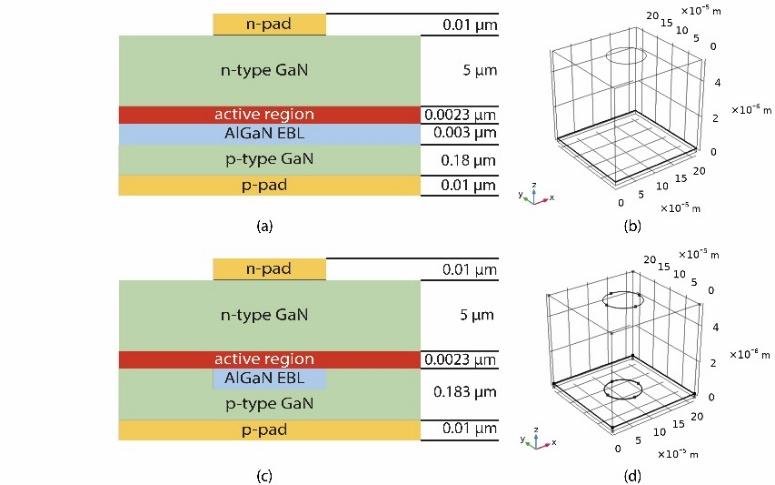
Sự phân cực bao gồm phân cực áp điện và phân cực tự phát, đây là một trong những nguyên nhân gây ra sự kết hợp lại không phát quang cao từ các giếng đa lượng tử (MQW) . Điều đó dẫn đến việc giảm hiệu quả sử dụng của diode phát quang. Các đặc tính phân cực của hợp kim nitrit với mô hình thành phần được sử dụng trong nghiên cứu này để nghiên cứu ảnh hưởng của sự phân cực đến các đặc tính của micro-LED. Để nghiên cứu cơ chế vật lý về cách hiệu ứng phân cực tại các giao diện bên trong của thiết bị ảnh hưởng đến đặc tính của micro-LED, mô phỏng số đã được thực hiện bằng trình mô phỏng COMSOL. Mô hình số ba chiều dựa trên phương trình khuếch tán trôi sóng mang, được sử dụng để thu được các đặc tính điện trong đèn LED siêu nhỏ. Hiệu suất lượng tử bên trong (IQE) được xác định bằng tỷ số giữa dòng bức xạ và dòng phun. Đường cong IQE và J-V của micro-LED với lớp chặn electron AlGaN tiêu chuẩn (EBL) được so sánh với kết quả thực nghiệm.

Hiệu ứng dòng điện tập trung cục bộ là một vấn đề phổ biến trong các cấu trúc diode phát quang, cả cấu trúc có và không có lớp chắn electron EBL, gây ra sự phân bố mật độ dòng điện không đồng đều trong cấu trúc diode phát quang. Việc tăng cường mật độ hạt điện tích cục bộ dẫn đến sự gia tăng việc kết hợp lại không phát quang rò rỉ điện tích. Sự tập trung dòng điện trong vùng dưới n-pad của lớp hoạt động trong cấu trúc diode phát quang gây ra sự rò rỉ dòng điện tích ở vùng này tăng nhanh hơn so với các vùng khác lân cận; hơn nữa, hầu hết từ khu vực này đều được hấp thụ bởi n-pad. Sự phân bố mật độ dòng điện đồng đều hơn là chìa khóa để đạt được diode phát quang hiệu suất cao. Trong đèn cấu trúc diode phát quang, lớp chắn electron EBL thường bao phủ toàn bộ lớp p-GaN, mặc dù chức năng chống lại sự tập trung dòng điện nghiêm trọng của nó chỉ áp dụng được ở vùng dưới n-pad của vùng hoạt động. Nó không hữu ích trong việc chặn sự rò rỉ điện tử ra ngoài vùng này mặc dù nó gây ra điện trở trong cao hơn và hiệu suất phun lỗ thấp hơn. Một thiết kế tốt hơn sẽ cho phép dòng điện lan truyền dễ dàng hơn từ n-pad đến vùng không nằm trong lớp chắn EBL. Điện trở nội tại nhỏ hơn sẽ dẫn đến sự phân bố mật độ dòng điện đồng đều hơn. Lớp chắn electron AlGaN EBL được thiết kế đặc biệt này sẽ cải thiện cả hiệu suất và độ suy giảm hiệu suất trong đèn LED SWQ. Đèn LED có cấu trúc EBL hình tròn này có khả năng hoạt động ở điện áp thấp hơn so với đèn LED có thiết kế EBL truyền thống.

### Mô hình vật lý

Hình 1 (a) và (b) thể hiện cấu trúc của thiết kế đèn LED InGaN/GaN SQW thẳng đứng với EBL Al0.1Ga0.9N dày 0,003-µm được kẹp giữa vùng hoạt động dày 0,0023-µm và vùng hoạt động 0,18-µm- Lớp GaN loại p dày, diện tích 212 × 212 µm2 (cấu trúc 1 (S1)). Lớp p-pad, n-pad và n-GaN lần lượt dày 0,01 µm, 0,01 µm và 5 µm. Vùng hoạt động bao gồm hai lớp rào cản GaN với một lớp hoạt động InGaN. Lớp hoạt động In0.19Ga0.81N và lớp rào cản GaN lần lượt dày 3 nm và 10 nm. Một n-pad hình tròn bán kính 40 µm được hình thành trên bề mặt n-GaN trên cùng [20]. Như có thể thấy trong Hình 1(c) và (d), hình dạng phẳng của EBL trong Hình 1(a) được thay đổi thành hình tròn Al0.1Ga0.9N EBL (cấu trúc 2 (S2)), chỉ trong vùng dưới n-pad. Đường kính giống như đường kính của lớp n-pad. Cấu hình pha tạp của lớp n-GaN, EBL và p-GaN lần lượt là 2 × 1018 cm-3, 6 × 1017 cm-3 và 2 × 1018 cm-3. Các lớp vùng hoạt động không được pha tạp. Độ dày của p-GaN là 0,183 µm. Cấu hình pha tạp của p-GaN, EBL và n-GaN lần lượt là 2×1018 cm-3, 2×1018 cm-3 và 6 × 1017 cm-3. Trong nghiên cứu này, thành phần EBL Al trong S1 và S2 thay đổi từ 0,05 đến 0,14.

****

****

Hình 1. Sơ đồ của đèn LED SWQ dọc với EBL truyền thống và EBL được thiết kế đặc biệt.

### Mô hình toán

Các phương trình Poisson, khuếch tán trôi và liên tục [9, 10] được trình bày bên dưới:

trong đó *ε* là độ thấm của vật liệu điện môi; *V* là điện thế; *n* là nồng độ hạt electron và *p* là nồng độ lỗ trống; đại diện cho nồng độ chất nhận ion hóa và đại diện cho nồng độ chất cho bị ion hóa; *Jn* và *Jp* lần lượt là mật độ dòng điện mang của electron và lỗ trống; và *µn* và *µp* lần lượt là độ linh động của electron và lỗ trống; *Rsrh* là dòng tái kết hợp Shockley-Read-Hall (SRH); *Rrad* là dòng kết hợp lại phát quang; *RAuger* là kết hợp lại Auger; *Rlk* là tổng tốc độ rò rỉ hạt tải điện, bao gồm rò rỉ hạt tải điện trực tiếp và điện tích. Trong đề tài này, tính di động của sóng mang được tính bằng mô hình di động Arora. Sự rò rỉ lỗ trống do phát xạ nhiệt bị bỏ qua do độ linh động của lỗ thấp hơn so với độ linh động của electron trong p-GaN pha tạp cao. Tốc độ kết hợp lại của cả electron và lỗ trống được giả định là như nhau trong đề tài này. Giá trị tỷ lệ bù băng tần giữa GaN và AlxGa*1-xN* là 0,5, giống như trong các nghiên cứu trước đây. Sự biến dạng của QW là do sự phân cực. Phân cực tổng bao gồm cả phân cực tự phát và phân cực điện áp. Mô hình được phát triển bởi Fiorentini *et al.* được sử dụng để bao gồm hiệu ứng phân cực trong các giao diện diode phát quang bên trong, đây là một hàm của thành phần và hằng số mạng của vật liệu. Hằng số mạng hợp kim được tính theo định luật Vegard. Sự phân cực dẫn đến điện tích bề mặt tại các bề mặt tiếp xúc giữa các vật liệu khác nhau, được định nghĩa như sau:

điện tích bề mặt; và lần lượt là các độ phân cực tự phát và điện áp. Tuy nhiên, giá trị điện tích bề mặt tìm được trong các thí nghiệm trong [24] nhỏ hơn từ 0,2 đến 0,8 so với giá trị tính theo phương trình (5). Trong các nghiên cứu lý thuyết và mô phỏng, điện tích bề mặt được tính theo phương trình (5) thường được giảm xuống theo hệ số phù hợp. Hệ số phù hợp được chọn là 0,4, giống như trong nghiên cứu trước đây của chúng tôi [24]. Điện tích bề mặt chỉ xuất hiện ở các bề mặt tiếp xúc giữa lớp rào cản QW/GaN, lớp rào cản GaN/EBL và lớp EBL/p-GaN trong nghiên cứu này. Tại các bề mặt tiếp xúc giữa các vật liệu khác nhau, điều kiện được áp dụng cho độ dịch chuyển của điện trường D, trong đó chỉ số dưới biểu thị lớp.

Tỷ lệ RSRH tái tổ hợp Shockley-Read-Hall (SRH) được đưa ra bởi:

trong đó *kB* là hằng số Boltzmann; *T* là nhiệt độ của mạng; *Et* là năng lượng bẫy; và *Ei* và *ni* là mức Fermi và mật độ sóng mang trong chất bán dẫn nội tại. là hệ số tái hợp SRH của electron và là hệ số tái hợp SRH của electron đối với lỗ trống. Tuổi thọ của chất mang SRH bao gồm hiệu ứng PSF, được phát triển trong nghiên cứu trước đây của chúng tôi [10], được đưa ra bởi

trong đó *n0* đại diện cho hệ số hiệu ứng PSF. Mật độ electron nền được chọn *N0* là 1 × 1017 cm-3 và số mũ không thứ nguyên = 0,4 được chọn để phù hợp với kết quả mô phỏng với IQE được đo bằng thực nghiệm. Hệ số PSF được đặt là 1 × 1019 cm-3 sao cho kết quả mô phỏng phù hợp với IQE đo bằng thực nghiệm [26-28]. Thuật ngữ này được sử dụng để giải quyết sự không chắc chắn phát sinh từ cấu trúc khiếm khuyết của tinh thể và tạp chất vật chất. Tuổi thọ của chất mang thử nghiệm ở mức tiêm rất thấp là từ 1 × 10-6 s đến 5 × 10-9 s. Ở đây, được chọn, như được tìm thấy trong một đèn LED có cùng độ dày và phần mol của lớp giếng, trong khi = 0,15 × 10-7s được chọn để tương ứng với IQE được đo bằng thực nghiệm.

Tốc độ tái hợp bức xạ [34] được tính bằng

trong đó *B* là hệ số bức xạ. Hệ số bức xạ dựa trên mật độ electron và hiệu ứng PSF [10, 35] được cho bởi

trong đó *B0*, bao gồm hệ số hiệu ứng PSF *n0*, là hệ số tái hợp bức xạ ở mức cung cấp cực thấp. *A* và *B* được chọn từ 1 × 10-11 đến 1,5 × 10-9 cm3·s-1; *B0* = 2,21 × 10-10 cm3·s-1 được sử dụng trong nghiên cứu này để phù hợp với đặc tính IQE thực nghiệm.

Tốc độ kết hợp lại Auger [38] được đưa ra bởi

trong đó *Cn* và *Cp* lần lượt là hệ số Auger của lỗ trống và electron. Hệ số Auger [35], trong đó *C0*, bao gồm hệ số hiệu ứng PSF n0, là hệ số ở mức bơm rất thấp. Giá trị của *Cn, p* nằm trong khoảng 1,4 × 10-30 cm6/s - 3,5 × 10-34 cm6/s. Giá trị được chọn của các thông số *Cn0* và *C*p0 trong nghiên cứu này là 8 × 10-31 cm6·s-1 để phù hợp với đặc tính IQE thực nghiệm. Tốc độ hạt tải bị rò rỉ trực tiếp được hỗ trợ bởi sự tái hợp Auger có thể được mô tả tương ứng và [40] đối với các electron và lỗ trống. Hệ số hạt tải điện rò rỉ trực tiếp có thể được mô tả bằng *Clk*, *n* = *l2Cn* và *Clk*, *p* = *l2Cp*. Số mũ không thứ nguyên l được giả sử bằng độ dày của lớp InGaN. Chất mang bị rò rỉ khác là do phát xạ nhiệt, trong đó biểu thị các electron bị rò rỉ và các lỗ trống bị rò rỉ.

IQElocal cục bộ có thể được mô tả bởi

Các dòng tái hợp được tính toán bằng cách tích hợp tốc độ phân phối tái hợp trên toàn bộ lớp hoạt động. IQE gộp có thể được mô tả bằng

trong đó Ir và Itotal lần lượt là tổng dòng tái hợp bức xạ và tổng dòng cung cấp.

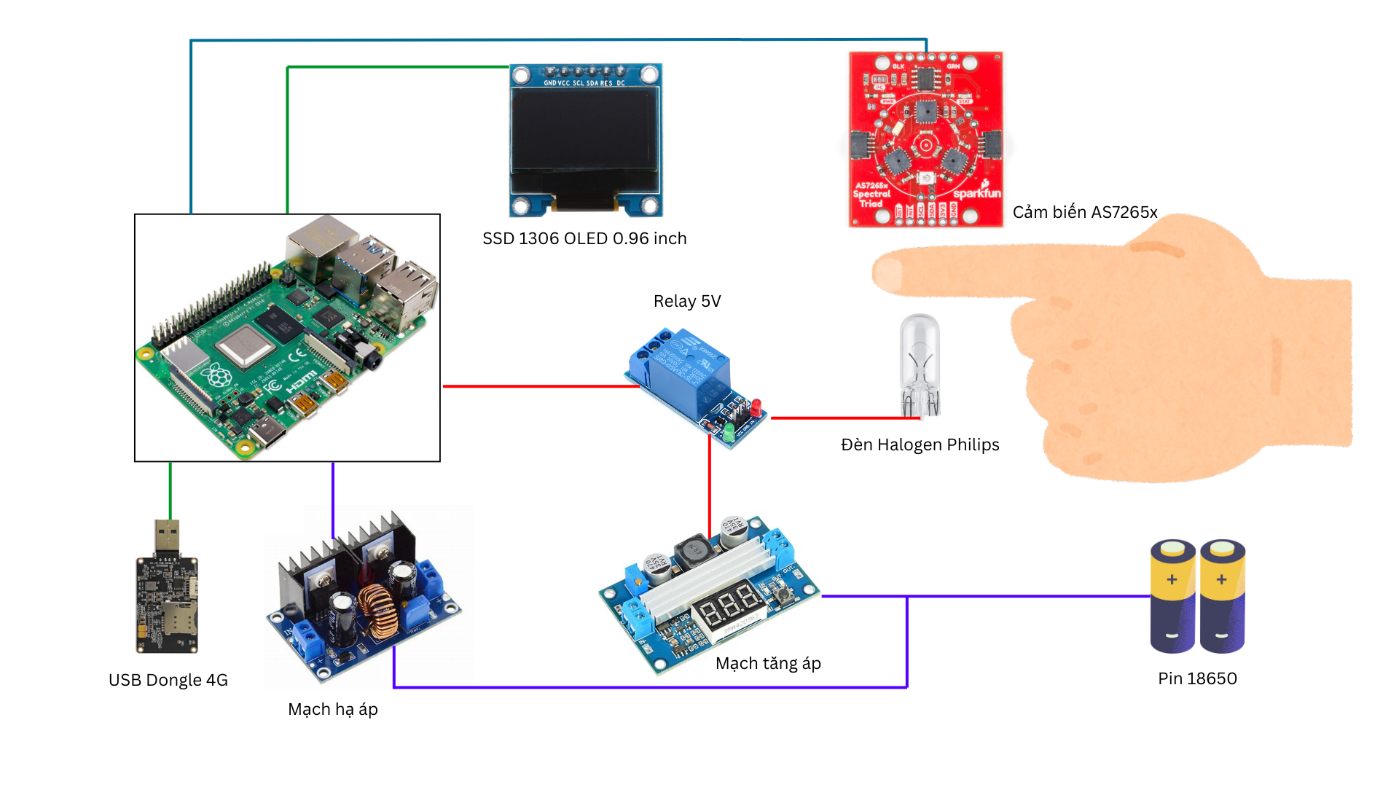
# CHƯƠNG 3: MỤC TIÊU VÀ PHƯƠNG PHÁP

## 3.1. Mục tiêu đặt ra

Mục tiêu của đề tài là phát triển một hệ thống dự đoán nồng độ glucose trong máu dựa trên tác động quang phổ của ánh sáng VIS – NIR lên các phân tử glucose. Đề tài tập trung vào việc nghiên cứu và xác định các bước sóng có ảnh hưởng đáng kể đến sự hấp thụ và tán xạ của glucose trong khoảng từ 410nm đến 940nm thông qua các thí nghiệm thực tế. Ngoài ra, hệ thống được tích hợp công nghệ IoT, cho phép người dùng theo dõi lịch sử đo lường và hiển thị trực quan các biểu đồ nồng độ glucose, hỗ trợ quản lý sức khỏe một cách hiệu quả. Toàn bộ hệ thống được thiết kế với chi phí tối ưu, không vượt quá 5 triệu đồng, đảm bảo tính khả thi và dễ tiếp cận.

## 3.2. Triển khai phần cứng

### 3.2.1. Tổng quan sơ đồ kết nối phần cứng



Hình 3. 1: Sơ đồ kết nối phần cứng

Hình 6 minh họa sơ đồ khối kết nối giữa các thành phần phần cứng trong hệ thống đo đường huyết không xâm lấn. Hệ thống bao gồm các thành phần chính như sau:

* Nguồn sáng Halogen (Philips W5W T10): Đóng vai trò cung cấp phổ sáng rộng từ 400 nm đến 1100 nm, phù hợp để chiếu sáng, đáp ứng yêu cầu thu thập dữ liệu từ các bước sóng liên quan đến vùng VIS - NIR.
* Cảm biến quang phổ AS7265x SparkFun: Có khả năng đo 18 bước sóng ánh sáng riêng lẻ trong khoảng từ 410 nm đến 940 nm, hỗ trợ phân tích phổ chính xác. Cảm biến giao tiếp với vi điều khiển qua giao thức I2C hoặc UART.
* Vi điều khiển Raspberry Pi 4B model B: Được chọn nhờ cấu hình mạnh mẽ, đáp ứng tốt yêu cầu chạy các mô hình máy học trong đề tài.
* Cảm biến điện dung TTP223B Mini: Được tích hợp để phát hiện ngón tay đặt vào, giúp tự động bật/tắt nguồn sáng, tối ưu hóa năng lượng, đảm bảo ngón tay đặt đúng vị trí trước khi thực hiện đo.
* USB Dongle 4G: Được sử dụng để khắc phục hạn chế về kết nối Internet thông qua WiFi của Raspberry Pi 4B, đảm bảo tín hiệu khi tích hợp công nghệ IoT, giúp hệ thống truyền tải dữ liệu đo lường hiệu quả.

### 3.2.2. Chi tiết các linh kiện phần cứng

### 3.2.2.1. Máy tính nhúng

Raspberry Pi 4 Model B đóng vai trò trung tâm trong hệ thống, chịu trách nhiệm xử lý dữ liệu từ cảm biến, điều khiển kích hoạt nguồn sáng và hiển thị kết quả đo trên màn hình OLED.

Sau khi nhận dữ liệu từ cảm biến quang phổ AS7265x, Raspberry Pi thực hiện phân tích các giá trị bước sóng và áp dụng mô hình máy học đã được huấn luyện để dự đoán nồng độ glucose trong máu. Với vi xử lý ARM Cortex-A72 và dung lượng RAM lớn, Raspberry Pi 4 Model B có đủ khả năng triển khai các mô hình học máy phức tạp, đảm bảo tính chính xác và hiệu quả trong quá trình xử lý và dự đoán dữ liệu.

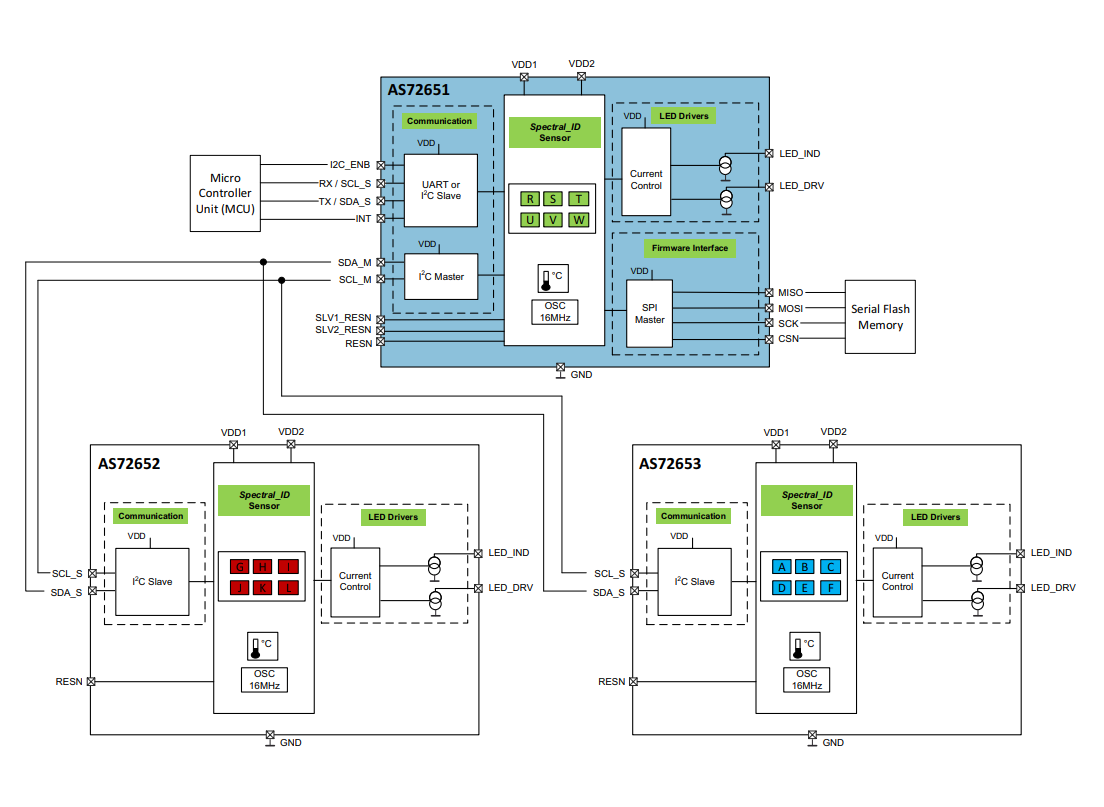
Thông số kỹ thuật của Raspberry Pi 4 Model B:

* Vi xử lý: Broadcom BCM 2711, Quad-core Cortex-A72 (ARM V8) 64-bit, tốc độ 1.8GHz
* Bộ nhớ RAM: 4GB LPDDR4-3200 SDRAM
* Mạng không dây: IEEE 802.11ac (2.4GHz và 5.0GHz), Bluetooth 5.0, BLE, Ethernet Gigabit
* Cổng USB: 2 cổng USB 3.0 và 2 cổng USB 2.0
* GPIO: 40 chân GPIO (General Purpose Input/Output)
* Cổng video: 2 cổng micro-HDMI (hỗ trợ 4K)
* Cổng hiển thị: MIPI DSI display port
* Cổng camera: MIPI CSI camera port
* Lưu trữ: Khe cắm thẻ micro-SD card để nạp hệ điều hành và lưu trữ dữ liệu
* Cổng cấp nguồn: USB-C (cung cấp nguồn 5V)
* Đầu ra 5V: Đầu ra 5V ở mỗi chân GPIO

### 3.2.2.2. Cảm biến quang phổ

Dựa trên các lý thuyết đã nghiên cứu, để khảo sát ảnh hưởng của ánh sáng đến phân tử glucose trong ngón tay, phổ sáng trong vùng VIS – NIR được áp dụng. Vì vậy, cảm biến sử dụng trong đề tài cần có khả năng thu thập tín hiệu từ nhiều bước sóng khác nhau trong dải bước sóng này. Với khả năng đo lường các bước sóng trong khoảng từ 410nm đến 940nm, cảm biến quang phổ AS7265x đã được lựa chọn sử dụng trong nghiên cứu này.

Cảm biến AS7265x được cấu thành từ ba cảm biến riêng biệt: AS72651, AS72652 và AS72653. Cảm biến này có khả năng thu thập dữ liệu qua sáu kênh bước sóng trong dải từ 410nm đến 940nm, bao phủ 18 dải tần số quang phổ riêng biệt, giúp đo đạc chính xác các đặc tính quang học của các phân tử glucose. Cảm biến hỗ trợ hai giao thức truyền dữ liệu chính là I2C và UART, giúp truyền tải hiệu quả thông tin quang phổ thu được đến các thiết bị xử lý. Hình 7 mô tả cấu tạo và cách bố trí các thành phần trong cảm biến AS7265x.



Hình 3. 2: Cấu tạo của cảm biến AS7265x

**Thông số kỹ thuật:**

* Kết hợp 3 cảm biến AS72651, AS72652, AS72653.
* Dãi bước sóng: 410nm – 940nm
* Số dải tần số quang phổ: 18 bước sóng.
* Độ chính xác: ±12%.
* Giao tiếp: I2C hoặc UART.
* Nguồn vào: 3.3V

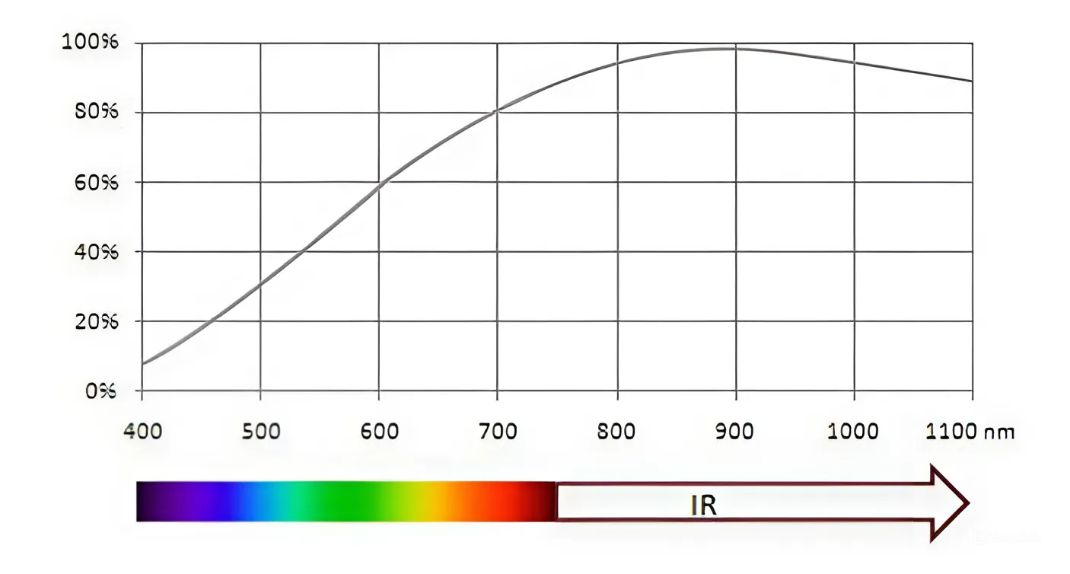
### 3.2.2.3. Nguồn sáng

Lựa chọn nguồn sáng và cảm biến là điểm mấu chốt trong nghiên cứu này, vì việc sử dụng bước sóng phù hợp sẽ cải thiện khả năng xuyên qua mô và giúp mô tả chính xác khả năng hấp thụ, tán xạ của các phân tử glucose trong quá trình đo lường. Để lựa chọn nguồn sáng phù hợp, đèn LEDs và đèn Halogen đã được so sánh ở bảng 2, đặc biệt chú trọng vào dải bước sóng phát ra của từng loại đèn.

Bảng 3. 1: So sánh đèn LED và đèn Halogen

|  |  |  |
| --- | --- | --- |
| Đặc điểm | Đèn Led | Đèn Halogen |
| Bước sóng | Cố định | Phát ra ánh sáng với dải bước sóng trải dài. |
| Nhiệt năng tỏa ra | Thấp | Cao |
| Tuổi thọ | Cao | Thấp |

Dù đèn LEDs có hiệu suất tốt hơn, với nhiệt năng tỏa ra thấp và tuổi thọ cao, nhưng trong nghiên cứu này, yếu tố bước sóng là lý do chính để lựa chọn nguồn sáng. Để khảo sát ảnh hưởng của các bước sóng trong dải VIS – NIR đối với nồng độ glucose, một nguồn sáng có dải bước sóng rộng sẽ cung cấp những thông tin chính xác hơn về sự hấp thụ và tán xạ ánh sáng trong mô. Vì vậy, đèn Halogen Philips W5W T10 đã được chọn lựa vì khả năng phát ra dải bước sóng rộng, đáp ứng tốt mục tiêu nghiên cứu của đề tài.



Hình 3. 3: Dãi quang phổ của bóng đèn Halogen

**Thông số kỹ thuật:**

* Điện áp hoạt động: 12V.
* Công suất: 5W.
* Cường độ sáng: 50Lm.
* Kích thước: 10mm x 26.8 mm.
* Tuổi thọ: ~400h.

### 3.2.2.4. Màn hình OLED

Để hiển thị kết quả dự đoán nồng độ glucose, đề tài lựa chọn màn hình OLED SSD 1306 0.96 inch. Màn hình này không chỉ mang lại chất lượng hiển thị sắc nét và rõ ràng ngay cả trong điều kiện ánh sáng ban ngày, mà còn tiết kiệm năng lượng và chi phí hợp lý.

Thông số kỹ thuật:

* Điện áp hoạt động: 2.2 – 5.5 VDC
* Công suất tiêu thụ: 0.04W
* Số điểm hiển thị: 128x64 điểm, cho phép hiển thị thông tin rõ ràng và chi tiết
* Kích thước: 0.96 inch,
* Giao tiếp: I2C (địa chỉ 0x3C hoặc 0x3D)

### 3.2.2.5. USB Dongle 4G

Mặc dù Raspberry Pi 4B có khả năng sử dụng WiFi để kết nối với Internet, việc sử dụng WiFi trong hệ thống IoT của đề tài vẫn gặp phải một số hạn chế. Đặc biệt, khi tín hiệu WiFi không ổn định hoặc không khả dụng, việc kết nối gặp khó khăn, dẫn đến dữ liệu đo không được cập nhật liên tục. Điều này có thể gây ra sự không nhất quán trong lịch sử đo nồng độ glucose, ảnh hưởng đến khả năng theo dõi đường huyết của người dùng.

Để khắc phục vấn đề này, thay vì sử dụng WiFi, giải pháp thay thế là sử dụng kết nối 4G. Công nghệ 4G đã được triển khai rộng rãi và phủ sóng hầu như khắp mọi nơi, đảm bảo tính ổn định trong việc truyền tải dữ liệu và hỗ trợ tính tiện dụng, di động của thiết bị, giúp người dùng theo dõi nồng độ glucose mọi lúc, mọi nơi.

USB Dongle 4G SIMCOM LTE CAT1 V2 được lựa chọn trong đề tài này vì khả năng cung cấp kết nối internet với tốc độ cao, đáp ứng yêu cầu truy cập internet ở khắp mọi nơi.

**Thông số kỹ thuật:**

* Nguồn cấp: USB-A 5V 1A
* Băng tần: LTE-CAT 1
* Tốc độ tải xuống (download): 10Mbps
* Tốc độ tải lên (upload): 5Mbps
* Khe đọc SIM: Hỗ trợ kết nối với SIM để duy trì kết nối 4G

### 3.2.2.6. Mạch tăng áp

Mạch tăng áp DC-DC LTC1871 được lựa chọn trong đề tài nhằm cung cấp nguồn điện ổn định cho đèn Halogen, đặc biệt trong các ứng dụng yêu cầu công suất lớn và hiệu suất cao. Với hiệu suất mạch đạt 84.1% khi điện áp đầu vào dưới 5V và trên 93% khi điện áp đầu vào từ 5V, LTC1871 đảm bảo việc chuyển đổi nguồn hiệu quả và tiết kiệm năng lượng. Vì đèn Halogen yêu cầu công suất lớn trong thời gian dài, nhiệt độ của mạch có thể tăng cao, nên hệ thống tản nhiệt được tích hợp nhằm duy trì hiệu suất ổn định và ngăn ngừa quá nhiệt. Bên cạnh đó, mạch còn tích hợp màn hình LED hiển thị giá trị điện áp đầu ra hiện tại, giúp dễ dàng điều chỉnh áp không cần sử dụng đồng hồ VOM.

**Thông số kỹ thuật:**

* Trọng lượng: 35 gam
* Kích thước: 67 x 42 x 18 mm
* Điện áp đầu vào: 3.5V – 30V
* Công suất tối đa: 100W
* Điện áp đầu ra: 3.5V – 30V
* Dòng vào: 6A

### 3.2.2.7. Mạch giảm áp

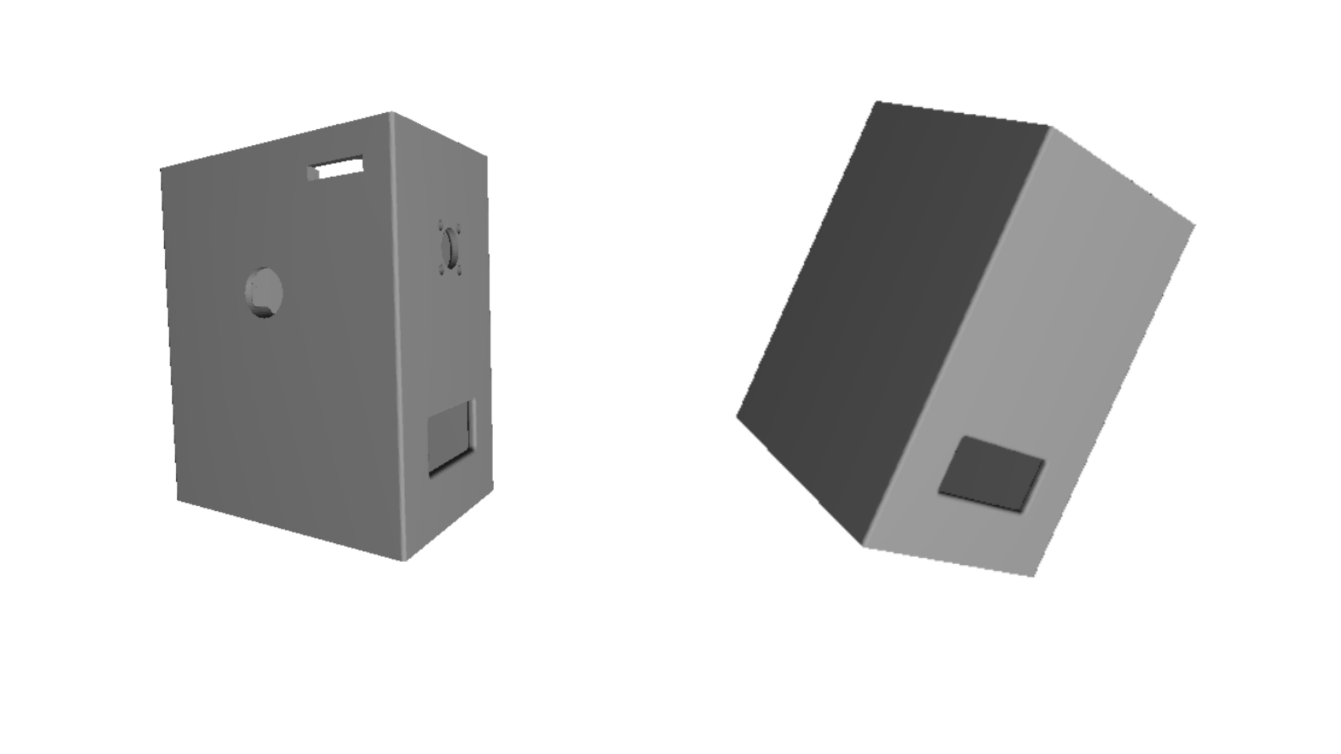
Raspberry Pi 4 Model B hoạt động với điện áp 5V, vì vậy mạch giảm áp DC LM2596 3A được sử dụng để hạ điện áp từ hai pin 18650 xuống mức 5V ổn định. Mạch LM2596 có thiết kế nhỏ gọn và khả năng giảm áp từ 30V xuống 1.5V, đồng thời đạt hiệu suất cao và ổn định lên đến 92%. Raspberry Pi 4 yêu cầu nguồn 5V/3A, và LM2596 3A có khả năng cung cấp dòng 3A, đáp ứng đủ nhu cầu điện năng cho Raspberry Pi.

**Thông số kỹ thuật:**

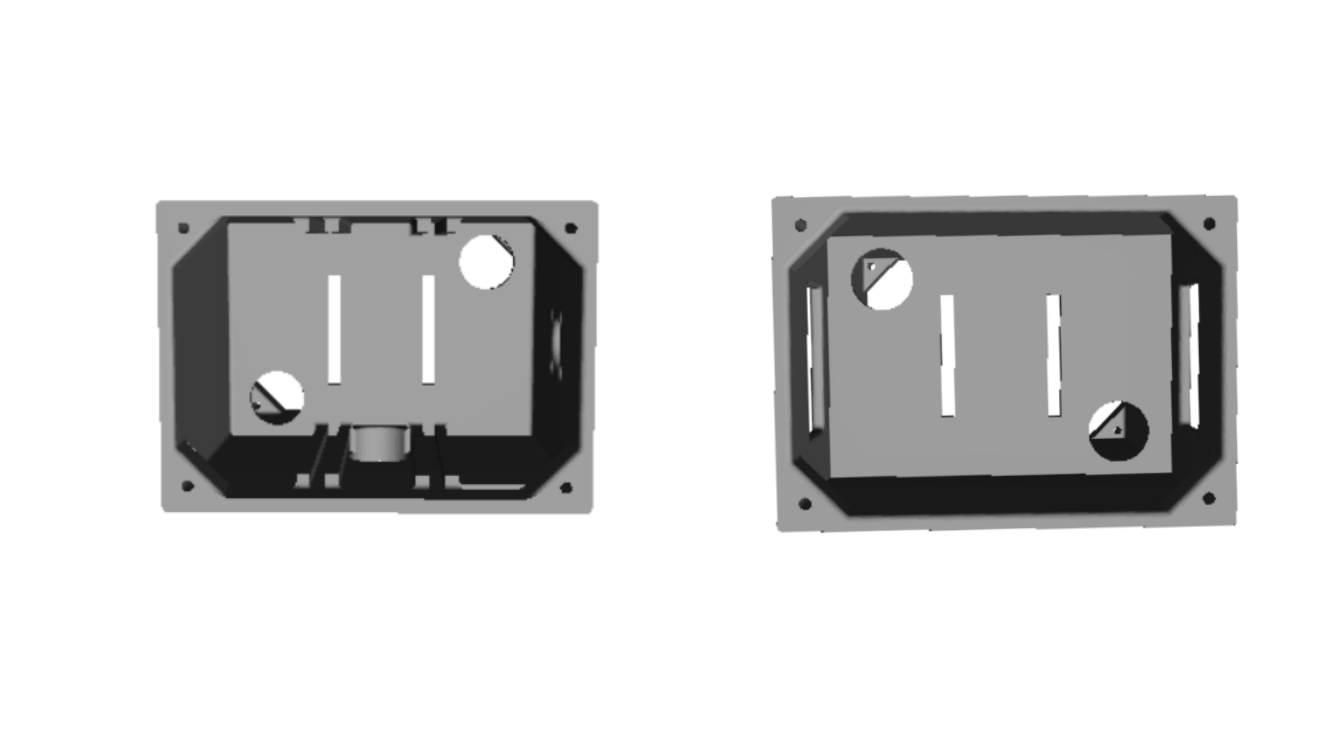
* Điện áp đầu vào: 3V – 30V.
* Điện áp đầu ra: Điều chỉnh từ 1.5V – 30V.
* Dòng tối đa: 3V.
* Hiệu suất: 92%.
* Công suất: 15W.

### 3.2.3. Thiết kế 3D đề tài

Hộp của thiết bị được chia thành 2 phần: thân trên và thân dưới như hình 49 mô tả 3D mặt ngoài của hệ thống và hình 50 thể hiện cấu trúc bên trong của hộp.



Hình 4. 33: Mô hình 3D mặt trước



Hình 4. 34: Mô hình 3D bên trong

Phần thân trên của thiết bị được thiết kế để chứa các thành phần như đèn Halogen, cảm biến, mạch giảm áp cho đèn Halogen, màn hình OLED SSD1306 0.96 inch và bo mạch chính. Kệ đặt tay được tích hợp ở thân trên, giúp người dùng dễ dàng định vị tay tại vị trí chính xác giữa đèn và cảm biến, từ đó đảm bảo độ ổn định và chính xác cho kết quả đo lường.

Phần thân dưới đảm nhận vai trò nguồn và trung tâm xử lý chính của thiết bị. Bao gồm pin, relay, Raspberry Pi 4 Model B với mạch giảm áp 5V phù hợp, cùng với USB dongle 4G.

### 3.2.4. Sơ đồ kết nối phần cứng

Thiết bị sử dụng pin Lithium-ion làm nguồn năng lượng chính, với hai pin Lithium-ion được kết nối nối tiếp. Hệ thống cung cấp nguồn 5V cho Raspberry Pi thông qua mạch giảm áp và 12V cho đèn Halogen thông qua mạch tăng áp. Cấu hình chi tiết và sơ đồ kết nối phần cứng được trình bày trong bảng 12.

Bảng 4. 9: Sơ đồ kết nối phần cứng

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Raspberry Pi | AS7265x | Relay | Oled | USB Dongle |
| 3.3V | VCC |  |  |  |
| 5V |  | VCC | VCC |  |
| GND | GND | GND | GND |  |
| GPIO17 |  | IN |  |  |
| SDA | SDA |  | SDA |  |
| SCL | SCL |  | SCL |  |
| USB |  |  |  | USB |

AS7265x và OLED SSD 1306 0.96 inch đều sử dụng phương thức I2C để truyền nhận tín hiệu. Với AS7265x sử dụng địa chỉ 0x49 và OLED SSD 1306 sử dụng địa chỉ là 0x3C.

## 3.3. Kịch bản lấy mẫu

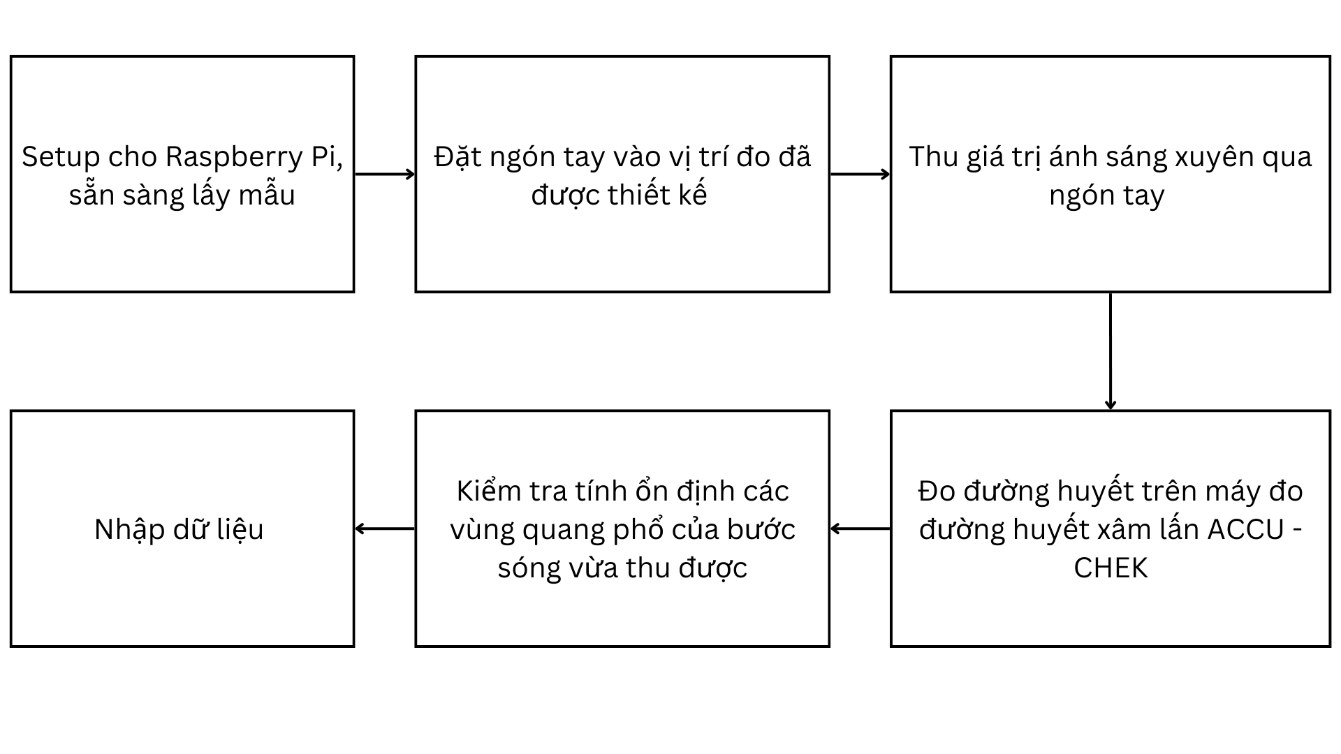
Quá trình lấy mẫu sẽ được thực hiện trong môi trường trong nhà, tránh ánh sáng tự nhiên và duy trì nhiệt độ phòng ổn định để đảm bảo các yếu tố ngoại vi không ảnh hưởng đến kết quả. Các mẫu sẽ được thu thập ngẫu nhiên trong ngày từ nhiều tình nguyện viên. Để đánh giá sự biến động của nồng độ glucose trong suốt ngày, quá trình đo sẽ được thực hiện tại bốn thời điểm chính trong ngày:

* Chưa ăn sáng.
* Từ 1 đến 3 giờ sau khi ăn sáng, trưa và chiều.
* Sau 3 giờ kể từ khi ăn xong bữa chính.
* Ở các thời điểm ngẫu nhiên

Đối với các phép đo xâm lấn sử dụng máy ACCU-CHEK, việc vệ sinh đầu ngón tay thật kỹ bằng cồn và bông là rất quan trọng trước khi thực hiện đo. Điều này nhằm loại bỏ tạp chất có thể làm sai lệch kết quả đo. Ngoài ra, cần đảm bảo lượng máu lấy đủ để tránh ảnh hưởng đến độ chính xác của kết quả và giảm thiểu nhiễu.

### 3.3.1. Quy trình lấy mẫu

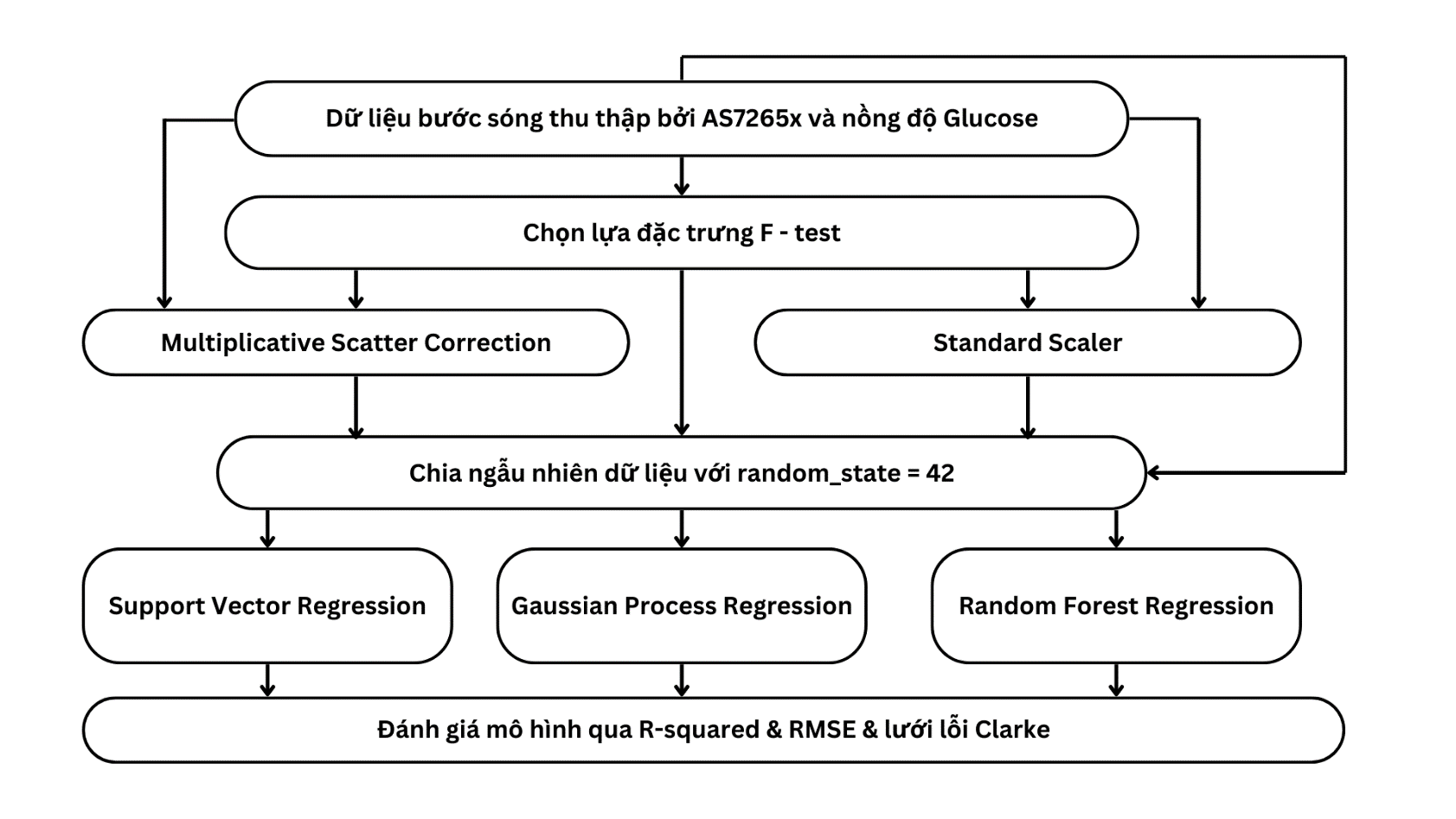
Quy trình lấy mẫu trong hệ thống đo nồng độ glucose không xâm lấn được thực hiện với Raspberry Pi đảm nhận vai trò điều khiển các thiết bị ngoại vi để thực hiện việc lấy mẫu. Một yếu tố quan trọng trong quy trình này là việc điều chỉnh chính xác ngón tay vào vị trí đã được thiết kế, vì sự thay đổi vị trí ngón tay sẽ ảnh hưởng trực tiếp đến giá trị bước sóng thu được. Sau khi ánh sáng chiếu vào ngón tay, một phần ánh sáng sẽ bị hấp thụ và tán xạ, và các ánh sáng còn lại sẽ được cảm biến quang phổ thu nhận. Dữ liệu đo từ cảm biến sẽ được ứng với kết quả của máy đo đường huyết xâm lấn ACCU-CHEK, thiết bị này cung cấp giá trị glucose thực tế của người đo tại thời điểm đó. Nhằm đảm bảo độ chính xác, biến động của các bước sóng đo được sẽ được theo dõi để đảm bảo tính tương thích với các nghiên cứu trước đó. Cuối cùng, dữ liệu thu được từ cảm biến sẽ bao gồm 18 bước sóng (X = {410nm, 435nm, 460nm,..., 940nm}) và giá trị nồng độ glucose (Y = {Glucose}) từ ACCU-CHEK, tạo thành bộ dữ liệu được sử dụng cho việc huấn luyện mô hình học máy để dự đoán nồng độ glucose không xâm lấn. Quy trình lấy mẫu của mô tả chi tiết ở hình 9.



Hình 3. 4: Quy trình lấy mẫu

## 3.4. Triển khai phần mềm

Dữ liệu cuối cùng sẽ gồm 18 bước sóng hay một tập hợp biến đầu vào X = {410nm,435nm, 460nm, 485nm, 510nm, 535nm, 560nm, 585nm, 610nm, 645nm, 680nm, 705nm, 730nm, 760nm, 810nm, 860nm, 900nm,,940nm} thu được từ cảm biến và nhãn y = {Glucose} thu được từ ACCU – CHEK. Sơ đồ mô tả triển khai phần mềm trong dự đoán nồng độ Glucose được thể hiện ở hình 3.5.



Hình 3. 5: Sơ đồ thể hiện quy trình dự đoán nồng độ glucose.

Tiền xử lý dữ liệu thô đóng vai trò quan trọng trong quá trình huấn luyện các mô hình máy học, giúp làm rõ các đặc trưng của biến độc lập và đánh giá tác động của chúng lên biến phụ thuộc. Việc xử lý dữ liệu đúng cách không chỉ giúp cải thiện độ chính xác của mô hình mà còn giúp đơn giản hóa tập dữ liệu, làm cho các thuật toán học máy dễ dàng huấn luyện và đạt hiệu quả cao hơn. Đề tài đề xuất 2 phương pháp tiền xử lý Muliplicative Scatter Correction và Standard Scaler và kiểm tra mức độ hiệu quả của 2 mô hình với việc không sử dụng bất kỳ phương pháp nào.

Các mô hình học máy được sử dụng trong nghiên cứu bao gồm Support Vector Regression (SVR), Gaussian Process Regression (GPR) và Random Forest Regression (RFR). Kết quả cuối cùng sẽ được đánh giá bằng hai phương pháp chính: phương pháp phân tích thông qua các chỉ số R² và RMSE, cùng với phương pháp lâm sàng Clarke Grid Error.

### 3.4.1. Tiền xử lý dữ liệu thô

Dữ liệu quang phổ VIS-NIR thô thường bị ảnh hưởng bởi nhiều yếu tố gây nhiễu, chẳng hạn như tán xạ ánh sáng, sự không đồng nhất của mẫu vật hoặc sai lệch từ thiết bị đo. Những yếu tố này làm cho tín hiệu trở nên khó phân tích và ảnh hưởng đến độ chính xác trong việc dự đoán nồng độ glucose. Để khắc phục vấn đề này, tiền xử lý dữ liệu đóng vai trò quan trọng trong việc loại bỏ các nhiễu không mong muốn và làm nổi bật các đặc trưng quan trọng của dữ liệu.

Các kỹ thuật tiền xử lý như hiệu chỉnh tán xạ (Multiplicative Scatter Correction - MSC) và chuẩn hóa dữ liệu (Standard Scaler - SS) được sử dụng để tối ưu hóa tín hiệu, giúp làm rõ mối quan hệ giữa bước sóng và nồng độ glucose. Việc áp dụng các phương pháp này không chỉ nâng cao hiệu suất của các mô hình phân tích và hiệu chuẩn, mà còn giúp tiết kiệm thời gian và tài nguyên, làm cho việc sử dụng quang phổ NIR trở nên hiệu quả và đáng tin cậy hơn.

Trong đề tài này, tập trung nghiên cứu và so sánh hai phương pháp tiền xử lý chính: hiệu chỉnh tán xạ MSC và chuẩn hóa đặc trưng SS, nhằm cải thiện độ chính xác và hiệu quả của hệ thống đo lường.

### 3.4.1.1. Multiplicative Scatter Correction

Multiplicative Scatter Correction [13] là một phương pháp trong phân tích quang phổ, đặc biệt là trong phân tích NIR được sử dụng để loại bỏ nhiễu do tán xạ trong dữ liệu quang phổ. MSC hoạt động với quy trình như sau:

* Thực hiện hồi quy từng phổ để tính toán được hệ số dịch và hệ số tỷ lệ so với phổ tham chiếu của toàn bộ tập dữ liệu qua công thức:

Với Xi là phổ đo của mẫu thứ i, Xm phổ trung bình của toàn bộ tập dữ liệu, hệ số dịch – biễu diễn sự chênh lệch giữa phổ đo được và phổ trung bình, hệ số tỷ lệ - biễu diễn sự khác biệt về tỷ lệ giữa phổ đo dược và phổ trung bình.

* Sau khi đã tính được và , phổ thu được từ cảm biến được chuẩn hóa bằng công thức:

là phổ đã được hiểu chỉnh của mẫu thứ i của tập dữ liệu, phổ đo được ban đầu, hệ số dịch, hệ số tỷ lệ. Công thức trên giúp loại bỏ các biến đổi về tỷ lệ ( và độ dịch gây ra bởi bởi tán xạ ánh sáng hoặc các yếu tố vật lý.

Phương pháp hiệu chỉnh tán xạ Multiplicative Scatter Correction (MSC) yêu cầu một quang phổ tham chiếu để bắt đầu quá trình xử lý. Lý tưởng nhất, quang phổ tham chiếu này sẽ là quang phổ không chịu ảnh hưởng của hiệu ứng tán xạ. Tuy nhiên, việc có được một quang phổ lý tưởng là không dễ dàng. Do đó, trong thực tế, quang phổ trung bình của toàn bộ tập dữ liệu thường được sử dụng làm tham chiếu cho thuật toán MSC.

Khi đo nồng độ glucose bằng phương pháp NIR, các yếu tố như kích thước ngón tay, góc đặt tay hoặc sự ảnh hưởng của melanin trong da có thể gây ra sự biến đổi ngẫu nhiên về hiệu ứng kích thước hạt và chiều dài quang học đi xuyên qua vật thể. Vì vậy, việc sử dụng quang phổ trung bình của tập dữ liệu giúp làm giảm ảnh hưởng của các yếu tố này, từ đó cải thiện độ chính xác của quá trình hiệu chỉnh và tăng tính ổn định trong kết quả đo.

### 3.4.1.2. Standard Scaler

Standard Scaler [14] là một kỹ thuật chuẩn hóa dữ liệu phổ biến trong học máy, đặc biệt quan trọng khi làm việc với các tập dữ liệu có các đặc trưng với đơn vị và phạm vi khác nhau. Mục tiêu chính của phương pháp này là đưa tất cả các đặc trưng của dữ liệu về cùng một thang đo, giúp chúng trở nên đồng nhất và dễ dàng hơn khi áp dụng vào các mô hình học máy. Cụ thể, phương pháp này sẽ chuẩn hóa dữ liệu sao cho giá trị trung bình của các đặc trưng bằng 0 và độ lệch chuẩn bằng 1.

Công thức chuẩn hóa của Standard Scaler được biểu diễn:

Trong đó:

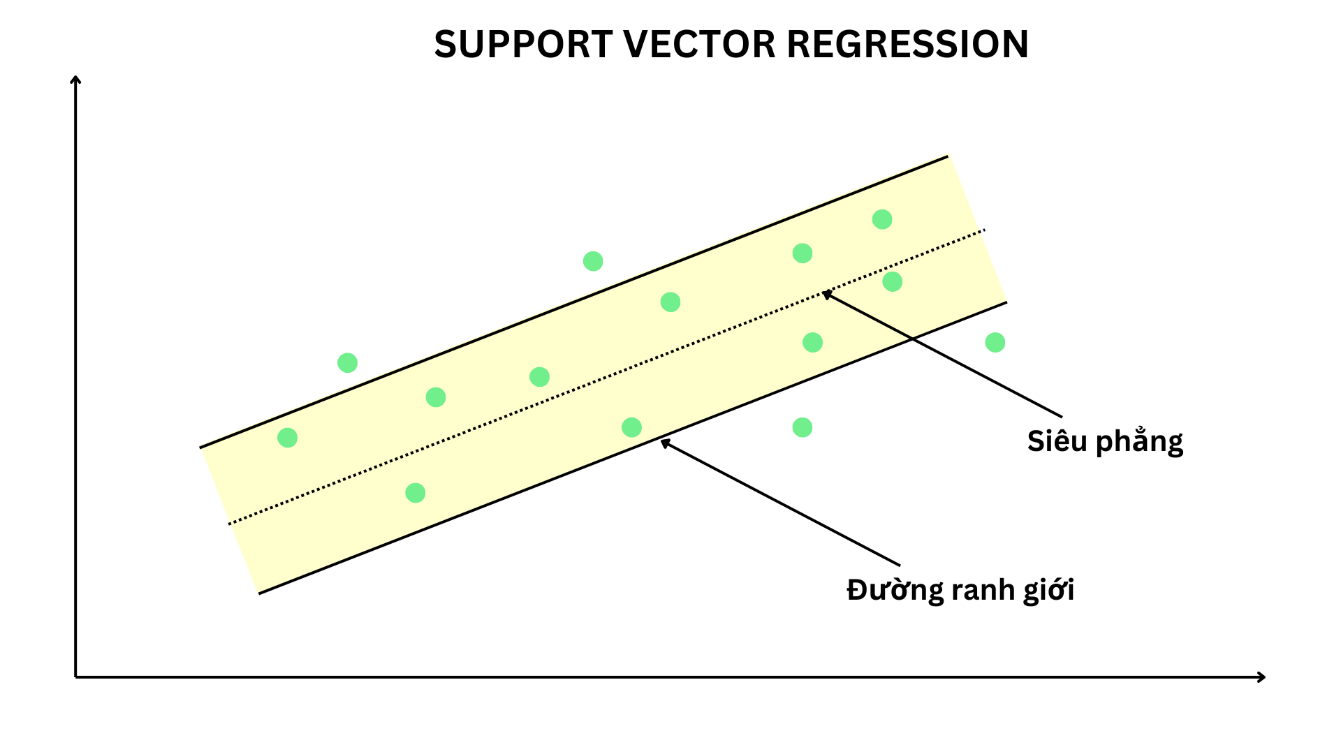
* : giá trị đặc trưng của tập dữ liệu gốc
* biểu diễn giá trị trung bình của đặc trưng.
* là độ lệch chuẩn của đặc trưng
* : giá trị chuẩn hóa của đặc trưng

Khi áp dụng Standard Scaler vào tập dữ liệu trong nghiên cứu quang phổ VIS-NIR, việc chuẩn hóa này đóng vai trò quan trọng trong việc loại bỏ các sự khác biệt về quy mô giữa các bước sóng và giúp các mô hình học máy có thể hiểu và phân tích dữ liệu một cách chính xác hơn. Việc chuẩn hóa cũng giúp cải thiện hiệu suất của các thuật toán học máy trong quá trình huấn luyện và dự đoán. Các mô hình Random Forest Regression, Support Vector Machine đều yêu cầu dữ liệu đầu vào có phân phối chuẩn để đạt được kết quả tối ưu.

### 3.4.2. Mô hình huấn luyện

### 3.4.2.1 Support Vector Regression

Support Vector Regression (SVR) [15] là một mô hình hồi quy dựa trên nguyên lý của Support Vector Machine (SVM), với mục tiêu chính là tìm ra một "vùng ống" trong không gian đặc trưng, nơi mà hầu hết các điểm dữ liệu huấn luyện có thể nằm trong đó với một mức sai số chấp nhận được (được xác định bởi tham số epsilon). Khác với các phương pháp hồi quy truyền thống chỉ ước lượng một đường hoặc mặt phẳng hồi quy, SVR tập trung vào việc xây dựng một vùng hồi quy tối ưu, giúp giảm thiểu tác động của các điểm dữ liệu nằm ngoài vùng này.



Hình 3. 6: Support Vector Regression

Một tham số quan trọng trong SVR là *C*, được sử dụng để điều chỉnh mức độ cân bằng giữa việc tối thiểu hóa sai số trên các điểm dữ liệu ngoài vùng ống và duy trì một mô hình đơn giản.

Tương tự như SVM, SVR sử dụng kỹ thuật "kernel trick" để ánh xạ dữ liệu từ không gian đầu vào sang một không gian đặc trưng cao hơn. Điều này cho phép SVR giải quyết hiệu quả cả các bài toán hồi quy tuyến tính lẫn phi tuyến tính. Các kernel phổ biến trong SVR bao gồm:

* **Kernel tuyến tính (Linear):** Phù hợp với các bài toán mà dữ liệu có mối quan hệ tuyến tính rõ ràng.
* **Kernel hàm cơ sở Radial (RBF):** Hiệu quả trong việc xử lý các bài toán phi tuyến tính, nơi dữ liệu không được phân bố theo tuyến tính.
* **Kernel đa thức (Polynomial):** Hữu ích khi dữ liệu có mối quan hệ phi tuyến bậc cao.

Kernel RBF là lựa chọn phổ biến trong nhiều bài toán nhờ khả năng xử lý hiệu quả các mối quan hệ phi tuyến phức tạp, trong khi kernel tuyến tính lại phù hợp hơn với dữ liệu có đặc trưng tuyến tính rõ ràng. Trong đề tài này, dữ liệu glucose chịu ảnh hưởng từ nhiều yếu tố như sự khác biệt về kích thước ngón tay, góc đặt tay không đồng nhất, và các yếu tố sinh học khác, dẫn đến sự biến đổi phi tuyến trong giá trị bước sóng thu được. Do đó, kernel RBF được lựa chọn làm phương pháp tối ưu để huấn luyện mô hình, đảm bảo khả năng xử lý các quan hệ phi tuyến của dữ liệu.

Công thức của SVR có thể được biểu diễn như sau:

Trong đó, đại diện cho dữ liệu huấn luyện, là dữ liệu phổ đầu vào được dự đoán, là hệ số Lagrange tương ứng với các vector hỗ trợ ở phía trên đường hồi quy, tức là các điểm mà có giá trị lớn hơn mức sai số cho phép (epsilon). là hệ số Lagrange liên quan đến các vector hỗ trợ ở phía dưới đường hồi quy, tức là các điểm có giá trị nhỏ hơn mức sai số cho phép. là hàm kernel và là tham số bias

### 3.4.2.2. Gaussian Process Regression

Gaussian Process Regression (GPR) [16] là một phương pháp hồi quy được sử dụng để mô hình hóa các hàm liên tục. GPR dựa trên lý thuyết của Gaussian Process (GP) - một mô hình xác suất giúp mô phỏng các hàm số.

Một Gaussian Process (GP) là một tập hợp vô hạn các biến ngẫu nhiên, trong đó mỗi tập con hữu hạn của các biến ngẫu nhiên này có phân phối chuẩn (Gaussian). Một GP có thể được mô tả bởi hàm “mean function” – biểu thị giá trị trung bình của process tại điểm *x* và kernel – mô tả sự tương quan giữa các điểm x và x’.

Trong đề tài này, mối quan hệ giữa các bước sóng ánh sáng và nồng độ glucose là một dạng hồi quy phi tuyến tính, trong đó biến mục tiêu (glucose) phụ thuộc vào các đặc trưng của bước sóng. GPR sẽ tìm cách mô hình hóa biến mục tiêu dưới dạng một Gaussian Process với các kernel function.

Phương trình dự đoán cho Gaussian Process Regression được định nghĩa như sau:

Trong đó:

Phương trình là phương trình mô tả phân phối xác xuất của giá trị dự đoán tại điểm mới . là phân phối xác suất của giá trị với điều kiện là dữ liệu huấn luyện đầu vào X và đầu ra y. là phân phối chuẩn với giá trị trung bình dự đoán của phân phối tại điểm , phương sai của phân phối tại điểm , biểu thị độ không chắc chắn của dự đoán tại .

Quá trình dự đoán giá trị glucose sử dụng mô hình Gaussian Process Regression với dữ liệu đầu vào bao gồm các giá trị từ 410nm đến 910nm và các giá trị glucose tương ứng.

### 3.4.2.3. Random Forest Regression

Theo định nghĩa 1.1 của [17], Random Forest là một mô hình phân loại bao gồm một tập hợp các mô hình phân loại có cấu trúc cây (decision tree) {), k = 1,...}, trong đó là các vector ngẫu nhiên độc lập và phân phối giống nhau.

Các giá trị quang phổ và giá trị glucose thường có mối quan hệ phức tạp, chịu ảnh hưởng từ các yếu tố như nhiễu từ môi trường đo hoặc sự không đồng nhất khi lấy mẫu trực tiếp từ ngón tay. Những yếu tố này làm cho mối quan hệ giữa dữ liệu đầu vào và đầu ra trở nên phi tuyến và khó mô hình hóa bằng các phương pháp tuyến tính thông thường. Do đó, Random Forest được dùng để huấn luyện và biểu diễn các mối quan hệ phi tuyến trong tập dữ liệu, giúp tăng độ chính xác của mô hình dự đoán.

Với bài toán hồi quy, Random Forest Regression được sử dụng thay thế cho Random Forest với nguyên tắc hoạt động tương tự nhưng có một vài khác biệt nhỏ.

Random Forest Regression được xây dựng từ tập hợp các cây quyết định hồi quy (regression trees), mỗi cây hoạt động độc lập trên các tập dữ liệu con được chọn ngẫu nhiên từ tập huấn luyện ban đầu. Giả sử có T cây và mỗi cây dự đoán một giá trị thì giá trị dự đoán cuối cùng là:

Phương pháp này giúp giảm thiểu hiện tượng overfitting bằng cách kết hợp dự đoán từ nhiều cây, đồng thời tăng khả năng tổng quát hóa và độ chính xác của mô hình trên dữ liệu mới.

### 3.4.3. Đánh giá mô hình

Bên cạnh việc đánh giá độ chính xác của mô hình máy học dựa trên các thước đo như RMSE và R2 , đối với mô hình dự đoán nồng độ glucose trong máu, hiệu suất được xem xét qua hai khía cạnh chính: độ chính xác phân tích và độ chính xác lâm sàn.

**Độ chính xác phân tích** tập trung vào việc đo lường mức độ chênh lệch giữa giá trị glucose dự đoán từ mô hình hoặc thiết bị và giá trị tham chiếu từ các phương pháp đo lường tiêu chuẩn. Các chỉ số như RMSE và R2 được sử dụng để đánh giá khả năng của mô hình trong việc đưa ra dự đoán chính xác trên các tập dữ liệu huấn luyện và kiểm thử. Mục tiêu chính của phân tích này là đảm bảo rằng mô hình có độ tin cậy cao về mặt thống kê, giúp cải thiện khả năng dự đoán trong môi trường nghiên cứu.

Ngược lại, **độ chính xác lâm sàng** tập trung vào việc đánh giá mức độ phù hợp của các dự đoán đối với điều trị thực tế của bệnh nhân. Một công cụ phổ biến để đánh giá này là Clarke Error Grid, chia kết quả dự đoán thành các vùng theo mức độ rủi ro, từ không đáng kể (vùng A) đến nguy hiểm (vùng E). Clarke Error Grid cung cấp một cái nhìn trực quan về sai số dự đoán, xác định những dự đoán có thể gây ảnh hưởng tiêu cực đến quyết định điều trị. Tỷ lệ phần trăm dữ liệu nằm trong các vùng ít rủi ro (A và B) cho thấy mức độ phù hợp của mô hình trong các tình huống thực tế, giúp hỗ trợ các bác sĩ đưa ra quyết định điều trị chính xác hơn và đảm bảo an toàn cho bệnh nhân.

### 3.4.3.1. Độ chính xác phân tích

Trong quá trình xây dựng mô hình máy học, để xem xét mô hình có chất lượng tốt hay không, có giải thích được tập dữ liệu không thì việc đánh giá mô hình là cần thiết. Trong đề tài, 2 chỉ số đánh giá mô hình được sử dụng là RMSE và R2

1. Root Mean Square Error (RMSE) một chỉ số thống kê phổ biến để đo lường mức độ sai lệch giữa giá trị thực tế và giá trị dự đoán của mô hình. Nó phản ánh độ chính xác của mô hình bằng cách tính toán sai số trung bình bình phương giữa các giá trị thực tế và giá trị mà mô hình dự đoán. Công thức của RMSE:

Trong đó, n là số lượng mẫu trong tập dữ liệu kiểm tra, y là giá trị thực tế tại mẫu thứ i và y’ là giá trị dự đoán của mô hình tại mẫu thứ i.

RMSE càng nhỏ thì mô hình càng chính xác, vì nó phản ánh mức độ sai lệch trung bình giữa các giá trị dự đoán và giá trị thực tế. Khi RMSE lớn, cho thấy mô hình có sai số lớn và cần phải được cải thiện.

1. Hệ số xác định () là một chỉ số thống kê dùng để đo lường mức độ mà biến độc lập giải thích được sự biến động của biến phụ thuộc trong mô hình hồi quy. Nó cho biết tỷ lệ phần trăm của biến động trong biến phụ thuộc mà mô hình hồi quy có thể giải thích. Công thức tính R2 được biểu diễn:

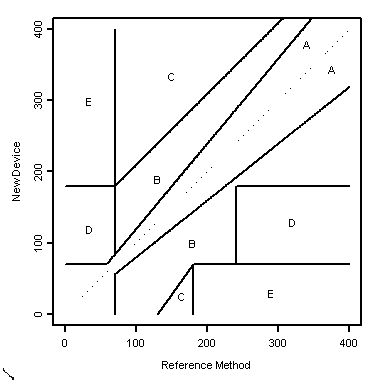
Trong đó, là giá trị thực tế của biến phụ thuộc tại mẫu i, là giá trị dự đoán của mô hình tại mẫu thứ i, là giá trị trung bình của biến phụ thuộc và n là số lượng mẫu.

### 3.4.3.2. Độ chính xác lâm sàn

Clarke Error Grid [1] được phát triển để đánh giá độ chính xác lâm sàng của các dự đoán nồng độ glucose trong máu của bệnh nhân so với giá trị glucose máu được thu từ một thiết bị đo chuẩn. Phương pháp này giúp phân tích sự khác biệt giữa kết quả đo từ thiết bị nghiên cứu và thiết bị chuẩn, từ đó xác định mức độ chính xác của thiết bị đo glucose.

Clarke Error Grid chia đồ thị phân tán giữa giá trị đo của thiết bị tham chiếu và thiết bị được đánh giá thành năm vùng cụ thể, mỗi vùng phản ánh mức độ ảnh hưởng của sai số đến quyết định điều trị:

* **Vùng A** chứa các giá trị nằm trong phạm vi 20% so với cảm biến tham chiếu, cho thấy sự chính xác chấp nhận được của dự đoán.
* **Vùng B** chứa các điểm ngoài phạm vi 20%, nhưng sai số này không dẫn đến việc điều trị sai.
* **Vùng C** chứa các điểm có thể dẫn đến việc điều trị không cần thiết, mặc dù sai số không gây nguy hiểm trực tiếp.
* **Vùng D** chứa các điểm sai nghiêm trọng trong việc phát hiện hạ đường huyết hoặc tăng đường huyết.
* **Vùng E** chứa các điểm có thể gây nhầm lẫn trong việc điều trị, như nhầm lẫn giữa hạ đường huyết và tăng đường huyết, có thể dẫn đến quyết định điều trị sai lệch.



Hình 3. 7: Clarke Grid Error

### 3.4.4. Lựa chọn đặc trưng F-test

F-test [18] là một kiểm định thống kê được sử dụng để đánh giá sự đóng góp của các đặc trưng vào mô hình hồi quy tổng thể. Kiểm định này giúp đánh giá mức độ ảnh hưởng của các đặc trưng đến sự thay đổi của biến mục tiêu. Nếu giá trị F-test lớn, điều đó cho thấy các đặc trưng có khả năng giải thích tốt sự biến động của biến mục tiêu, tức là chúng có ảnh hưởng đáng kể đến kết quả dự đoán của mô hình.

Công thức tính F-test được biểu diễn như sau:

Trong đó:

* SST (Total Sum of Squares) là tổng phương sai của biến mục tiêu, đo lường sự biến động tổng thể của giá trị biến mục tiêu.
* SSE (Sum of Squared Errors) là tổng sai số, đo lường sự sai lệch giữa giá trị thực tế và giá trị dự đoán từ mô hình.
* p là số lượng biến độc lập trong mô hình.
* n là số lượng mẫu dữ liệu.

F-test giúp đánh giá mức độ ảnh hưởng của các bước sóng (features) đến giá trị biến mục tiêu (như nồng độ glucose). Trong bài toán dự đoán nồng độ glucose, F-test được sử dụng để xác định các bước sóng có ảnh hưởng mạnh nhất đến giá trị glucose, từ đó ưu tiên chọn các bước sóng quan trọng nhất, giúp giảm thiểu độ phức tạp tính toán khi huấn luyện các mô hình máy học.

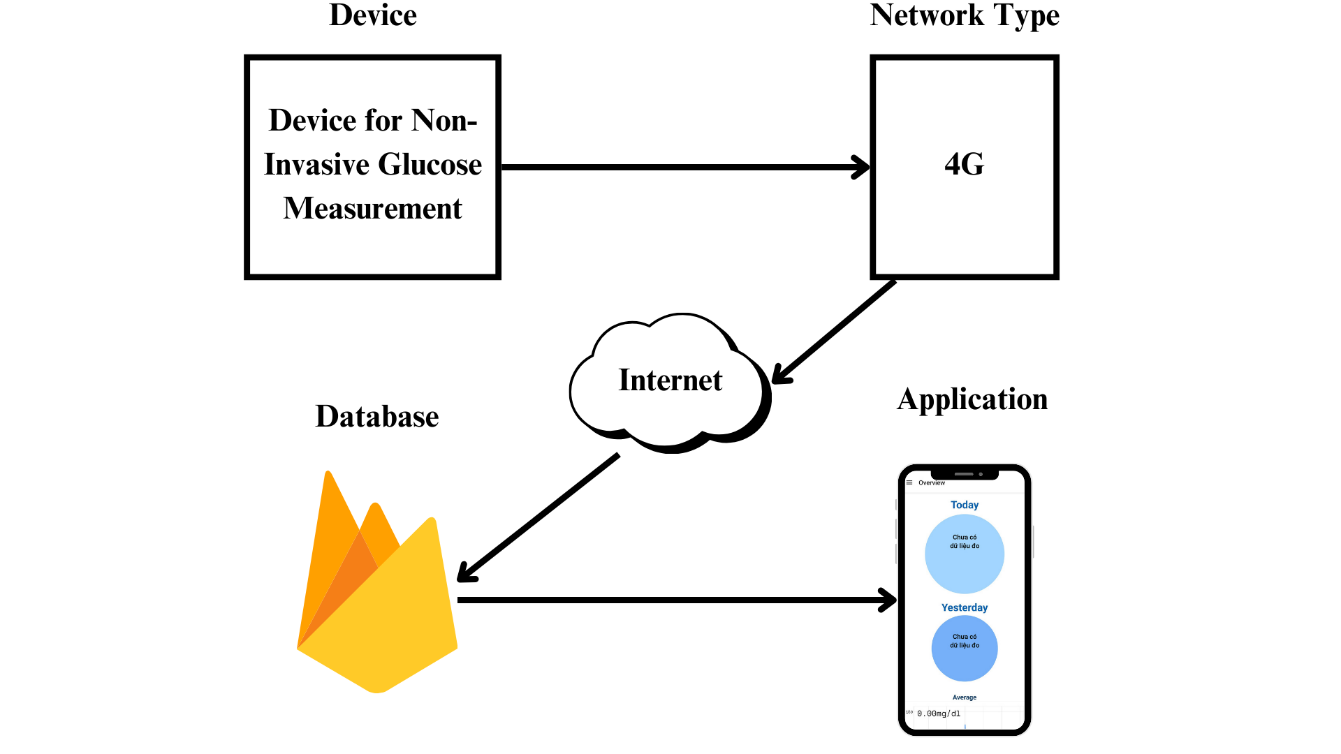
## 3.5. IoMT

IoMT (Internet of Medical Things) là một hệ sinh thái tích hợp các thiết bị y tế, cảm biến và các sản phẩm theo dõi sức khỏe, tất cả đều được kết nối thông qua mạng Internet. Hệ thống này đóng vai trò quan trọng trong việc nâng cao chất lượng chăm sóc sức khỏe và hỗ trợ các bác sĩ, bệnh nhân trong quá trình theo dõi và điều trị.

IoMT giúp thu thập và truyền tải dữ liệu y tế từ các cảm biến và thiết bị đo lường, với khả năng xử lý và hiển thị dữ liệu theo thời gian thực. Điều này giúp giảm độ trễ trong việc cung cấp thông tin quan trọng, từ đó giúp bác sĩ và người sử dụng có thể nắm bắt tình trạng sức khỏe kịp thời.

Khả năng giám sát từ xa của IoMT không chỉ mang lại sự thuận tiện cho bệnh nhân mà còn tạo điều kiện cho bác sĩ theo dõi liên tục các chỉ số sức khỏe như nhịp tim, huyết áp, mức đường huyết và nồng độ oxy trong máu. Nhờ vào các ứng dụng di động, bác sĩ có thể dễ dàng đánh giá tình trạng sức khỏe của bệnh nhân mọi lúc, mọi nơi, giúp đưa ra những lời khuyên kịp thời và chẩn đoán chính xác hơn.

Trong phạm vi đề tài, thiết bị đo đường huyết không xâm lấn sẽ được kết nối với mạng 4G, giúp thiết bị có thể truy cập Internet một cách linh hoạt. Dữ liệu đo được sẽ được lưu trữ trên Cloud Firestore và truyền trực tiếp đến ứng dụng trên điện thoại thông qua kết nối mạng. Mô hình IoMT trong đề tài này được thể hiện trong hình 14.



Hình 3. 8: Mô hình IoMT triển khai trong đề tài

### 3.5.1. Phương thức 4G

Fourth-Generation (4G) là công nghệ truyền dữ liệu không dây kế thừa từ các mạng 2G và 3G, cung cấp tốc độ truyền tải dữ liệu nhanh hơn, có thể lên đến 1 Gb/s. Với tốc độ này, 4G không chỉ đáp ứng nhu cầu truyền tải lượng dữ liệu lớn mà còn hỗ trợ các ứng dụng Internet of Things (IoT), giúp kết nối các thiết bị di động và thiết bị thông minh trong một mạng lưới rộng lớn.

Tính năng phủ sóng rộng và tín hiệu mạnh của 4G giúp khắc phục các hạn chế của mạng WiFi trong việc sử dụng hệ thống đo đường huyết, đặc biệt là khi không phụ thuộc vào vị trí của thiết bị và độ ổn định của tín hiệu WiFi khi xa router. Điều này đảm bảo sự ổn định trong việc kết nối Internet để truyền tải dữ liệu lên cơ sở dữ liệu.

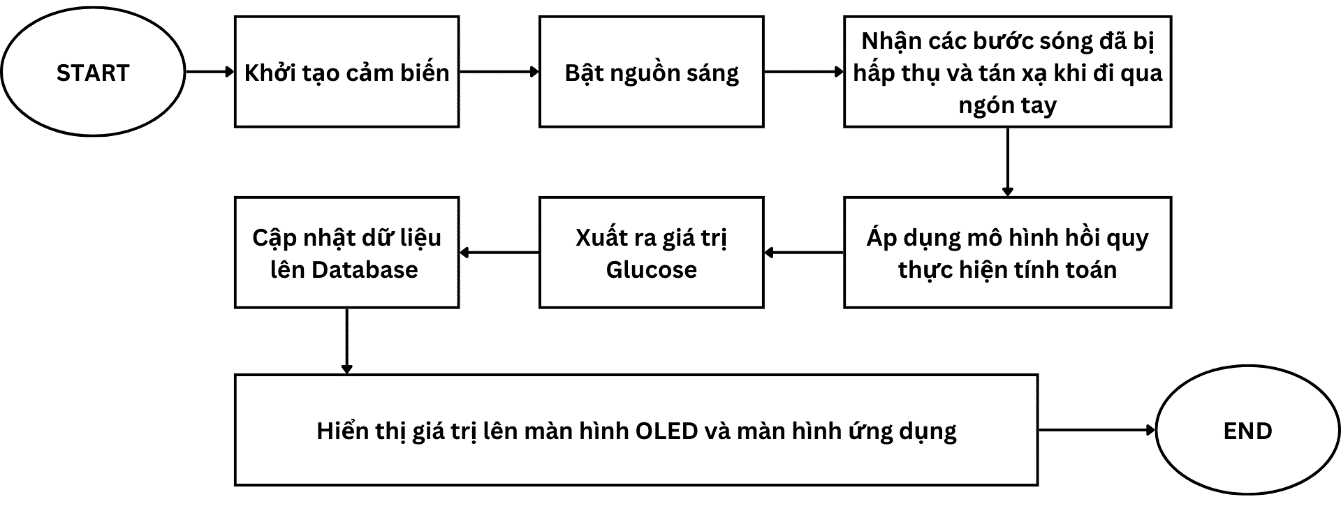
Với điều kiện địa lý tại Việt Nam, sim 4G VietNamobile được chọn trong đề tài nhờ vào tầm phủ sóng rộng và tín hiệu ổn định, đáp ứng tốt các yêu cầu về kết nối và truyền tải dữ liệu của hệ thống.

### 3.5.2. Firebase

Firebase là nền tảng BaaS (Backend as a Service) cung cấp mô hình dịch vụ điện toán đám mây, giúp các nhà phát triển ứng dụng có thể quản lý và vận hành back-end mà không cần phải xây dựng từ đầu. Firebase cung cấp một loạt các dịch vụ và công cụ hỗ trợ phát triển ứng dụng một cách nhanh chóng và hiệu quả.

Trong đề tài, sau khi thực hiện dự đoán giá trị nồng độ Glucose, giá trị này sẽ được đẩy lên Firestore Cloud – một dịch vụ của Firebase dùng để lưu trữ dữ liệu. Dữ liệu sẽ được gửi thông qua các Certificate từ file JSON của Firestore, đảm bảo quá trình lưu trữ và truyền tải dữ liệu diễn ra an toàn và chính xác.

## 3.6. Tổng quan sơ đồ giải thuật của thiết bị



Hình 3. 9: Sơ đồ giải thuật của thiết bị.

Sơ đồ giải thuật của thiết bị được mô tả trong hình ... quy trình bắt đầu việc khởi tạo cảm biến, sau đó đèn Halogen được bật để phát ánh sáng xuyên qua ngón tay. Cảm biến thu nhận các giá trị bước sóng của ánh sáng sau khi bị hấp thụ và tán xạ qua ngón tay, dữ liệu này sẽ được sử dụng làm đầu vào cho mô hình hồi quy đã được huấn luyện. Mô hình tiến hành dự đoán mức Glucose và toàn bộ quy trình xử lý được thực hiện trên Raspberry Pi 4B. Kết quả dự đoán sau đó được cập nhật lên Cloud Database (Firestore) để lưu trữ, đồng thời hiển thị trên màn hình OLED của thiết bị và giao diện ứng dụng điện thoại. Ngoài ra, dữ liệu cũng được lưu trữ dưới dạng file PDF trên ứng dụng, giúp người dùng dễ dàng theo dõi và quản lý thông tin.

# CHƯƠNG 4: KẾT QUẢ VÀ THẢO LUẬN

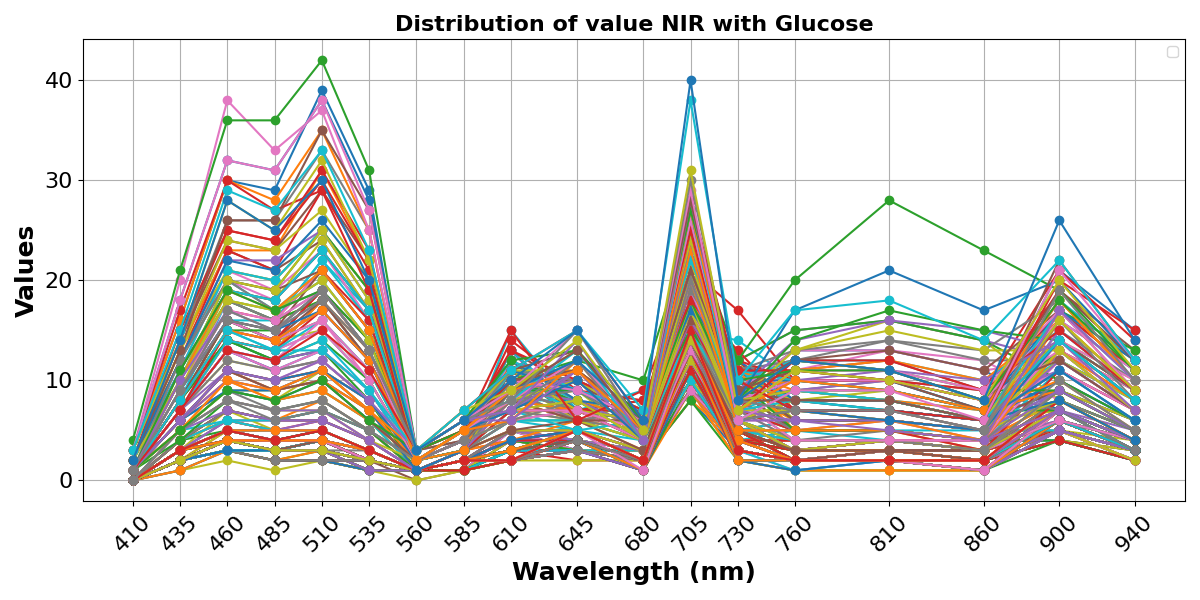
## 4.1. Tập dữ liệu

Tập dữ liệu được xây dựng dựa trên mẫu thử của 102 tình nguyện viên trong đó 5 tình nguyện viên được lấy mẫu thường xuyên vào các thời gian như khi chưa ăn sáng, trước bữa ăn, sau bữa ăn một đến hai tiếng và lấy mẫu ngẫu nhiên.

Tập dữ liệu được thu thập từ các tình nguyện viên có tình trạng sức khỏe tốt, không mắc bệnh lý nghiêm trọng hoặc chỉ có một số bệnh nền ít ảnh hưởng đến sức khỏe tổng thể. Các tình nguyện viên tham gia lấy mẫu có độ tuổi dao động từ 17 đến 25 và một số cá nhân ở độ tuổi 48.

Tuy nhiên, hạn chế của tập dữ liệu này là sự thiếu đa dạng về độ tuổi và bệnh lý. Cụ thể, độ tuổi của các tình nguyện viên chủ yếu tập trung vào nhóm thanh thiếu niên và người trưởng thành trẻ, trong khi các nhóm tuổi khác, đặc biệt là người cao tuổi chưa được triển khai lấy mẫu nhiều. Bên cạnh đó, tập dữ liệu chưa bao gồm nhiều loại bệnh lý khác nhau, điều này có thể ảnh hưởng đến khả năng tổng quát kết quả khi áp dụng các mô hình học máy hồi quy giá trị glucose cho các đối tượng có tình trạng sức khỏe khác nhau.

Tập dữ liệu thu thập bao gồm: 431 dòng, 19 cột trong đó 18 cột đầu tiên là giá trị của bước sóng được thu bởi AS7265x trải dài từ 410nm – 940nm. Giá trị cột cuối cùng là giá trị glucose được đo bởi máy đo xâm lấn ACCU CHEK Instant.



Hình 4. 1: Phân phối VIS - NIR với giá trị Glucose

Hình 15 biểu diễn mối quan hệ giữa giá trị VIS-NIR và các mức Glucose, với ánh sáng xuyên qua mạnh nhất ở các bước sóng 410 *nm* – 535 *nm*, 705 *nm*, 810 *nm*, và 900 *nm*. Ngoài các bước sóng đặc trưng cho giá trị Glucose, các yếu tố như góc đặt tay và kích thước ngón tay cũng có thể ảnh hưởng đến kết quả đo. Sự thay đổi liên tục của tín hiệu ở các bước sóng khác nhau trong dãi 410 *nm* – 940 *nm* được sử dụng để xây dựng các mô hình hồi quy để dự đoán giá trị glucose.

Giá trị của các bước sóng được tổng hợp lại cụ thể ở bảng 3.

Bảng 4. 1: Phạm vi giá trị các bước sóng.

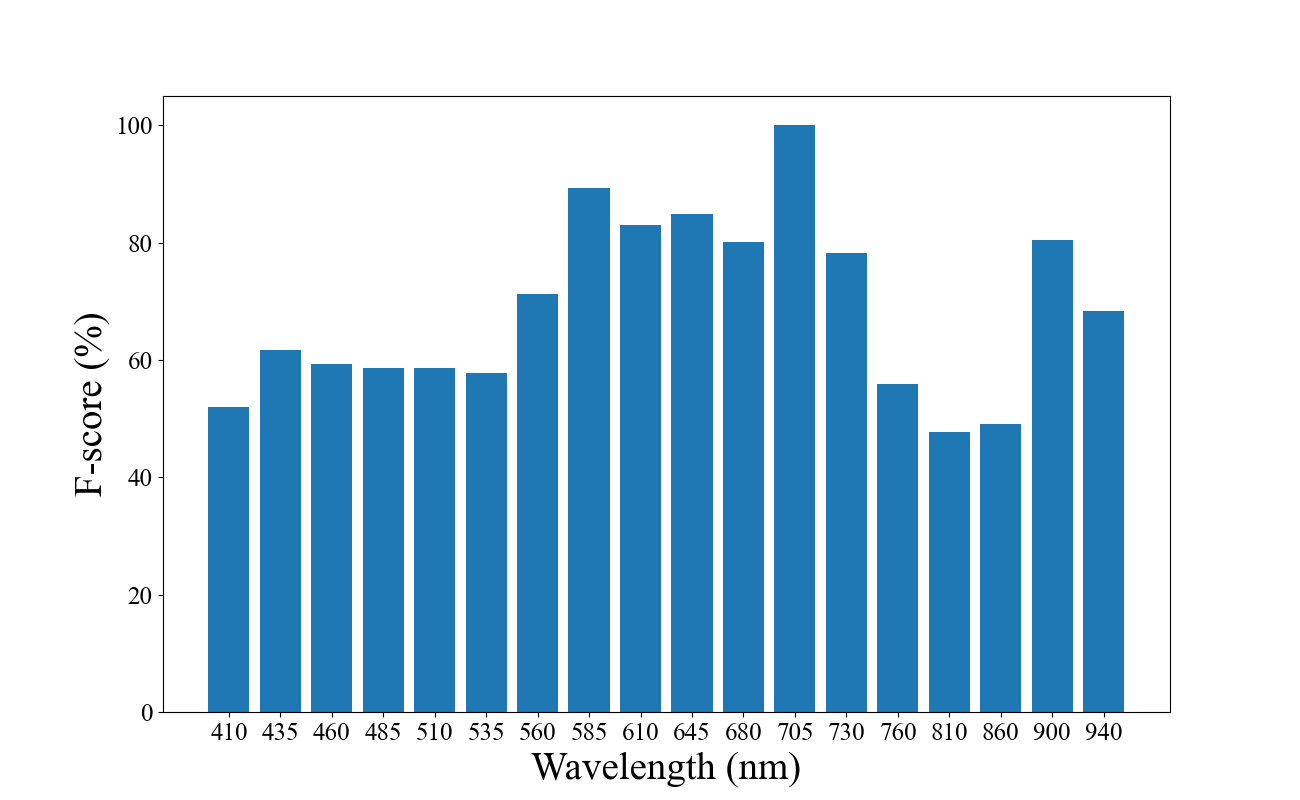
|  |  |  |
| --- | --- | --- |
| STT | Tên thuộc tính | Phạm vi giá trị |
| 1 | 410 *nm* | (0, 4) |
| 2 | 435 *nm* | (1, 21) |
| 3 | 460 *nm* | (2, 38) |
| 4 | 485 *nm* | (1, 36) |
| 5 | 510 *nm* | (2, 42) |
| 6 | 535 *nm* | (1, 31) |
| 7 | 560 *nm* | (0, 3) |
| 8 | 585 *nm* | (1, 7) |
| 9 | 610 *nm* | (2, 15) |
| 10 | 645 *nm* | (2, 15) |
| 11 | 680 *nm* | (1, 10) |
| 12 | 705 *nm* | (8, 40) |
| 13 | 730 *nm* | (2, 17) |
| 14 | 760 *nm* | (1, 20) |
| 15 | 810 *nm* | (1, 28) |
| 16 | 860 *nm* | (1, 23) |
| 17 | 900 *nm* | (4, 26) |
| 18 | 940 *nm* | (2, 15) |
| 19 | Glucose | (79, 176) |

## 4.2. Lựa chọn đặc trưng F-test

Như đã trình bày trong phần trước, kỹ thuật lựa chọn đặc trưng F-test được áp dụng trong nghiên cứu nhằm đánh giá mức độ ảnh hưởng của các biến độc lập đến biến phụ thuộc. Cụ thể, F-test giúp xác định những bước sóng nào có sự tác động mạnh mẽ nhất đến giá trị Glucose, từ đó giúp loại bỏ những bước sóng không có ảnh hưởng đáng kể, giảm thiểu độ phức tạp của mô hình.

Hình 4.29 minh họa giá trị F-test của các bước sóng trong nghiên cứu, cho thấy bước sóng 705 nm có giá trị F-test cao nhất, biểu thị mức ảnh hưởng lớn nhất đến giá trị Glucose. Ngược lại, bước sóng 810 nm có giá trị F-test thấp nhất, cho thấy sự tác động của nó đối với biến phụ thuộc là yếu

Dựa trên bảng kết quả F-test, đề tài đã chọn ra 7 bước sóng có ảnh hưởng mạnh nhất đến giá trị Glucose, bao gồm: 585 nm, 610 nm, 645 nm, 680 nm, 705 nm, 730 nm và 900 nm. Những bước sóng này được lựa chọn để huấn luyện và kiểm thử mô hình, qua đó so sánh hiệu quả của mô hình khi áp dụng kỹ thuật F-test với toàn bộ bước sóng



Hình 4. 29: F-Score của 18 bước sóng

Bảng 12 trình bày các giá trị R² và RMSE khi áp dụng F-test vào quá trình phân tích. Nhìn chung, độ chính xác của các mô hình hồi quy không có sự thay đổi đáng kể khi sử dụng toàn bộ 18 bước sóng, cho thấy rằng việc sử dụng tất cả các bước sóng không mang lại sự cải thiện rõ rệt về mặt hiệu quả dự đoán.

Dựa trên kết quả này, trước khi áp dụng các mô hình hồi quy để dự đoán nồng độ Glucose, chúng ta hoàn toàn có thể giảm số lượng bước sóng đầu vào từ 18 xuống chỉ còn 7 bước sóng. Việc giảm bớt số lượng bước sóng không chỉ giúp giảm thiểu độ phức tạp của mô hình mà còn giúp giảm đáng kể khối lượng tính toán cần thiết. Bên cạnh đó, việc tối ưu hóa số lượng đặc trưng đầu vào này cũng góp phần nâng cao hiệu suất của các mô hình hồi quy, giúp cải thiện tốc độ xử lý và độ chính xác trong việc dự đoán nồng độ Glucose.

## 4.3. Tiền xử lý dữ liệu

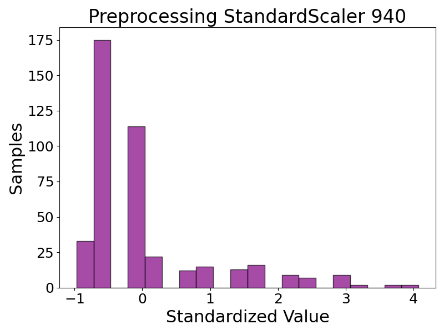
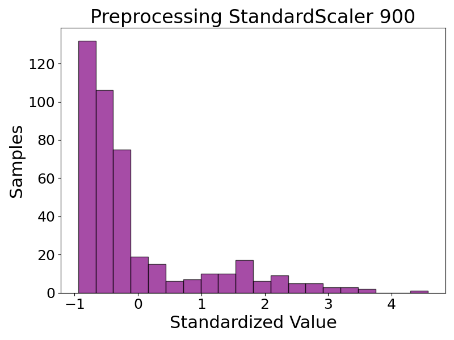
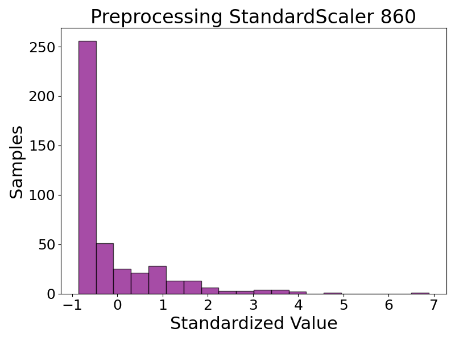
Dữ liệu trong tập huấn luyện sẽ được xử lý bằng hai phương pháp tiền xử lý: Standard Scaler và Multiplicative Scatter Correction (MSC). Nghiên cứu sẽ so sánh hiệu quả của từng phương pháp tiền xử lý với nhau, đồng thời đối chiếu với trường hợp không áp dụng bất kỳ phương pháp nào. Mục tiêu là đánh giá tác động của các kỹ thuật tiền xử lý đối với độ chính xác và khả năng dự đoán của mô hình.

### 4.3.1. Standard Scaler

### 4.3.1.1. Toàn bộ bước sóng

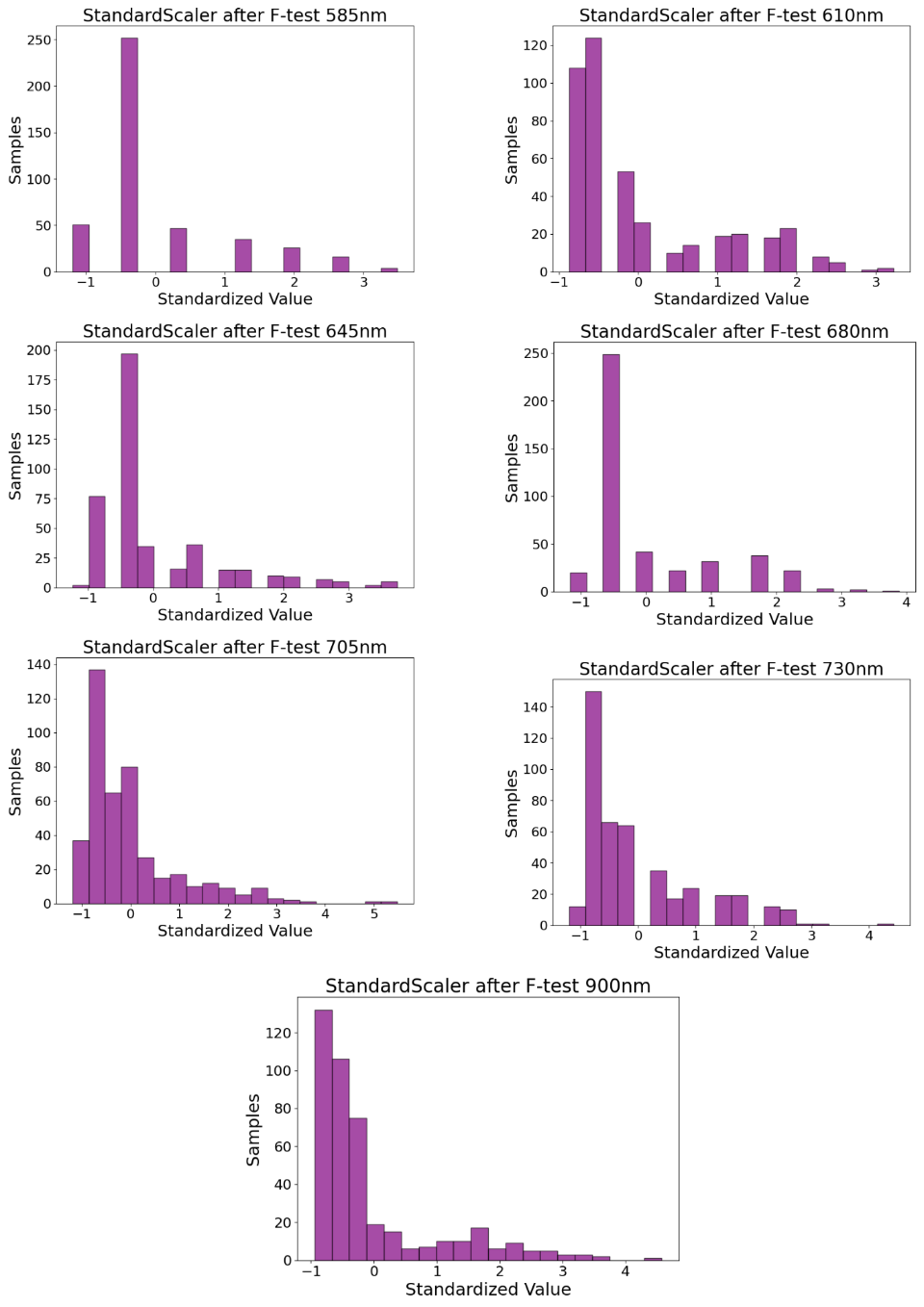
Dữ liệu sẽ được xử lý bằng phương pháp Standard Scaler, như đã trình bày trước đó, để chuẩn hóa các đặc trưng về cùng một thang đo. Hình 17 đến hình 25 trình bày dữ liệu dưới dạng biểu đồ cột, thể hiện các bước sóng từ 410nm đến 940nm trước và sau khi áp dụng phương pháp Standard Scaler.





Hình ...: Giá trị Standard Scaler ở toàn bộ bước sóng

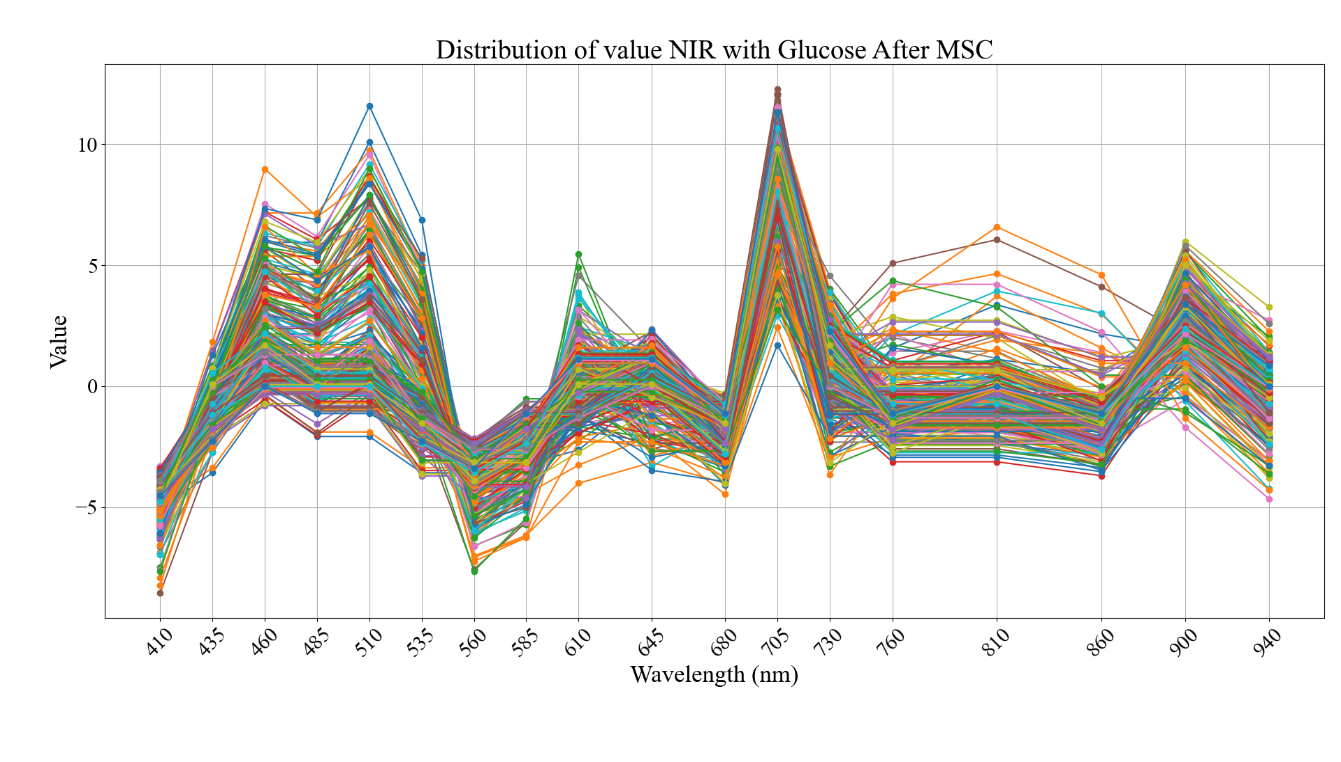
**4.3.1.2. Các bước sóng được chọn bởi F-test**

****

Hình ...: Giá trị Standard Scaler ở 7 bước sóng được chọn bởi F-test

### 4.2.2. Multiplicative Scatter Correction

Khác với StandardScaler, phương pháp Multiplicative Scatter Correction (MSC) thực hiện hiệu chỉnh toàn bộ dữ liệu dựa trên một phổ tham chiếu. Trong nghiên cứu này, phổ tham chiếu được chọn là giá trị trung bình của toàn bộ phổ dữ liệu. Hình 4.12 biểu diễn mối quan hệ giữa các giá trị phổ VIS-NIR và nồng độ glucose sau khi áp dụng MSC.



Hình 4. 12: Xu hướng giá trị VIS – NIR đến Glucose khi sử dụng MSC

## 4.3. Đánh giá mô hình học

### 4.3.1. Độ chính xác phân tích

Bảng 4. 4: Kết quả mô hình học máy ở tập dữ liệu gốc

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Bước sóng | Tập Training | | Tập Testing | |
| R­2 | RMSE | R­2 | RMSE |
| Random Forest | Toàn bộ bước sóng | 0.85 | 7.78 | 0.80 | 8.63 |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* | 0.87 | 7.26 | 0.84 | 7.71 |
| GPR | Toàn bộ bước sóng | 0.97 | 1.45 | 0.76 | 9.68 |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* | 0.90 | 6.18 | 0.79 | 9.02 |
| SVR | Toàn bộ bước sóng | 0.92 | 5.58 | 0.76 | 9.43 |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* | 0.88 | 7.05 | 0.74 | 9.74 |

Bảng 4. 4: Kết quả mô hình học máy ở tập dữ liệu sử dụng Standard Scaler

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Bước sóng | Tập Training | | Tập Testing | |
| R­2 | RMSE | R­2 | RMSE |
| Random Forest | Toàn bộ bước sóng | 0.87 | 7.23 | 0.81 | 8.31 |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* | 0.87 | 7.40 | 0.84 | 7.68 |
| GPR | Toàn bộ bước sóng | 0.99 | 1.45 | 0.79 | 8.64 |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* | 0.90 | 6.18 | 0.78 | 8.97 |
| SVR | Toàn bộ bước sóng | 0.49 | 14.68 | 0.64 | 11.56 |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* | 0.76 | 9.96 | 0.57 | 12.59 |

Bảng 4. 4: Kết quả mô hình học máy ở tập dữ liệu sử dụng Multiplicative Scatter Correction

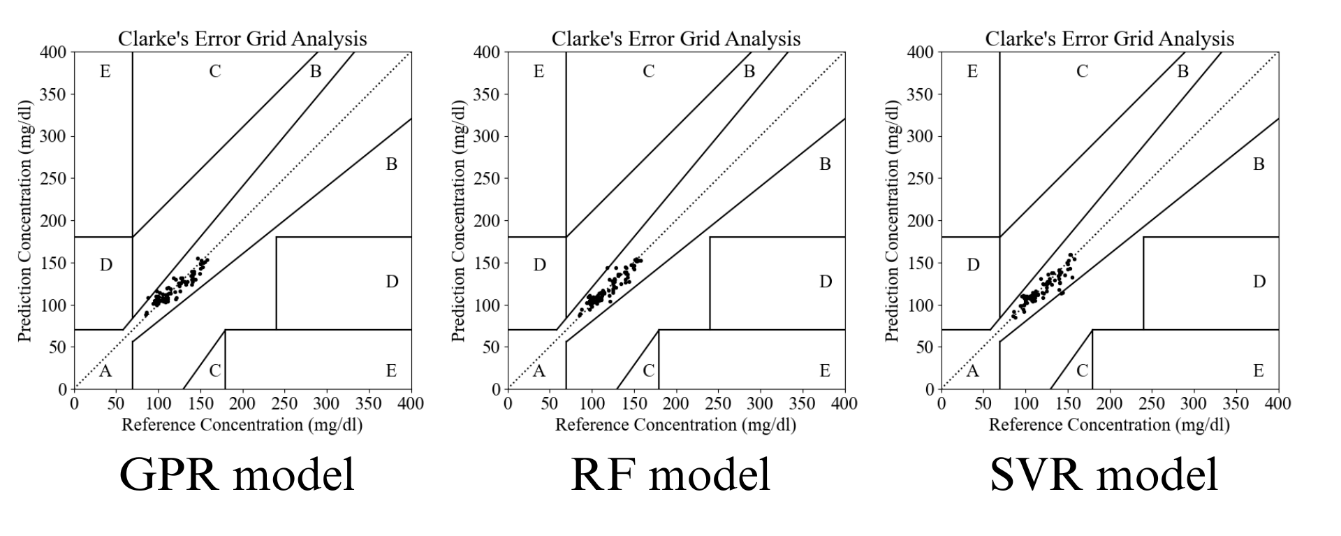
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Bước sóng | Tập Training | | Tập Testing | |
| R­2 | RMSE | R­2 | RMSE |
| Random Forest | Toàn bộ bước sóng | 0.94 | 4.66 | 0.67 | 11.02 |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* | 0.87 | 7.23 | 0.71 | 10.27 |
| GPR | Toàn bộ bước sóng |  |  |  |  |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* |  |  |  |  |
| SVR | Toàn bộ bước sóng |  |  |  |  |
| 585 *nm*, 610 *nm*, 645 *nm*, 680 *nm*, 705 *nm*, 730 *nm* và 900 *nm* |  |  |  |  |

### 4.3.2. Độ chính xác lâm sàn

Bên cạnh các đánh giá phân tích, trong phát triển các phương pháp dự đoán giá trị Glucose, đánh giá dựa trên Clarke Grid Error là cần thiết.

Bảng 4. 5: Giá trị Clarke Grid Error của các mô hình với Standard Scaler

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Vùng | | | | |
| A | B | C | D | E |
| Random Forest | 98.851% | 1.149% | 0% | 0% | 0% |
| GPR | 98.851% | 1.149% | 0% | 0% | 0% |
| SVR | 97.701% | 2.299% | 0% | 0% | 0% |

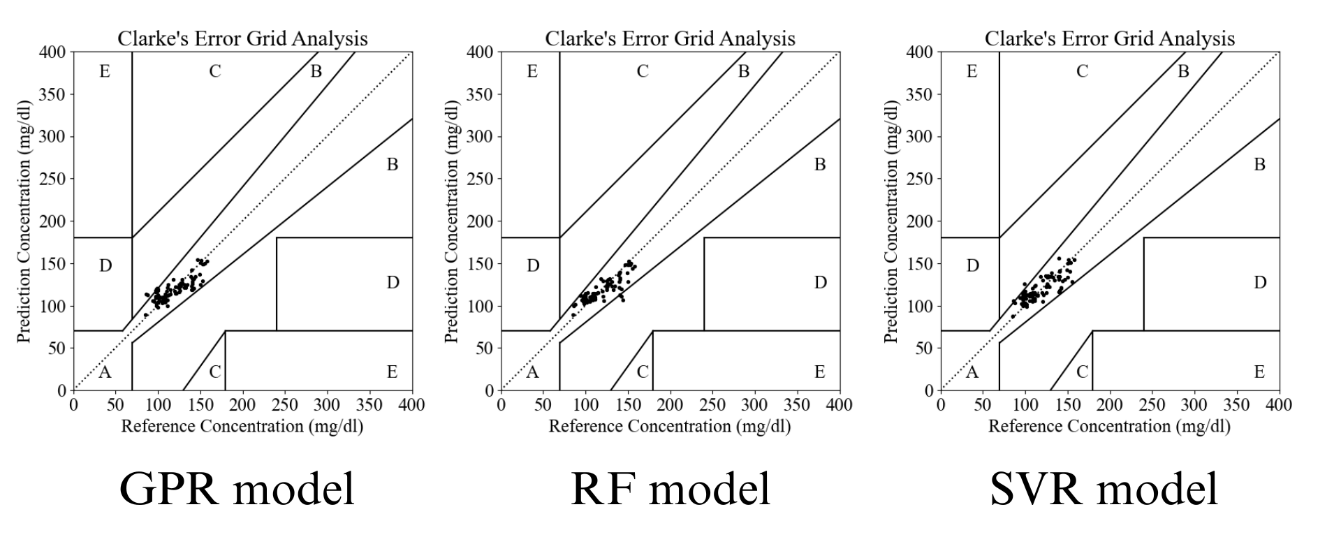


Hình 4. 26: Phân phối trong CGE với phương pháp Standard Scaler

Giá trị phần trăm các điểm dự đoán được thể hiện chi tiết ở bảng 9 và hình 46 Với phương pháp tiền xử lý Standard Scaler, cả ba mô hình đều có các điểm dự đoán chủ yếu tập trung vào vùng A (trên 97%), trong khi không có điểm nào rơi vào các vùng C, D, E. Điều này chứng tỏ các mô hình này biểu diễn tốt giá trị nồng độ glucose, khi được đánh giá thông qua lưới lỗi Clarke.

Bảng 4. 6: Giá trị Clarke Grid Error của các mô hình với MSC

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Vùng | | | | |
| A | B | C | D | E |
| Random Forest | 95.402% | 4.598% | 0% | 0% | 0% |
| GPR | 95.402% | 4.598% | 0% | 0% | 0% |
| SVR | 97.701% | 2.299% | 0% | 0% | 0% |

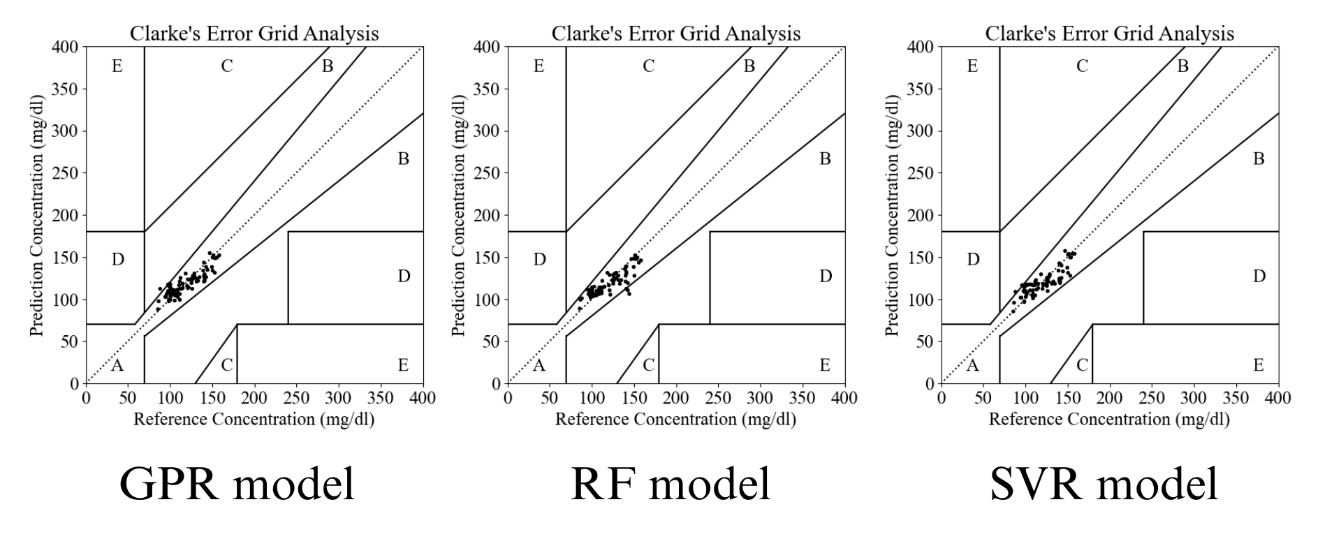


Hình 4. 27: Phân phối trong CGE với phương pháp MSC

Tương tự phương pháp Standard Scaler, các điểm dự đoán chủ yếu nằm ở vùng A và không xuất hiện ở các vùng nguy hiểm C, D, E. Tuy nhiên, tỷ lệ các điểm trong vùng A giảm nhẹ đối với hai mô hình Random Forest và GPR (giảm khoảng 3%), trong khi SVR giữ nguyên tỷ lệ so với Standard Scaler. Dù vậy, xét về tổng quan, cả ba mô hình vẫn biểu diễn tốt các giá trị dự đoán.

Bảng 4. 7: Giá trị Clarke Grid Error của các mô hình khi kết hợp 2 phương pháp tiền xử lý

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mô hình | Vùng | | | | |
| A | B | C | D | E |
| Random Forest | 95.402% | 4.598% | 0% | 0% | 0% |
| GPR | 96.552% | 3.448% | 0% | 0% | 0% |
| SVR | 94.253% | 5.747% | 0% | 0% | 0% |



Hình 4. 28: Phân phối trong CGE với phương pháp MSC và Standard Scaler

Khi kết hợp cả hai phương pháp Standard Scaler và Multiplicative Scatter Correction (MSC), các mô hình vẫn thể hiện khả năng dự đoán tốt giá trị glucose trong lưới lỗi Clarke. Tuy nhiên, độ chính xác có sự giảm nhẹ so với khi chỉ áp dụng riêng lẻ từng phương pháp, dù sự khác biệt này không ảnh hưởng đáng kể đến hiệu quả tổng thể của các mô hình.

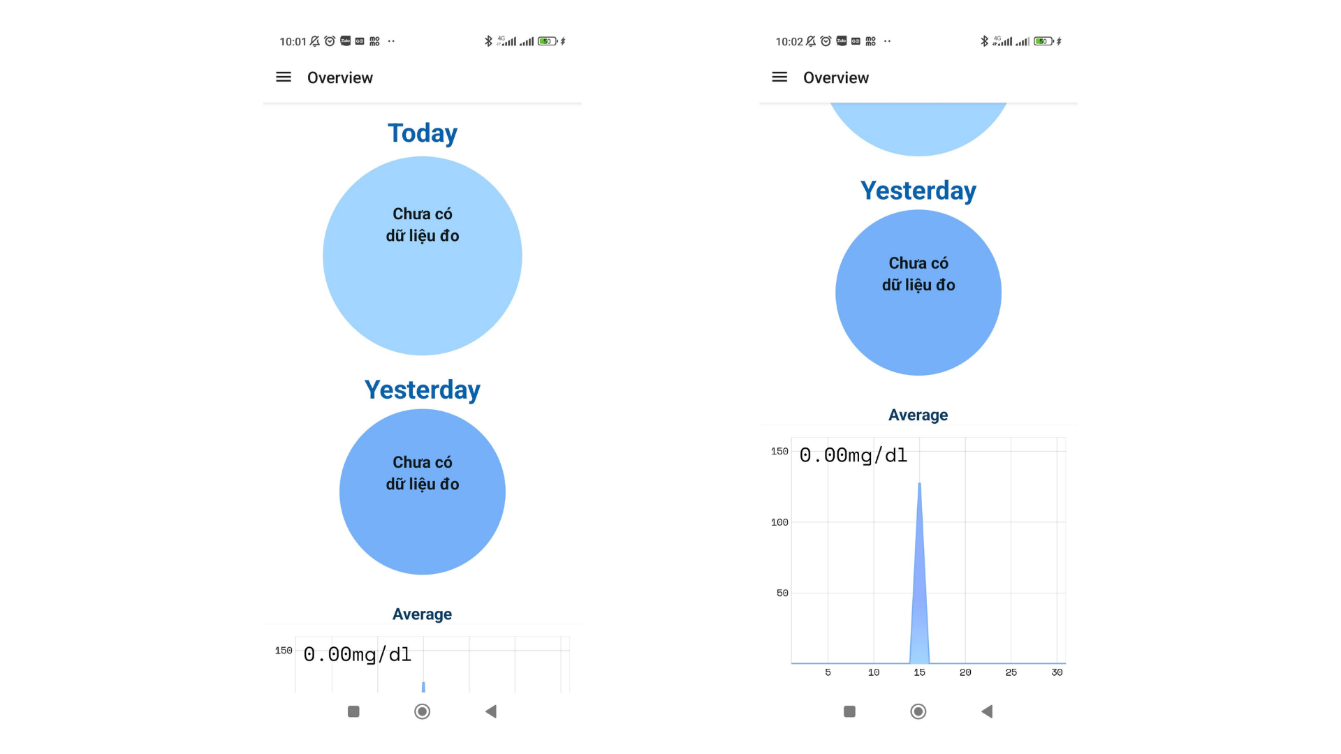
## 4.6. Ứng dụng điện thoại

Hình 51 và hình 52 minh họa biểu tượng và giao diện chính của ứng dụng ProSugr. Giao diện được thiết kế thân thiện và dễ sử dụng, mang lại trải nghiệm trực quan cho người dùng. Trên màn hình chính, ứng dụng hiển thị các thông tin quan trọng như giá trị đường huyết mới nhất được đồng bộ từ Firebase, giá trị đường huyết gần nhất của ngày hôm qua, và biểu đồ theo dõi xu hướng đường huyết trong 30 ngày gần đây.

Giao diện này giúp người dùng nhanh chóng có cái nhìn tổng quan về tình trạng đường huyết của mình trong một khoảng thời gian nhất định. Từ đó, họ có thể đưa ra các nhận định sơ bộ về tình trạng sức khỏe liên quan đến bệnh tiểu đường và có những hành động kịp thời, chẳng hạn như điều chỉnh lối sống hoặc tham khảo ý kiến bác sĩ chuyên khoa khi cần thiết.



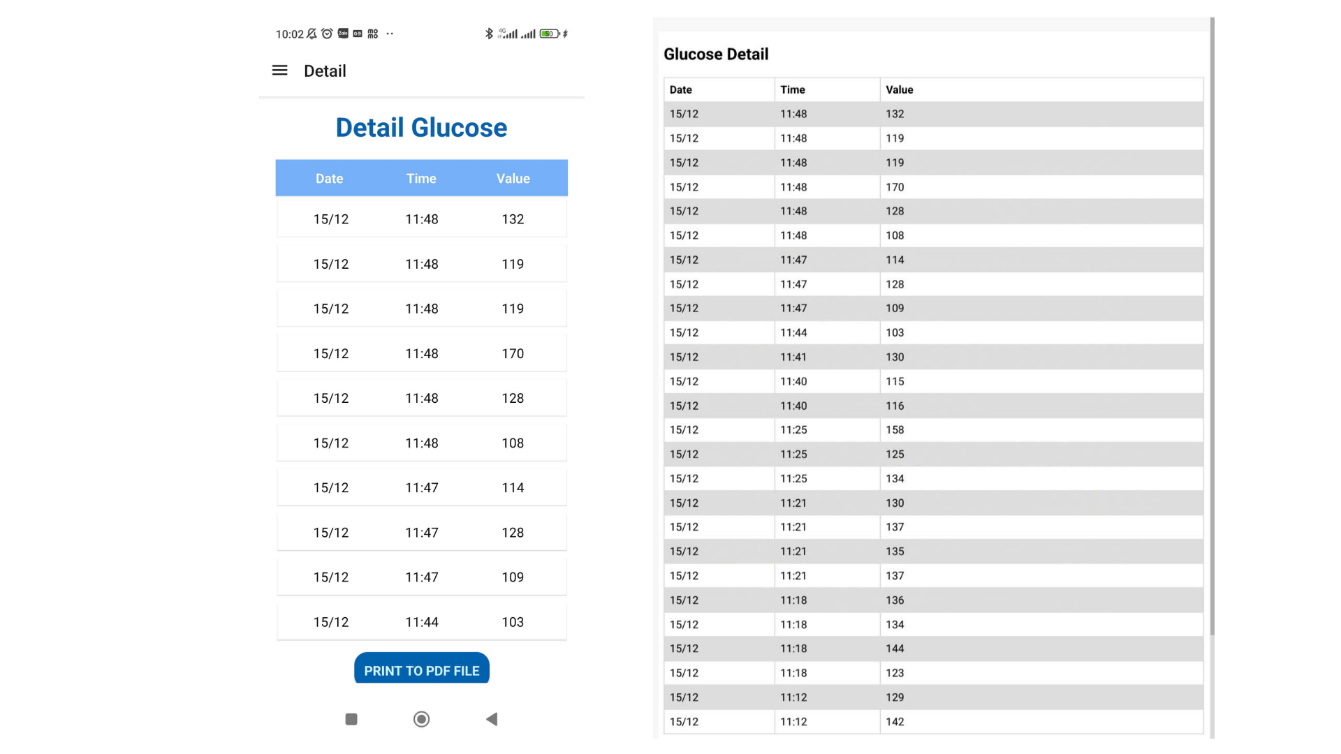
Hình 4. 35: Ứng dụng ProSugr theo dõi nồng độ đường huyết.



Hình 4. 36: Giao diện chính của ứng dụng

Hình 53 biểu diễn mục “Detail” trong ứng dụng, nơi hiển thị thông tin chi tiết về từng lần đo đường huyết. Các chi tiết bao gồm ngày đo, giờ đo chính xác đến từng phút, cùng với giá trị Glucose tương ứng. Mục này được thiết kế để tự động cập nhật khi thiết bị đo đường huyết không xâm lấn gửi dữ liệu mới lên Firebase.

Ngoài ra, ứng dụng cho phép người dùng xuất toàn bộ lịch sử đo lường thành tệp PDF. Tính năng này rất hữu ích cho bệnh nhân cần theo dõi chỉ số đường huyết trong thời gian dài hoặc muốn cung cấp tài liệu cho bác sĩ khi thăm khám, giúp bác sĩ dễ dàng phân tích và đưa ra nhận định chính xác về tình trạng sức khỏe của bệnh nhân.



Hình 4. 37: Ứng dụng ở mục “Detail”

## 4.7. Đánh giá mô hình

Đóng gói của thiết bị đo nồng độ đường huyết không xâm lấn cụ thể ở hình 52. Với kết quả ở trong cả đánh giá phân tích và đánh giá lâm sàn. Mô hình máy học Random Forest được tải vào Raspberry Pi để thực hiện dự đoán giá trị Glucose



Hình 4. 38: Đóng gói thiết bị

Bảng 4. 10 – Độ chênh lệch giá trị đo từ thiết bị và ACCU CHEK

|  |  |  |
| --- | --- | --- |
| **Giá trị đo từ ACCU CHEK**  **(mg/dl)** | **Giá trị đo từ thiết bị**  **(mg/dl)** | **Độ chênh lệch (%)** |
| 113 | 114 | 0.877 |
| 98 | 103 | 4.854 |
| 117 | 121 | 3.305 |
| 107 | 123 | 13.008 |
| 105 | 106 | 0.943 |
| 113 | 112 | 0.892 |
| 91 | 102 | 10.784 |
| 102 | 104 | 1.923 |
| 94 | 110 | 14.552 |
| **Trung bình chênh lệch** | | 5.681 |

Bảng 14 trình bày sự so sánh về độ chênh lệch giữa giá trị đo thực tế từ thiết bị và từ ACCU-CHEK. Mặc dù đề tài đã đạt được những dự đoán tương đối chính xác, nhưng vẫn tồn tại độ sai lệch. Nguyên nhân chính có thể xuất phát từ việc vị trí đặt tay của thiết bị chưa được tối ưu hóa, cùng với sự hạn chế trong tính đa dạng của dữ liệu huấn luyện, khiến các mô hình chưa thể dự đoán chính xác trên những ngón tay mới. Độ chênh lệch trung bình của thiết bị đo được ghi nhận là 5,681%, dựa trên dữ liệu từ 9 tình nguyện viên khác nhau.

# CHƯƠNG 5: KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

## 5.1. Kết luận

Sau một kỳ nghiên cứu và phát triển, đề tài đã hoàn thiện việc thiết kế và xây dựng một hệ thống đo nồng độ đường huyết không xâm lấn dựa trên công nghệ quang phổ cận hồng ngoại và trí tuệ nhân tạo. Đồng thời, công nghệ IoT được tích hợp để hỗ trợ người dùng theo dõi và đánh giá nồng độ glucose trong thời gian dài một cách thuận tiện.

Trong quá trình thực hiện, nhóm nghiên cứu đã thu thập được bộ dữ liệu gồm 431 dòng và 19 cột, tương ứng với 18 bước sóng và giá trị glucose đo lường. Ba mô hình học máy, bao gồm **Support Vector Regression**, **Gaussian Process Regression**, và **Random Forest Regression**, được áp dụng để định lượng glucose. Kết quả cho thấy cả ba mô hình đạt hiệu quả cao với hệ số R² đều trên 80%, trong khi độ chính xác theo lưới lỗi Clarke ở vùng A đạt trên 95%. Sau khi so sánh, **Random Forest Regression** được lựa chọn làm mô hình chính nhờ vào độ chính xác vượt trội và khả năng xử lý tốt các kiểu dữ liệu phi tuyến.

Ngoài ra, thông qua phân tích **F-test**, nghiên cứu đã xác định được 7 bước sóng quan trọng ảnh hưởng lớn đến giá trị glucose, bao gồm: **585 nm, 610 nm, 645 nm, 680 nm, 705 nm, 730 nm, và 900 nm**. Khi áp dụng chỉ 7 bước sóng này để huấn luyện và đánh giá, độ chính xác không chênh lệch đáng kể so với việc sử dụng toàn bộ 18 bước sóng, chứng minh hiệu quả và tính tối ưu của lựa chọn này.

Tuy nhiên, đề tài vẫn gặp một số hạn chế. Số lượng mẫu dữ liệu chưa đủ lớn và đa dạng, ảnh hưởng đến khả năng khái quát hóa của mô hình. Bên cạnh đó, thiết kế đóng gói của thiết bị chưa được tối ưu, dẫn đến sai số đáng kể trong dự đoán nồng độ glucose. Đây là những điểm cần được cải thiện trong các nghiên cứu và phát triển tiếp theo.

**5.1.1. Những điều đạt được**

Đề tài đã thành công đạt được các mục tiêu đề ra bằng việc hiện thực hóa mô hình đo nồng độ đường huyết không xâm lấn, với độ chênh lệch giữa trung bình kết quả đo của thiết bị và ACCU-CHEK dưới 8%. Đồng thời, việc tích hợp công nghệ IoT vào thiết bị đã nâng cao tính ứng dụng thực tế, giúp người dùng dễ dàng theo dõi và quản lý nồng độ đường huyết trong thời gian dài.

Bên cạnh đó, đề tài đã ứng dụng các cơ sở lý thuyết vào ứng dụng thực tiễn, hoàn thiện việc đóng gói phần cứng của thiết bị, đảm bảo tính đồng bộ và tiện lợi. Đề tài cũng đã triển khai và so sánh hiệu quả của ba mô hình học máy, từ đó chọn mô hình tối ưu để tích hợp lên Raspberry Pi, phục vụ việc dự đoán nồng độ glucose.

Thông qua phân tích **F-test** trên toàn bộ tập dữ liệu, đề tài đã đánh giá mức độ ảnh hưởng của từng biến độc lập và xác định được 7 bước sóng quan trọng nhất tác động lớn đến giá trị glucose, bao gồm: **585 nm, 610 nm, 645 nm, 680 nm, 705 nm, 730 nm, và 900 nm.** Những phát hiện này là nền tảng quan trọng để cải tiến và tối ưu hóa hệ thống trong tương lai.

**5.2. Hướng phát triển đề tài**

Hướng phát triển của đề tài được đề xuất tăng thu thập dữ liệu với sự tham gia của nhiều tình nguyện viên có đa dạng đặc điểm và bệnh lý, đặc biệt là bệnh nhân tiểu đường. Việc này bao gồm tăng cường các tình huống như kích thước và vị trí ngón tay để cải thiện độ chính xác trong dự đoán giá trị glucose.

Ngoài ra, phát triển thêm các chức năng hữu ích như cảnh báo khi nồng độ glucose vượt ngưỡng an toàn (quá cao hoặc quá thấp), đồng thời cung cấp các khuyến nghị về lượng đường cần bổ sung thông qua ứng dụng di động. Điều này sẽ nâng cao trải nghiệm người dùng, khi dữ liệu đo từ thiết bị được truyền về ứng dụng một cách nhanh chóng và tiện lợi.

Hoàn thiện hệ thống IoMT (Internet of Medical Things), tạo sự kết nối chặt chẽ giữa bác sĩ, bệnh nhân, và thiết bị đo nhằm nâng cao hiệu quả theo dõi và quản lý sức khỏe từ xa.

Hiện tại, nhóm đang nghiên cứu thêm về tác động của quang phổ cận hồng ngoại (NIR) đến huyết áp, với kế hoạch tích hợp chức năng đo huyết áp vào cùng hệ thống đo đường huyết trên cùng một bộ dữ liệu. Điều này sẽ gia tăng tính tiện lợi và mở rộng khả năng ứng dụng của sản phẩm trong việc theo dõi sức khỏe toàn diện.

TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | P. a. J. A. M. a. M. S. Jain, "Everything you wanted to know about noninvasive glucose measurement and control," *arXiv preprint arXiv:2101.08996,* 2021. |
| [2] | S. P. K. S. P. M. O. K. D. B. S. C. B. A. C. J. M. J. P. M. R. A. W. S. J. S. H. W. Z. P. B. C. K. S. B. E. M. D. Sun H, "IDF Diabetes Atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045.," *Diabetes Res Clin Pract,* 2022. |
| [3] | D. a. M. H. K. Mohajan, "Hyperglycaemia among Diabetes Patients: A Preventive Approach," *Innovation in Science and Technology,* vol. 2, pp. 27-33, 2023. |
| [4] | P. J. Hantzidiamantis, A. O. Awosika and S. L. Lappin., Physiology, Glucose, StatPearls Publishing, 2024. |
| [5] | M. S. a. H. K. S. a. D. S. a. K. S. a. A. E. O. a. R. M. A. a. H. M. A. a. U. M. J. a. P. M.-G. Rahman, "Role of insulin in health and disease: an update," *International journal of molecular sciences,* vol. 22, p. 6403, 2021. |
| [6] | V. P. a. C. W.-Y. Rachim, "Wearable-band type visible-near infrared optical biosensor for non-invasive blood glucose monitoring," *Sensors and Actuators B: Chemical,* vol. 286, pp. 173 - 180, 2019. |
| [7] | M. a. C. D. P. a. R. R. C. a. Q. S. Shokrekhodaei, "Non-invasive glucose monitoring using optical sensor and machine learning techniques for diabetes applications," *IEEE Access,* vol. 9, pp. 73029--73045, 2021. |
| [8] | P. a. M. R. a. J. A. M. Jain, "A precise non-invasive blood glucose measurement system using NIR spectroscopy and Huber’s regression model," *Optical and Quantum Electronics,* vol. 51, p. 51, 2019. |
| [9] | Z. a. R. M. a. L. K. Nie, "Blood glucose prediction based on imagingphotoplethysmography in combination with Machine learning," *Biomedical Signal Processing and Control,* vol. 79, p. 104179, 2023. |
| [10] | S. D. a. B. K. Parag Narkhede, "NIR Based Non-Invasive Blood Glucose," *Indian Journal of Science and Technology,* vol. 9, pp. 1-7, 2016. |
| [11] | J. a. R. A. a. S. V. a. M. B. M. Yadav, "Prospects and limitations of non-invasive blood glucose monitoring using near-infrared spectroscopy," *Biomedical signal processing and control,* vol. 18, pp. 214-227, 2015. |
| [12] | I. a. S. J. Oshina, "Beer--Lambert law for optical tissue diagnostics: current state of the art and the main limitations," *Journal of biomedical optics,* vol. 26, pp. 100901--100901, 2021. |
| [13] | J.-M. a. M. A. a. M. F. Roger, "Preprocessing NIR spectra for Aquaphotomics," *Molecules,* vol. 27, p. 6795, 2022. |
| [14] | M. M. a. M. M. P. a. S. P. K. a. G. K. D. a. S. Z. Ahsan, "Effect of data scaling methods on machine learning algorithms and model performance," *Technologies,* vol. 9, p. 52, 2021. |
| [15] | F. a. O. L. J. Zhang, Support vector regression, Elsevier, 2020. |
| [16] | E. a. S. M. a. K. A. Schulz, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of mathematical psychology,* vol. 85, pp. 1-16, 2018. |
| [17] | L. BREIMAN, "Random Forests," *Machine learning,* vol. 45, pp. 5-32, 2001. |
| [18] | R. a. P. I. R. a. A. S. S. a. S. M. a. N. J. J. Dhanya, "F-test feature selection in Stacking ensemble model for breast cancer prediction," *Procedia Computer Science,* vol. 171, pp. 1561-1570, 2020. |
| [19] | Y. Jung, "Multiple predicting K-fold cross-validation for model selection," *Journal of nonparametric statistics,* vol. 30, pp. 197-215, 2018. |
| [20] | R. a. K. S. a. S. S. a. M. Y. Kasahara, "Noninvasive glucose monitoring using mid-infrared absorption spectroscopy based on a few wavenumbers," *Biomedical optics express,* vol. 9, pp. 289 - 302, 2017. |