

ĐẠI HỌC ĐÀ NẴNG
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA ĐIỆN

ĐỒ ÁN TỐT NGHIỆP
NGÀNH: KỸ THUẬT ĐIỀU KHIỂN & TỰ ĐỘNG HÓA

**NGHIÊN CỨU VÀ TRIỂN KHAI HỆ THỐNG HỆ
THỐNG THU THẬP DỮ LIỆU CỦA ĐMTMN
DỰA TRÊN NỀN TẢNG IoT VÀ AI PHỤC VỤ
CHO VIỆC ĐÁNH GIÁ HIỆU SUẤT**

Giáo viên hướng dẫn: TS. NGUYỄN KIM ÁNH

KS. LÊ MINH HOÀNG HẠC

Sinh viên thực hiện: NGÔ MINH HIẾN - 105200404

ĐỖ ĐOÀN TUẤN HÙNG - 105200407

TRỊNH HOÀNG KHOA - 105200412

Lớp: 20TDHCLC2

Đà Nẵng, 2025

TÓM TẮT

Tên đề tài: *Nghiên cứu và triển khai hệ thống thu thập dữ liệu của ĐMTMN dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất*

Sinh viên thực hiện:

1) Ngô Minh Hiến

Số thẻ SV: 105200404

Lớp: 20TDHCLC2

2) Đỗ Đoàn Tuấn Hùng

Số thẻ SV: 105200407

Lớp: 20TDHCLC2

3) Trịnh Hoàng Khoa

Số thẻ SV: 105200412

Lớp: 20TDHCLC2

Trước bối cảnh khủng hoảng năng lượng và biến đổi khí hậu, việc giám sát và dự báo hiệu suất hệ thống điện mặt trời trở nên ngày càng quan trọng nhằm phát hiện sớm các suy giảm hiệu suất và tối ưu hóa vận hành. Tuy nhiên, các phương pháp truyền thống còn hạn chế do thiếu dữ liệu môi trường đồng bộ và khả năng thích ứng kém với điều kiện thực tế.

Đề án đề xuất một hệ thống giám sát và dự báo hiệu suất điện mặt trời mái nhà dựa trên nền tảng IoT và học máy. Hệ thống tích hợp dữ liệu từ cảm biến môi trường (bức xạ, nhiệt độ, độ ẩm, nhiệt độ mô-đun PV) với dữ liệu điện từ inverter, sử dụng phần cứng nhúng chi phí thấp (Raspberry Pi), phần mềm mã nguồn mở (Node-RED, InfluxDB, Grafana) và truyền dữ liệu qua mạng 4G.

Hai mô hình học máy SVR-RBF và XGBoost được huấn luyện trên dữ liệu thực nghiệm thu thập trong 3 tháng tại một nhà máy ở Đà Nẵng. Kết quả cho thấy cả hai mô hình đều dự báo công suất chính xác, trong đó XGBoost có hiệu suất ổn định hơn.

Hệ thống vận hành ổn định, hỗ trợ trực quan hóa dữ liệu và giám sát từ xa, phù hợp triển khai tại các khu vực khó tiếp cận.

NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP

TT	Họ tên sinh viên	Số thẻ SV	Lớp	Ngành
1	Ngô Minh Hiến	105200404	20TDHCLC2	Kỹ thuật Điều khiển và Tự động hóa
2	Đỗ Đoàn Tuấn Hùng	105200407	20TDHCLC2	Kỹ thuật Điều khiển và Tự động hóa
3	Trịnh Hoàng Khoa	105200412	20TDHCLC2	Kỹ thuật Điều khiển và Tự động hóa

1. Tên đề tài đồ án:

Nghiên cứu và triển khai hệ thống thu thập dữ liệu của ĐMTMN dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất

2. Đề tài thuộc diện: Có ký kết thỏa thuận sở hữu trí tuệ đối với kết quả thực hiện

3. Các số liệu và dữ liệu ban đầu:

Số liệu nhà máy Hòa Vang (Đà Nẵng) và các tài liệu tham khảo.

4. Nội dung các phần thuyết minh và tính toán:

a. Phần chung:

TT	Họ tên sinh viên	Nội dung
1	Ngô Minh Hiến	Tổng quan về ước lượng công suất điện mặt trời mái nhà dựa trên nền tảng IoT
2	Trịnh Hoàng Khoa	
3	Đỗ Đoàn Tuấn Hùng	
4	Ngô Minh Hiến	Kết quả triển khai và phân tích
5	Trịnh Hoàng Khoa	
6	Đỗ Đoàn Tuấn Hùng	
7	Ngô Minh Hiến	Kết luận chung và hướng phát triển
8	Trịnh Hoàng Khoa	
9	Đỗ Đoàn Tuấn Hùng	

b. Phần riêng:

TT	Họ tên sinh viên	Nội dung
1	Ngô Minh Hiến	Xây dựng hệ thống IoT
2	Trịnh Hoàng Khoa	Cơ sở để xây dựng mô hình ước lượng công suất và hệ thống đánh giá hiệu suất

3	Đỗ Đoàn Tuấn Hùng	Xây dựng mô hình học máy ước lượng và tích hợp vào hệ thống IoT
---	-------------------	-----------------------------------------------------------------

5. Các bản vẽ, đồ thị (ghi rõ các loại và kích thước bản vẽ):

a. Phần riêng:

TT	Họ tên sinh viên	Nội dung
1	Ngô Minh Hiến	Kiến trúc hệ thống IoT

6. <i>Họ tên người hướng dẫn:</i>	<i>Phần/ Nội dung:</i>
TS. Nguyễn Kim Ánh	
KS. Lê Minh Hoàng Hạc	

7. Ngày giao nhiệm vụ đồ án: 24/02/2025

8. Ngày hoàn thành đồ án: 15/06/2025

Đà Nẵng, ngày 15 tháng 6 năm 2025

Trưởng Bộ môn Tự động hóa

Người hướng dẫn

TS. Giáp Quang Huy

TS. Nguyễn Kim Ánh

KS. Lê Minh Hoàng Hạc

PHIẾU KIỂM SOÁT TIẾN ĐỘ LÀM ĐỒ ÁN TỐT NGHIỆP

(Phiếu dành cho người hướng dẫn/sinh viên)

Họ tên sinh viên: Ngô Minh Hiến Số thẻ SV : 105200404
Đỗ Đoàn Tuấn Hùng Số thẻ SV : 105200407
Trịnh Hoàng Khoa Số thẻ SV : 105200412

Tên đề tài ĐATN: *Nghiên cứu và triển khai hệ thống thu thập dữ liệu của ĐMTMN dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất*

Họ tên người HD: TS. Nguyễn Kim Ánh Đơn vị: Trường Đại học Bách Khoa – ĐHQĐN

Tuần	Ngày	Khối lượng		GVHD ký tên
		đã thực hiện (%)	tiếp tục thực hiện (%)	
1	12/02/2025	<ul style="list-style-type: none">Nhận đề tàiTìm hiểu nghiên cứu sơ lược về hệ thống	<ul style="list-style-type: none">Xác định giải pháp và hướng tiếp cận	
2	18/02/2025	<ul style="list-style-type: none">Tìm hiểu các thuật toán học máy phù hợpXác định kiến trúc tổng thể	<ul style="list-style-type: none">Đọc, tham khảo tài liệu trên internet	
3	20/02/2025	<ul style="list-style-type: none">Thiết kế kiến trúc hệ thống IoT	<ul style="list-style-type: none">Viết báo cáo chương 1Xác định các thiết bị phần cứng IoT	
4	24/02/2025	Duyệt lần 1: Đánh giá khối lượng hoàn thành 25 % : Được tiếp tục làm ĐATN <input type="checkbox"/> Không tiếp tục thực hiện ĐATN <input type="checkbox"/>		
5	27/02/2025	<ul style="list-style-type: none">Xác định các thiết bị phần cứng IoTXác định các thuật toán học máy khả thi	<ul style="list-style-type: none">Thi công phần cứng hệ thống con IoTThử nghiệm các thuật toán	
6	01/03/2025	<ul style="list-style-type: none">Tích hợp hệ thống IoT vào hệ thống ĐMTMN	<ul style="list-style-type: none">Thu thập dữ liệu cho mô hình học máyXây dựng hệ thống HMI giám sátViết nội dung chương 2 và 4	
7	31/03/2025	Duyệt lần 2: Đánh giá khối lượng hoàn thành 60 % : Được tiếp tục làm ĐATN <input type="checkbox"/> Không tiếp tục thực hiện ĐATN <input type="checkbox"/>		
8	02/04/2025	<ul style="list-style-type: none">Lựa chọn được những thuật toán phù hợp	<ul style="list-style-type: none">Viết nội dung chương 3	
9	24/04/2025	<ul style="list-style-type: none">Thu thập toàn bộ dữ liệu cần thiết cho mô hình học máy	<ul style="list-style-type: none">Huấn luyện các mô hình học máy dựa trên dữ liệu thu thập	

10	19/05/2025	Duyệt lần 3: Đánh giá khối lượng hoàn thành 85 % : Được tiếp tục làm ĐATN <input type="checkbox"/> Không tiếp tục thực hiện ĐATN <input type="checkbox"/>		
11	20/05/2025	- Hiệu chỉnh và tối ưu các mô hình học máy đã chọn	- Tích hợp mô hình học máy vào hệ thống Rashberry	
12	10/05/2025	- Hoàn thành chương 3	- Xây dựng Dashboard quan sát công suất dự báo - Tiếp tục hoàn thiện báo cáo	
13	10/06/2025	- Tối ưu và hoàn thiện hệ thống - Hoàn thành đồ án		

LỜI NÓI ĐẦU VÀ CẢM ƠN

Lời đầu tiên, nhóm sinh viên xin chân thành cảm ơn quý thầy cô Khoa Điện, Trường Đại học Bách khoa – Đại học Đà Nẵng, đã tận tình giảng dạy, truyền đạt những kiến thức nền tảng và chuyên sâu trong suốt quá trình học tập vừa qua.

Đặc biệt, nhóm xin gửi lời cảm ơn sâu sắc đến TS. Nguyễn Kim Ánh, giảng viên hướng dẫn chính, cùng KS. Lê Minh Hoàng Hạc – Công ty TNHH Gia Lực, người đồng hành và hỗ trợ chuyên môn trong suốt quá trình thực hiện đề tài.

Mặc dù nhóm đã nỗ lực hết sức trong việc nghiên cứu và triển khai, nhưng do thời gian và kiến thức còn hạn chế, báo cáo chắc chắn không tránh khỏi thiếu sót. Nhóm rất mong nhận được những góp ý quý báu từ quý thầy cô để hoàn thiện đồ án tốt hơn.

Chúng em xin chân thành cảm ơn!

LỜI CAM ĐOAN LIÊM CHÍNH HỌC THUẬT

Nhóm sinh viên thực hiện đề án tốt nghiệp với tên đề tài: “*Nghiên cứu và triển khai hệ thống thu thập dữ liệu của ĐMTMN dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất*”, xin cam đoan rằng:

1. Toàn bộ nội dung trong đề án là kết quả nghiên cứu và triển khai thực tế của nhóm dưới sự hướng dẫn của TS. Nguyễn Kim Ánh và KS. Lê Minh Hoàng Hạc.
2. Các số liệu, tài liệu và thông tin tham khảo đều được trích dẫn rõ ràng, đúng quy định về đạo đức và liêm chính học thuật.
3. Nhóm hoàn toàn chịu trách nhiệm trước nhà trường nếu có bất kỳ hành vi sao chép, gian lận hay vi phạm học thuật nào được phát hiện.
4. Dữ liệu được sử dụng trong đề án không được sử dụng lại nếu không được sự cho phép của Người hướng dẫn TS. Nguyễn Kim Ánh.

Sinh viên thực hiện 1

Sinh viên thực hiện 2

Sinh viên thực hiện 3

Ngô Minh Hiến

Đỗ Đoàn Tuấn Hùng

Trịnh Hoàng Khoa

MỤC LỤC

TÓM TẮT	i
NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP	ii
PHIẾU KIỂM SOÁT TIẾN ĐỘ LÀM ĐỒ ÁN TỐT NGHIỆP	iv
LỜI NÓI ĐẦU VÀ CẢM ƠN	vii
LỜI CAM ĐOAN LIÊM CHÍNH HỌC THUẬT	viii
MỤC LỤC	ix
DANH SÁCH CÁC BẢNG	xi
DANH SÁCH CÁC HÌNH VẼ	xii
DANH SÁCH CÁC KÝ HIỆU, CHỮ VIẾT TẮT	xiii
MỞ ĐẦU	1
Chương 1: TỔNG QUAN VỀ ƯỚC LƯỢNG CÔNG SUẤT ĐIỆN MẶT TRỜI MÁI NHÀ DỰA TRÊN NỀN TẢNG IOT	4
1.1. Giới thiệu	4
1.2. Các nguyên nhân ảnh hưởng đến hiệu suất của hệ thống ĐMTMN	6
1.2.1. Bức xạ mặt trời và phổ năng lượng	6
1.2.2. Ảnh hưởng của nhiệt độ tế bào và môi trường đến hiệu suất PV	7
1.2.3. Ảnh hưởng của bụi bẩn và khí dung đến hiệu suất hệ thống PV	9
1.2.4. Lão hóa và suy giảm dài hạn của mô-đun PV	11
1.3. Các phương pháp tiếp cận	13
1.4. Đánh giá hiệu suất online	13
1.5. Kết luận	14
Chương 2: CƠ SỞ ĐỀ XÂY DỰNG MÔ HÌNH ƯỚC LƯỢNG CÔNG SUẤT VÀ HỆ THỐNG ĐÁNH GIÁ HIỆU SUẤT	15
2.1. Giới thiệu	15
2.2. Mô hình dự báo hiệu suất điện mặt trời	15
2.2.1. Mô hình vật lý	15
2.2.2. Mô hình học máy	17
2.2.3. Đánh giá và lựa chọn mô hình	21
2.3. Giải pháp thu thập dữ liệu và giám sát	22
2.3.1. Giải pháp thu thập và giám sát cục bộ	22

2.3.2. Giải pháp thu thập và giám sát từ xa	23
2.4. Kết luận	25
Chương 3: XÂY DỰNG MÔ HÌNH HỌC MÁY ƯỚC LƯỢNG HIỆU SUẤT	26
3.1. Giới thiệu	26
3.2. Mô hình học máy (đề xuất)	27
3.2.1. Mô hình SVR-Linear	27
3.2.2. Mô hình XGBoosting	29
3.3. Tiêu chí đánh giá: RMSE, MAE, R^2	30
3.4. Huấn luyện và đánh giá mô hình	32
3.4.1. Kết quả huấn luyện mô hình học máy	32
3.5. Kết luận	36
Chương 4: XÂY DỰNG HỆ THỐNG IOT VÀ TÍCH HỢP MÔ HÌNH ƯỚC LƯỢNG	38
4.1. Giới thiệu	38
4.2. Kiến trúc hệ IoT	38
4.2.1. Kiến trúc phần cứng	38
4.2.2. Phương thức truyền thông và lưu trữ	56
4.3. Tích hợp mô hình AI vào hệ thống IoT	60
4.4. Xây dựng giao diện giám sát theo thời gian thực	62
4.5. Kết luận	63
Chương 5: KẾT QUẢ TRIỂN KHAI VÀ PHÂN TÍCH	64
5.1. Giới thiệu	64
5.2. Hệ thống thực nghiệm	64
5.3. Đánh giá kết quả	69
5.4. Thảo luận kết quả và tính ứng dụng thực tế	69
KẾT LUẬN	70

DANH SÁCH CÁC BẢNG

Bảng 4.1. GPIO và pin của Rashberry Pi 4B	44
Bảng 4.2. Thông số Rashberry Pi 4B	45
Bảng 4.3. Thông số kỹ thuật của Inverter SUN2000-100KTL-M1	47
Bảng 4.4. Thông số kỹ thuật của Smartlogger 3000A	49
Bảng 4.5. Thông số kỹ thuật Router wifi 4G	50
Bảng 4.6. Thông số kỹ thuật của RK330-02	52
Bảng 4.7. Thông số kỹ thuật của RK200-04	54
Bảng 4.8. Thông số kỹ thuật của Meanwell EDR-75-24	55

DANH SÁCH CÁC HÌNH VẼ

Hình 3.1. Biểu đồ phân tán giữa sản lượng thực tế và sản lượng dự báo của mô hình SVR-RBF	33
Hình 3.2. So sánh sản lượng điện đo được và dự báo theo thời gian trong ngày 23/04	33
Hình 3.3. Độ lệch dự báo giữa dự báo và thực tế theo thời gian (Error = Prediction – Actual)	34
Hình 3.4. Biểu đồ phân tán giữa sản lượng điện thực tế và dự báo từ mô hình XGBoost	35
Hình 3.5. So sánh diễn biến công suất đo được và dự báo trong ngày 23/04	36
Hình 3.6. Độ lệch giữa giá trị dự báo và thực tế theo thời gian (Error = Prediction – Actual)	36
Hình 4.1. Sơ đồ khối của giải pháp IoT được đề xuất [44]	40
Hình 4.2. Lưu đồ thuật toán của giải pháp dựa trên IoT được đề xuất [44]	41
Hình 4.3. Raspberry Pi 4B	42
Hình 4.4. Giới thiệu chân Raspberry Pi4	42
Hình 4.5. Cấu hình thực nghiệm của kiến trúc IoT được đề xuất [44]	43
Hình 4.6. Inverter SUN2000-100KTL-M1	46
Hình 4.7. Smartlogger 3000A	49
Hình 4.8. Gateway router WiFi 4G	50
Hình 4.9. RK330-02	52
Hình 4.10. RK200-04	53
Hình 4.11. Bộ nguồn Meanwell EDR-75-24	55
Hình 4.12. Mô tả truyền dữ liệu của Modbus RTU	57
Hình 4.13. Host scan order trên 1 thiết bị cảm biến sử dụng RS485	57
Hình 4.14. Cấu trúc kết nối điện hình	58
Hình 4.15. Cơ chế dự báo của mô hình AI trong hệ thống IoT	62
Hình 5.1. Giao diện Dashboard của hệ thống	65
Hình 5.2. Công suất dự báo của mô hình AI hiển thị trên Dashboard	66
Hình 5.3. Dữ liệu điện thời gian thực thu thập từ hệ thống PV, bao gồm cả nhiệt độ biến tần	68

DANH SÁCH CÁC KÝ HIỆU, CHỮ VIẾT TẮT

CHỮ VIẾT TẮT:

- ABS: Acrylonitrile Butadiene Styrene
- ABC: Artificial Bee Colony
- AC: Dòng điện xoay chiều (Alternating Current)
- AI: Trí tuệ nhân tạo (Artificial Intelligence)
- ANN: Mạng nơ-ron nhân tạo (Artificial Neural Network)
- CBM: Bảo trì theo điều kiện (Condition-Based Maintenance)
- CE: Conformité Européenne
- CNN: Convolutional Neural Network
- CRC-16: Cyclic Redundancy Check
- DC: Dòng điện một chiều (Direct Current)
- DDM: Mô hình hai diode (Double Diode Model)
- DE: Differential Evolution
- DHCP: Dynamic Host Configuration Protocol
- DI: Digital Input
- DIN: Deutsches Institut für Normung
- DMZ: Demilitarized Zone
- DO: Digital Output
- DoS: Denial of Service
- EEPROM: Electrically Erasable Programmable Read-Only Memory
- EN: European Standard
- EVA: Ethylene-vinyl acetate
- FCC: Federal Communications Commission
- 4G: Mạng di động thế hệ thứ tư (Fourth Generation)
- GA: Genetic Algorithm
- GHI: Bức xạ toàn phần trên mặt phẳng ngang (Global Horizontal Irradiance)
- GPIO: General Purpose Input/Output
- GND: Ground
- GPU: Graphics Processing Unit
- IC: Inter-Integrated Circuit
- IoT: Internet vạn vật (Internet of Things)

- IP: Ingress Protection
- IPsec: Internet Protocol Security
- IRENA: International Renewable Energy Agency
- Isc: Dòng ngắn mạch (Short-circuit current)
- JSON: JavaScript Object Notation
- LAN: Local Area Network
- LCD: Liquid Crystal Display
- LED: Diode phát sáng (Light Emitting Diode)
- LID: Light-Induced Degradation
- LoRa: Long Range
- LTE: Long-Term Evolution
- LSTM: Long Short-Term Memory
- L2TP: Layer 2 Tunneling Protocol
- MAE: Sai số tuyệt đối trung bình (Mean Absolute Error)
- MBAP Header: Modbus Application Protocol Header
- MBUS: Meter Bus
- Mbps: Megabits per second
- ML: Học máy (Machine Learning)
- Modbus ASCII: Modbus American Standard Code for Information Interchange
- Modbus RTU: Modbus Remote Terminal Unit
- Modbus TCP: Modbus Transmission Control Protocol
- MPP: Điểm công suất tối đa (Maximum Power Point)
- MPPT: Maximum Power Point Tracking
- MSE: Sai số bình phương trung bình (Mean Square Error)
- MQTT: Message Queuing Telemetry Transport
- NAT: Network Address Translation
- NOCT: Nhiệt độ của mô-đun dưới điều kiện thử nghiệm tiêu chuẩn
- NREL: National Renewable Energy Laboratory
- OS: Operating System
- OTA: Over-The-Air
- PCA: Principal Component Analysis
- PDU: Protocol Data Unit
- PID: Potential-Induced Degradation
- PLC: Programmable Logic Controller

- PPTP: Point-to-Point Tunneling Protocol
- PSO: Particle Swarm Optimization
- PV: Quang điện (Photovoltaic)
- PWM: Pulse Width Modulation
- QoS: Quality of Service
- RBF: Radial Basis Function
- RF: Random Forest
- RH: Relative Humidity
- RJ45: Registered Jack 45
- RMSE: Sai số căn phương bình quân (Root Mean Square Error)
- RoHS: Restriction of Hazardous Substances
- RS-485: Recommended Standard 485
- R²: Hệ số xác định (Coefficient of Determination)
- SCADA: Supervisory Control and Data Acquisition
- SCL: Serial Clock Line
- SD: Secure Digital
- SDA: Serial Data Line
- SDIO: Secure Digital Input/Output
- SDM: Mô hình một diode (Single Diode Model)
- SIM: Subscriber Identity Module
- SCLK: Serial Clock
- SoC: System on Chip
- SPD: Thiết bị chống sét lan truyền (Surge Protective Device)
- SPI: Serial Peripheral Interface
- SPI Firewall: Stateful Packet Inspection Firewall
- SSD: Solid State Drive
- STC: Standard Test Conditions
- SVR: Hồi quy véc-tơ hỗ trợ (Support Vector Regression)
- SVM: Máy vector hỗ trợ (Support Vector Machine)
- TDM: Mô hình ba diode (Triple Diode Model)
- TCP/IP: Transmission Control Protocol/Internet Protocol
- THD: Tổng độ méo hài (Total Harmonic Distortion)
- TLS/SSL: Transport Layer Security/Secure Sockets Layer
- TRRS: Tip-Ring-Ring-Sleeve
- UART: Universal Asynchronous Receiver-Transmitter

- UL: Underwriters Laboratories
- USB: Universal Serial Bus
- UV: Ultraviolet
- Voc: Điện áp mạch hở (Open-circuit voltage)
- VPN: Virtual Private Network
- WAN: Wide Area Network
- WEP: Wired Equivalent Privacy
- WEEE: Waste Electrical and Electronic Equipment
- WiFi: Wireless Fidelity
- WLAN: Wireless Local Area Network
- WPA-PSK: Wi-Fi Protected Access Pre-Shared Key
- WPA2-PSK: Wi-Fi Protected Access 2 Pre-Shared Key
- XGBoost: Extreme Gradient Boosting
- XML: Extensible Markup Language

MỞ ĐẦU

1. Mục đích thực hiện đề tài:

Đề tài này tập trung vào việc nghiên cứu, xây dựng và triển khai một giải pháp giám sát và dự báo hiệu suất cho hệ thống điện mặt trời mái nhà, dựa trên nền tảng Internet vạn vật (IoT) kết hợp với các thuật toán học máy hiện đại như Support Vector Regression (SVR) và Extreme Gradient Boosting (XGBoost). Hệ thống được thiết kế nhằm tích hợp các cảm biến môi trường (bức xạ mặt trời, nhiệt độ không khí, độ ẩm, nhiệt độ bề mặt mô-đun PV) cùng với dữ liệu điện năng được thu thập từ biến tần (inverter), tạo thành một cơ sở dữ liệu đồng bộ, đầy đủ và có độ tin cậy cao phục vụ cho việc phân tích và dự báo hiệu suất hệ thống.

Mục đích chính của đề tài là đáp ứng nhu cầu cấp thiết trong việc giám sát và nâng cao hiệu quả vận hành hệ thống điện mặt trời mái nhà – đặc biệt trong bối cảnh khủng hoảng năng lượng và biến đổi khí hậu toàn cầu, nơi năng lượng tái tạo đang ngày càng giữ vai trò trọng yếu. Bằng cách triển khai cơ chế bảo trì theo điều kiện (Condition-Based Maintenance – CBM) thông qua việc so sánh công suất thực tế với công suất lý tưởng được dự báo, hệ thống đề xuất có khả năng phát hiện sớm các bất thường, từ đó hỗ trợ bảo trì chủ động, giảm thiểu tổn thất hiệu suất và tối ưu hóa vận hành.

Ngoài ra, đề tài còn góp phần làm nền tảng cho các nghiên cứu chuyên sâu hơn về năng lượng tái tạo và ứng dụng IoT – AI, hướng tới xây dựng hệ thống điện mặt trời thông minh, hiệu quả và đáng tin cậy – đặc biệt phù hợp với điều kiện địa lý, kinh tế của Việt Nam, bao gồm các khu vực nông thôn, miền núi và hải đảo.

2. Mục tiêu đề tài:

Đề tài hướng tới việc đạt được các mục tiêu cụ thể sau:

- Thiết kế kiến trúc hệ thống giám sát điện mặt trời dựa trên IoT, với phần cứng chi phí thấp (Raspberry Pi), phần mềm mã nguồn mở (Node-RED, InfluxDB, Grafana) và sử dụng kết nối di động 4G để đảm bảo khả năng giám sát từ xa.

- Xây dựng hệ thống thu thập dữ liệu đồng bộ giữa các thông số môi trường (bức xạ, nhiệt độ, độ ẩm, nhiệt độ PV) và dữ liệu điện năng từ biến tần, tạo ra bộ dữ liệu chất lượng cao phục vụ cho việc phân tích hiệu suất và dự báo.
- Phát triển và đánh giá các mô hình học máy (SVR và XGBoost) trong nhiệm vụ dự báo công suất lý tưởng của hệ thống PV dựa trên điều kiện vận hành thực tế.
- Tích hợp mô hình học máy vào hệ thống IoT, triển khai dự báo công suất theo thời gian thực và hiển thị kết quả trên nền tảng trực quan Grafana.

3. Phạm vi và đối tượng nghiên cứu

Đối tượng nghiên cứu:

- Hệ thống điện mặt trời mái nhà (rooftop PV).
- Các yếu tố môi trường ảnh hưởng đến hiệu suất PV (bức xạ, nhiệt độ, độ ẩm, bụi...).
- Dữ liệu điện: công suất, dòng, điện áp, nhiệt độ biến tần.
- Các thuật toán học máy SVR, XGBoost và công nghệ IoT.
- Phần mềm mã nguồn mở: Node-RED, InfluxDB, Grafana.
- Phần cứng nhúng: Raspberry Pi, SmartLogger, cảm biến công nghiệp.

Phạm vi nghiên cứu:

- Địa điểm: Nhà máy điện mặt trời tại Hòa Vang, Đà Nẵng.
- Dữ liệu: Thu thập 5 phút/lần từ 5h đến 19h hàng ngày, liên tục trong ≈ 52 ngày.
- Mô hình dự báo: Dự báo công suất ngắn hạn (short-term forecasting).
- Phân tích hiệu suất: Đánh giá trực tuyến (online) theo thời gian thực.
- Kỹ thuật: Thiết kế hệ thống IoT, giao tiếp Modbus/MQTT, lưu trữ, trực quan hóa dữ liệu và tích hợp mô hình học máy.

4. Phương pháp nghiên cứu:

Đề tài kết hợp giữa nghiên cứu lý thuyết và thực nghiệm triển khai, cụ thể:

- Nghiên cứu tổng quan: Tìm hiểu các phương pháp ước lượng công suất điện mặt trời (vật lý, thống kê, học máy) và các công nghệ IoT liên quan.

- Thiết kế hệ thống: Xây dựng kiến trúc IoT đa tầng, lựa chọn phần cứng (Raspberry Pi, cảm biến, SmartLogger) và giao thức truyền thông (Modbus RTU/TCP, MQTT).
- Thu thập và xử lý dữ liệu: Triển khai thực nghiệm tại Đà Nẵng, thu thập dữ liệu điện và môi trường trong ≈ 52 ngày, xử lý dữ liệu đầu vào (lọc nhiễu, chuẩn hóa).
- Huấn luyện và đánh giá mô hình AI: Áp dụng SVR và XGBoost để dự báo công suất, đánh giá mô hình bằng RMSE, MAE, R^2 .
- Tích hợp hệ thống: Cài đặt Node-RED, InfluxDB, Grafana; triển khai mô hình dự báo thời gian thực và giao diện giám sát trực quan.
- Phân tích và kết luận: Đánh giá hiệu quả hệ thống, khả năng ứng dụng và đề xuất hướng phát triển trong tương lai.

5. Cấu trúc của đồ án tốt nghiệp:

Đồ án tốt nghiệp bao gồm các phần chính sau:

- TÓM TẮT, NHIỆM VỤ ĐỒ ÁN, CAM KẾT HỌC THUẬT
- MỞ ĐẦU: Mục đích, mục tiêu, phạm vi, phương pháp, cấu trúc đồ án
- 6 chương nội dung:
 1. Tổng quan về ước lượng công suất PV dựa trên IoT
 2. Cơ sở mô hình hóa và hệ thống giám sát hiệu suất
 3. Xây dựng mô hình học máy dự báo hiệu suất
 4. Thiết kế hệ thống IoT và tích hợp mô hình
 5. Kết quả thực nghiệm và phân tích đánh giá
 6. Kết luận và hướng phát triển
- PHỤ LỤC, TÀI LIỆU THAM KHẢO

Chương 1: TỔNG QUAN VỀ ƯỚC LƯỢNG CÔNG SUẤT ĐIỆN MẶT TRỜI MÁI NHÀ DỰA TRÊN NỀN TẢNG IOT

1.1. Giới thiệu

Trong bối cảnh khủng hoảng năng lượng và biến đổi khí hậu toàn cầu, việc tìm kiếm và phát triển các nguồn năng lượng tái tạo, đặc biệt là năng lượng mặt trời, đang trở thành ưu tiên hàng đầu của nhiều quốc gia trên thế giới. Theo báo cáo của IRENA [1], đến năm 2030, năng lượng mặt trời được dự báo sẽ chiếm đến 49% tổng công suất điện tái tạo toàn cầu với mức công suất ước tính đạt 5.654 GW. Điều này phản ánh rõ sự chuyển đổi mạnh mẽ từ năng lượng hóa thạch sang năng lượng sạch trong cơ cấu năng lượng toàn cầu.

Cùng với sự phát triển nhanh chóng của điện mặt trời, hiệu suất các mô-đun quang điện (PV) đã liên tục được cải thiện, trong khi chi phí sản xuất ngày càng giảm đáng kể nhờ những tiến bộ về công nghệ và quy mô sản xuất [2]. Sự cải tiến này đã tạo điều kiện thuận lợi cho việc triển khai rộng rãi điện mặt trời từ quy mô lớn, công nghiệp cho tới các hệ thống hộ gia đình, đặc biệt tại các quốc gia như Trung Quốc, nơi công suất lắp đặt PV có thể lên đến 1.300 GW vào năm 2050 [2].

Tuy nhiên, song hành cùng sự mở rộng nhanh chóng của các hệ thống điện mặt trời, nhu cầu giám sát hiệu quả hoạt động và dự báo công suất nhằm phát hiện sớm các hiện tượng suy giảm hiệu suất đã trở nên cấp thiết. Trong thực tế vận hành, việc theo dõi chỉ các chỉ số hiệu suất điện đơn thuần là chưa đủ để hỗ trợ ra quyết định bảo trì tối ưu. Thay vào đó, chiến lược bảo trì theo điều kiện (Condition-Based Maintenance – CBM) yêu cầu sự so sánh liên tục giữa công suất thực tế và công suất lý tưởng của hệ thống trong các điều kiện vận hành tối ưu. Để đạt được điều này, khả năng dự báo công suất chính xác dựa trên các điều kiện vận hành thực tế đóng vai trò then chốt.

Hiện nay, việc dự báo và ước lượng công suất hệ thống PV thường dựa vào các mô hình vật lý hoặc toán học [3][4][5]. Tuy nhiên, các mô hình này còn nhiều hạn chế, như thiếu dữ liệu thực tế từ cảm biến, khó phản ánh chính xác ảnh hưởng của các yếu tố môi trường như biến đổi bức xạ, nhiệt độ, độ ẩm và bụi bẩn. Do đó, việc sử dụng các thuật toán học máy (Machine Learning – ML) đã trở thành xu hướng được quan tâm rộng rãi nhờ khả

năng xử lý hiệu quả các mối quan hệ phi tuyến giữa các tham số đầu vào và công suất đầu ra [6].

Mặc dù học máy mang lại nhiều triển vọng, hiệu quả của mô hình dự báo công suất PV phụ thuộc trực tiếp vào chất lượng và độ đầy đủ của dữ liệu đầu vào. Tuy nhiên, hầu hết các hệ thống giám sát hiện nay vẫn chỉ tập trung vào các thông số điện như dòng điện, điện áp DC và công suất đầu ra từ inverter, mà ít quan tâm đến các yếu tố môi trường như bức xạ mặt trời, nhiệt độ bề mặt pin, độ ẩm môi trường – những yếu tố tác động trực tiếp và mạnh mẽ đến hiệu suất PV. Sự thiếu hụt hoặc mất đồng bộ trong dữ liệu môi trường và điện đã gây khó khăn trong việc phân tích chi tiết và dự báo chính xác hiệu suất hệ thống.

Hơn nữa, phần lớn các giải pháp hiện hữu sử dụng phần mềm đóng hoặc nền tảng đám mây cố định, khiến việc tùy biến chức năng và mở rộng tích hợp dữ liệu bị hạn chế đáng kể. Việc thu thập đồng thời các thông số môi trường và điện năng trong các hệ thống này thường đòi hỏi thêm các cảm biến và bộ chuyển đổi (ADC), dẫn đến chi phí triển khai cao và giảm tính linh hoạt của hệ thống.

Trước những hạn chế trên, đề án này tập trung vào việc nghiên cứu, xây dựng và triển khai một giải pháp giám sát và dự báo hiệu suất cho hệ thống điện mặt trời mái nhà dựa trên nền tảng Internet vạn vật (IoT) và ứng dụng các thuật toán học máy tiên tiến như Support Vector Regression (SVR) và Extreme Gradient Boosting (XGBoost). Hệ thống sẽ tích hợp đầy đủ các cảm biến đo môi trường (bức xạ, nhiệt độ, độ ẩm, nhiệt độ tế bào PV) với dữ liệu điện thu thập từ inverter, nhằm cung cấp một nền tảng dữ liệu đồng bộ, đầy đủ và đáng tin cậy cho việc phân tích và dự báo hiệu suất.

Kiến trúc hệ thống được xây dựng dựa trên phần cứng nhúng giá rẻ (Raspberry Pi), phần mềm mã nguồn mở (Node-RED, InfluxDB, Grafana), và sử dụng kết nối mạng di động 4G LTE để đảm bảo khả năng giám sát từ xa hiệu quả, đặc biệt phù hợp với điều kiện triển khai tại các khu vực vùng sâu, vùng xa. Đây là nền tảng quan trọng để tiến tới áp dụng phương pháp bảo trì theo điều kiện, giúp phát hiện sớm và ngăn ngừa hiệu quả các vấn đề kỹ thuật tiềm ẩn trong hệ thống điện mặt trời mái nhà tại Việt Nam.

1.2. Các nguyên nhân ảnh hưởng đến hiệu suất của hệ thống ĐMTMN

1.2.1. Bức xạ mặt trời và phổ năng lượng

Bức xạ mặt trời (solar irradiance) là yếu tố đầu vào trực tiếp và có ảnh hưởng quan trọng nhất đến công suất đầu ra tức thời của hệ thống PV. Cường độ bức xạ càng cao thì tổng năng lượng khả dụng cho quá trình chuyển đổi quang điện càng lớn. Tuy nhiên, mối quan hệ giữa bức xạ và hiệu suất mô-đun không hoàn toàn tuyến tính. Hiệu suất tế bào PV phụ thuộc không chỉ vào cường độ (W/m^2) mà còn vào phổ quang (spectral distribution) [7], góc tới (angle of incidence) [8], và độ đồng nhất theo thời gian của bức xạ.

Về mặt lý thuyết, mô-đun PV hoạt động hiệu quả nhất khi bức xạ mặt trời tới bề mặt đạt giá trị tương đương với điều kiện STC (1000 W/m^2 , phổ AM 1.5). Tuy nhiên, trong môi trường thực tế, cường độ bức xạ dao động đáng kể theo thời gian trong ngày, theo mùa và đặc biệt là theo điều kiện khí tượng tức thời như mây, sương mù, và bụi khí quyển. Sự thay đổi này khiến mô-đun PV thường xuyên vận hành dưới ngưỡng công suất danh định, dẫn đến sai số lớn khi sử dụng dữ liệu STC để dự báo sản lượng năng lượng.

Bức xạ thường được biểu diễn qua hai thông số chính: Bức xạ toàn phần trên mặt phẳng ngang (*Global Horizontal Irradiance – GHI*) và Bức xạ trực tiếp vuông góc với mặt trời (*Direct Normal Irradiance – DNI*). GHI bao gồm cả thành phần trực tiếp và khuếch tán của bức xạ mặt trời trên mặt phẳng ngang, trong khi DNI chỉ tính thành phần bức xạ trực tiếp tới bề mặt vuông góc với tia sáng mặt trời.

Tùy thuộc vào cấu hình hệ thống, mỗi dạng bức xạ lại đóng vai trò khác nhau trong việc quyết định hiệu suất vận hành. Cụ thể, các hệ thống PV cố định (fixed-tilt) chủ yếu phụ thuộc vào GHI, trong khi các hệ thống có cơ chế theo dõi mặt trời (solar tracking) tận dụng tốt hơn bức xạ DNI do khả năng tự điều chỉnh góc nghiêng theo vị trí mặt trời trong ngày.

Bên cạnh đó, vị trí địa lý và thời điểm trong năm là những yếu tố quyết định đến sự thay đổi theo không gian và thời gian của bức xạ mặt trời. Tại các khu vực gần xích đạo (vĩ độ từ -30° đến $+30^\circ$), bức xạ mặt trời thường có cường độ cao và ổn định hơn so với các vùng cận cực, nơi góc tới mặt trời lớn và thời gian chiếu sáng trong ngày biến động mạnh theo mùa. Ngoài ra, vào mùa hè, mặt trời lên cao hơn và thời gian chiếu sáng trong ngày kéo dài, góp phần làm tăng tổng lượng bức xạ nhận được.

Cũng cần lưu ý rằng quỹ đạo elip của Trái đất quanh Mặt trời gây ra những biến động nhỏ theo chu kỳ hàng năm về khoảng cách giữa hai thiên thể. Vào thời điểm cận nhật (perihelion), Trái đất gần Mặt trời hơn, dẫn đến lượng bức xạ nhận được cao hơn một chút so với thời điểm viễn nhật (aphelion). Tuy nhiên, sự chênh lệch này là tương đối nhỏ (khoảng 3%) và thường không ảnh hưởng đáng kể đến tính toán tổng thể trong các mô hình ước lượng công suất PV.

Từ các phân tích trên, có thể thấy rằng bức xạ mặt trời là một thông số biến thiên mạnh theo cả không gian và thời gian, và việc mô hình hóa chính xác đại lượng này là điều kiện tiên quyết để dự báo hiệu suất hệ thống điện mặt trời một cách tin cậy, đặc biệt trong các ứng dụng học máy. Các mô hình dự báo công suất PV hiện nay thường sử dụng GHI (hoặc DNI nếu có hệ tracking) làm đầu vào chính để huấn luyện thuật toán, cùng với các biến số thời tiết khác như nhiệt độ, độ ẩm hay tốc độ gió.

Với các hệ thống PV có cơ chế theo dõi mặt trời, thành phần trực tiếp (DNI) có vai trò chủ đạo, trong khi hệ thống cố định góc nghiêng chủ yếu khai thác bức xạ toàn phần trên mặt phẳng ngang (GHI). Ở các khu vực có tỷ lệ mây cao như miền Bắc Việt Nam, thành phần khuếch tán có thể chiếm tới 50–60% tổng GHI, làm giảm hiệu quả của các hệ thống có thiết kế tối ưu cho bức xạ trực tiếp.

Ngoài ra, sự thay đổi bức xạ theo thời gian – đặc biệt là sự không ổn định trong quy mô phút (intra-hour variability) – gây ra hiện tượng dao động công suất (power ramping), ảnh hưởng đến cả chất lượng điện lưới lẫn hiệu quả hoạt động của bộ biến tần. Điều này đòi hỏi các mô hình dự báo bức xạ phải có độ phân giải cao và tích hợp các yếu tố như tốc độ chuyển động mây, sự thay đổi vị trí mặt trời và ảnh hưởng địa hình (terrain-induced shading).

1.2.2. Ảnh hưởng của nhiệt độ tế bào và môi trường đến hiệu suất PV

Nhiệt độ tế bào quang điện (PV cell temperature) là một trong những tham số môi trường có tác động trực tiếp và đáng kể nhất đến hiệu suất chuyển đổi năng lượng của mô-đun PV. Đối với tế bào quang điện, nhiệt độ tăng lại gây suy giảm hiệu quả hoạt động. Nhiệt độ là một trong những yếu tố môi trường ảnh hưởng rõ rệt và tiêu cực đến hiệu suất của cell pin mặt trời. Khi nhiệt độ cell PV tăng, hiệu suất chuyển đổi quang điện có xu

hướng giảm do sự suy giảm điện áp mạch hở (V_{oc}) và hệ số điền đầy (Fill Factor), trong khi dòng ngắn mạch (I_{sc}) chỉ tăng nhẹ [9][10].

$$P_m = (FF) I_{sc} V_{sc} \quad (1.1)$$

Nhiệt độ tế bào không trùng với nhiệt độ không khí (ambient temperature), mà cao hơn đáng kể do hấp thụ bức xạ mặt trời. Mối liên hệ này được thể hiện thông qua công thức tính gần đúng:

$$T_{cell} = T_{air} + \frac{NOCT - 20}{800} E \quad (1.2)$$

Với: T_{cell} là nhiệt độ tế bào quang điện, T_{air} là nhiệt độ môi trường, $NOCT$ là nhiệt độ của mô-đun dưới điều kiện thử nghiệm tiêu chuẩn, và suất bức xạ mặt trời tính bằng W/m^2 . Nhiệt độ $NOCT$ trung bình đối với các mô-đun silicon tinh thể (cSi) thương mại là $46,5^\circ C$, với khoảng tứ phân vị nằm trong khoảng từ $45,5$ đến $47,5^\circ C$,

Công thức này cho thấy rằng vào buổi trưa với bức xạ $1000 W/m^2$ và nhiệt độ không khí $35^\circ C$, nhiệt độ tế bào có thể lên tới $65-70^\circ C$. Điều này làm rõ rằng trong nhiều giờ trong ngày, mô-đun PV hoạt động trong trạng thái không tối ưu về nhiệt, dẫn đến tổn thất hiệu suất liên tục.

Hơn nữa, nhiệt độ cao kéo dài không chỉ ảnh hưởng hiệu suất tức thời mà còn làm tăng tốc độ lão hóa (degradation rate) của các thành phần vật liệu [11]. Nhiệt độ cao có ảnh hưởng đáng kể đến độ bám dính giữa lớp encapsulant (EVA) và kính trong mô-đun quang điện, từ đó dẫn đến hiện tượng tách lớp (delamination). Theo nghiên cứu được trích dẫn trong Oreski et al. (2019)[12], độ bám dính (adhesive energy G_c) của giao diện EVA/kính giảm gần như tuyến tính từ $25^\circ C$ đến $50^\circ C$ và giảm mạnh khi nhiệt độ vượt quá $60^\circ C$, do lúc này EVA trở nên mềm đáng kể. Ở $70^\circ C$, lực bóc (peel force) của cả EVA và ionomer giảm hơn 10 lần so với điều kiện bình thường. Điều này chứng tỏ rằng nhiệt độ là một yếu tố then chốt làm suy yếu kết cấu cơ học của encapsulant và thúc đẩy quá trình delamination trong điều kiện vận hành thực tế. Một trong những hiện tượng được ghi nhận rộng rãi trong phân tích lão hóa vật liệu là gia tốc Arrhenius – theo đó, cứ mỗi $10^\circ C$ tăng thêm trong nhiệt độ vận hành, tốc độ suy giảm hiệu suất có thể tăng gấp đôi. Hiện tượng này phản ánh đặc trưng của quá trình suy giảm do nhiệt kích hoạt, thường gặp ở các vật liệu polymer như EVA, Tedlar và cả lớp kính trong mô-đun quang điện. Trong nghiên cứu của Rajput và

cộng sự (2017) [13], các tác giả đã áp dụng mô hình Arrhenius để đánh giá hệ số gia tốc suy giảm (acceleration factor) của các vật liệu trong điều kiện khí hậu hỗn hợp tại Ấn Độ. Kết quả cho thấy khi sử dụng giá trị năng lượng hoạt hóa 50.1 kJ/mol, nhiệt độ tương đương (T_{eq}) và nhiệt độ hiệu dụng (T_{eff}) có thể đạt đến 49.4°C và 54.95°C tương ứng, dẫn đến gia tăng đáng kể tốc độ suy giảm vật liệu. Nghiên cứu khẳng định rằng nhiệt độ cao kéo dài có thể làm tăng đáng kể tốc độ lão hóa, đặc biệt trong các điều kiện chiếu xạ mạnh và nhiệt độ môi trường cao.

Trong các hệ thống điện mặt trời quy mô lớn, việc kiểm soát nhiệt độ vận hành của tế bào PV đóng vai trò quan trọng đối với hiệu suất tổng thể. Nhiệt độ tế bào tăng cao có thể làm giảm điện áp mạch hở (V_{oc}) và hiệu suất chuyển đổi năng lượng, với mức tổn hao điển hình khoảng 0.5% cho mỗi 1°C tăng thêm [14]. Một số nghiên cứu đã chứng minh rằng giải pháp làm mát bằng nước, như phun nước phía trước mô-đun, có thể cải thiện đáng kể hiệu suất hệ thống [14]. Ngoài ra, cấu trúc cơ học của hệ thống như kiểu lắp đặt, độ thông thoáng phía sau mô-đun cũng có ảnh hưởng đến nhiệt độ vận hành và qua đó gián tiếp đến hiệu suất [14].

Tóm lại, nhiệt độ là yếu tố ảnh hưởng kép – vừa gây tổn thất công suất tức thời, vừa ảnh hưởng lâu dài đến độ tin cậy của hệ thống PV. Việc tích hợp cảm biến đo nhiệt độ tế bào, kết hợp mô hình hóa động lực nhiệt và dữ liệu thời tiết theo thời gian thực là hướng đi bắt buộc nếu muốn giám sát và tối ưu hóa hiệu suất vận hành trong điều kiện khí hậu khắc nghiệt như ở Việt Nam.

1.2.3. Ảnh hưởng của bụi bẩn và khí dung đến hiệu suất hệ thống PV

Soiling – hiện tượng bụi và các hạt khí dung (aerosols) tích tụ trên bề mặt mô-đun PV – là một trong những nguyên nhân hàng đầu dẫn đến suy giảm hiệu suất quang điện nhưng thường bị đánh giá thấp trong giai đoạn thiết kế và dự báo vận hành hệ thống. Theo nghiên cứu thực nghiệm của Hussain et al. (2017) [15], hiệu suất của mô-đun quang điện có thể giảm từ 20% đến hơn 70% khi có sự tích tụ của các loại bụi môi trường khác nhau trên bề mặt. Loại bụi có kích thước hạt nhỏ như vỏ trấu (~10 μm) được xác định là gây tổn thất công suất nghiêm trọng nhất, với mức giảm lên đến 75.88% ở mức chiếu xạ 750 W/m^2 . Nghiên cứu chỉ ra rằng kích thước hạt bụi càng nhỏ thì khả năng chắn bức xạ mặt trời càng lớn, từ đó làm giảm đáng kể công suất đầu ra của mô-đun PV. Ngoài ra, một lớp bụi mỏng

5g cũng có thể làm giảm công suất từ 13% đến 20%, cho thấy ảnh hưởng rõ rệt ngay cả ở mức độ ô nhiễm nhẹ

Cơ chế ảnh hưởng của bụi đến hiệu suất PV rất đa chiều. Trước hết, các hạt bụi trên bề mặt mô-đun làm suy giảm truyền dẫn quang (optical transmittance) [16] thông qua ba quá trình vật lý chính: (i) tán xạ (scattering), (ii) hấp thụ (absorption), và (iii) phản xạ ngược (back reflection). Sự hiện diện của lớp bụi làm giảm lượng ánh sáng tới tế bào bán dẫn, kéo theo giảm dòng điện ngắn mạch (I_{sc}) – đại lượng phụ thuộc trực tiếp vào cường độ bức xạ khả dụng. Điện áp hở mạch (V_{oc}) ít bị ảnh hưởng, dẫn đến dịch chuyển điểm công suất tối đa (MPP) về phía dòng điện thấp hơn, gây tổn thất tổng thể công suất đầu ra.

Thứ hai, bụi thường không phân bố đều trên bề mặt mô-đun mà tích tụ thành các vùng cục bộ, gây ra hiện tượng che bóng từng phần (partial shading) trên một số tế bào quang điện [17]. Khi một tế bào bị che khuất không hoàn toàn, phần không bị che sẽ rơi vào trạng thái phân cực ngược, trong khi vẫn tiếp nhận dòng điện từ chuỗi các tế bào nối tiếp. Trạng thái này dẫn đến dòng điện cục bộ lớn và phát nhiệt mạnh với mật độ lên tới 1079.62 W/m^2 , làm xuất hiện điểm nóng (hotspot) với nhiệt độ cao hơn khu vực xung quanh khoảng [18]. Sự chênh lệch dòng điện giữa các tế bào nối tiếp không chỉ làm giảm hiệu suất toàn hệ mà còn gây dòng rò cục bộ và stress nhiệt, dẫn đến nguy cơ nứt vỡ tế bào, hư hỏng diode bypass và suy thoái sớm lớp encapsulant. Theo các phân tích bằng ảnh nhiệt (infrared thermography), vùng bụi tích tụ cục bộ có thể gây ra chênh lệch nhiệt độ lên đến $10\text{--}15^\circ\text{C}$ so với phần còn lại của mô-đun, được xem là dấu hiệu cảnh báo sớm cho tổn thất năng lượng ẩn và rủi ro hư hỏng vĩnh viễn [17].

Ngoài ra, bản chất hóa học và kích thước hạt của bụi cũng quyết định mức độ ảnh hưởng đến hiệu suất PV [19]. Theo kết quả nghiên cứu thực nghiệm trong phòng thí nghiệm của Jamil et al. (2015) [20], bụi có thể làm giảm công suất đầu ra tối đa của mô-đun PV tới 57.78% khi mật độ bụi tăng từ 0 đến 200 g/m^2 . Ngoài ra, bụi có kích thước hạt nhỏ và màu sáng cho thấy khả năng làm giảm hiệu suất nghiêm trọng hơn so với bụi hạt to hoặc màu tối. Cụ thể, bụi mịn như bột mì làm giảm dòng ngắn mạch (I_{sc}) tới 29.76%, trong khi bụi có màu sáng như bột ớt hoặc bột nghệ cũng làm giảm công suất đáng kể hơn so với các loại bụi có màu sẫm như cà phê hay bột cà ri đen. Trong khi đó, nghiên cứu của Wang et al. (2019) [21] cho thấy các hạt bụi chứa muối như NaCl có khả năng hút ẩm từ không khí, tạo thành lớp dung dịch mỏng khi gặp sương đêm. Sau khi nước bay hơi, lớp này kết tinh

thành màng muối bám chặt trên bề mặt mô-đun, gây giảm truyền sáng và hiệu suất phát điện, đồng thời rất khó loại bỏ bằng phương pháp vệ sinh thông thường, thậm chí cả chổi quét, và yêu cầu giải pháp làm sạch hóa học hoặc cơ giới chuyên dụng

Mức độ tổn thất công suất do bụi phụ thuộc chặt chẽ vào:

- Mật độ bụi (g/m^2),
- Chu kỳ làm sạch (số ngày),
- Đặc điểm khí hậu (tần suất mưa, độ ẩm, tốc độ gió),
- Góc nghiêng và hướng lắp đặt mô-đun,
- Vật liệu và đặc tính bề mặt kính (glass texture, hydrophobicity).

Một nghiên cứu tại UAE cho thấy cứ mỗi gram bụi/ m^2 tăng thêm, hiệu suất giảm khoảng 0.4–0.5%. Tuy nhiên, mối quan hệ này không tuyến tính, mà có xu hướng bão hòa (saturation) khi lớp bụi đạt đến ngưỡng cản sáng gần tuyệt đối (opacity threshold). Nghiên cứu [22] đã ghi nhận chỉ với 10 g/m^2 bụi đã có thể làm giảm công suất đầu ra tới 34%. Tuy nhiên, khi lượng bụi tiếp tục tăng, mức độ suy giảm hiệu suất chậm lại rõ rệt và có xu hướng tiệm cận.

Về mặt kinh tế, chi phí vệ sinh thủ công hệ thống điện mặt trời thường dao động từ 1.5 đến 4 USD/kWp mỗi năm, tùy thuộc vào phương pháp thực hiện, điều kiện tiếp cận và đặc điểm môi trường. Với các nhà máy quy mô lớn, tổng chi phí vệ sinh có thể lên đến hàng chục nghìn USD mỗi MW mỗi năm, như được ghi nhận tại Algeria và Saudi Arabia, đặc biệt trong điều kiện bụi tích tụ trên 7% sẽ gây tổn thất sản lượng đáng kể nếu không được xử lý kịp thời [23]. Do đó, bài toán tối ưu hóa lịch làm sạch (cleaning scheduling) trở nên cấp thiết, nhằm cân bằng giữa chi phí vận hành và giá trị điện năng thu hồi sau khi vệ sinh.

Tóm lại, ảnh hưởng của bụi và khí dung là một yếu tố động, phức tạp và không thể xem nhẹ trong thiết kế và vận hành hệ thống PV.

1.2.4. Lão hóa và suy giảm dài hạn của mô-đun PV

Một trong những yếu tố kỹ thuật quan trọng cần được tính đến trong thiết kế và vận hành hệ thống điện mặt trời là sự suy giảm hiệu suất của mô-đun PV theo thời gian. Mặc dù các nhà sản xuất thường cam kết mô-đun vẫn duy trì 80% công suất ban đầu sau 25 năm, tương đương với tốc độ suy giảm trung bình khoảng 0.7–0.8%/năm, nhưng trên thực

tế, nhiều nghiên cứu đã chỉ ra rằng tỷ lệ này phụ thuộc mạnh vào điều kiện khí hậu và môi trường vận hành, và có thể gia tăng đáng kể trong các vùng có điều kiện khắc nghiệt.

Dữ liệu từ báo cáo thử nghiệm thực địa của Carigiet et al. (2021) [24] tại Thụy Sĩ cho thấy các mô-đun tinh thể silicon đa tinh thể (multi c-Si) có thể duy trì tốc độ suy giảm rất thấp, chỉ 0.18–0.29%/năm, nếu được lắp đặt, giám sát và bảo trì trong điều kiện lý tưởng. Điều này cho thấy giá trị suy giảm được công bố bởi nhà sản xuất là một giá trị danh định, mang tính tiêu chuẩn, và thực tế vận hành có thể tốt hơn – hoặc xấu hơn – tùy điều kiện thực tế.

Theo tổng hợp của Fouad et al. (2017) [14] và Sharma & Chandel (2013) [25], các yếu tố bên ngoài làm gia tăng tỷ lệ suy giảm bao gồm:

- Bức xạ UV cao: thúc đẩy quá trình lão hóa vật liệu encapsulant, gây vàng hóa (yellowing), giảm độ truyền sáng đến cell.
- Nhiệt độ môi trường cao, kết hợp với chu kỳ nhiệt – ẩm liên tục: gây co giãn không đồng đều, dẫn đến vi nứt (microcrack) và mỏi hàn (solder fatigue).
- Xâm nhập độ ẩm (moisture ingress): phá hủy lớp chống thấm hoặc backsheet, tạo phản ứng ăn mòn mối nối hàn và hình thành acid acetic từ EVA.
- Tích lũy bụi và chất ô nhiễm bề mặt: gây hotspot cục bộ, tăng điện trở nối tiếp, và làm tăng nhiệt độ hoạt động không đồng đều.
- Hiệu ứng PID (Potential-Induced Degradation) và LID (Light-Induced Degradation): xảy ra mạnh mẽ hơn khi mô-đun vận hành trong môi trường có điện áp cao và độ ẩm lớn – thường gặp ở các nhà máy điện mặt trời nổi lưới công suất lớn.

Báo cáo của Omazic et al. (2019) [26] cho thấy tại các vùng sa mạc khô nóng, nơi nhiệt độ môi trường ban ngày vượt 40°C và bức xạ UV rất mạnh, mô-đun sử dụng encapsulant EVA thông thường có thể bị suy giảm sớm ngay trong 5–7 năm đầu, với tốc độ có thể vượt 1.2–1.5%/năm, cao hơn đáng kể so với cam kết của nhà sản xuất

Ngoài các yếu tố vật lý – môi trường, các sai lệch trong khâu thiết kế – lắp đặt như: góc nghiêng không tối ưu, không đồng nhất mô-đun (mismatch), tổn thất dây dẫn cao, hệ thống MPPT vận hành sai điểm công suất... cũng làm gia tăng hao hụt tích lũy theo thời gian.

1.3. Các phương pháp tiếp cận

Hiện nay, để dự báo và đánh giá hiệu suất hệ thống điện mặt trời (PV), có ba hướng tiếp cận chính:

- Phương pháp vật lý (physical-based): Xây dựng các mô hình toán học dựa trên đặc tính vật lý của các tế bào quang điện. Các mô hình diode, phương trình lý thuyết về chuyển đổi quang điện thường được sử dụng. Tuy nhiên, các phương pháp này đòi hỏi nhiều tham số kỹ thuật chi tiết từ datasheet và điều kiện vận hành lý tưởng (STC). Khi áp dụng vào thực tế, các mô hình vật lý thường gặp khó khăn do khó phản ánh đầy đủ tác động phi tuyến của các yếu tố môi trường như nhiệt độ biến động, bụi bẩn và góc tới không đồng đều của ánh sáng.
- Phương pháp thống kê, toán học (statistical-based): Sử dụng các mô hình hồi quy tuyến tính và phi tuyến (như hồi quy đa thức, hồi quy logistic), hoặc các kỹ thuật phân tích chuỗi thời gian (ARIMA) để mô hình hóa mối quan hệ giữa các biến đầu vào môi trường (bức xạ, nhiệt độ...) và công suất đầu ra. Ưu điểm là tính toán nhanh, dễ cài đặt nhưng khả năng xử lý phi tuyến và nhiễu trong dữ liệu thực tế còn hạn chế, hiệu quả dự báo không cao khi môi trường vận hành có nhiều biến động.
- Phương pháp dựa trên học máy (Machine Learning - ML): Các thuật toán ML như mạng nơ-ron nhân tạo (ANN), hồi quy véc-tơ hỗ trợ (SVR), XGBoost và Random Forest đang được ứng dụng rộng rãi nhờ khả năng xử lý hiệu quả các mối quan hệ phi tuyến giữa đầu vào và đầu ra. Phương pháp này có thể thích ứng tốt với biến động dữ liệu, đặc biệt hiệu quả khi có đủ lượng dữ liệu thực tế được thu thập đồng bộ và đầy đủ từ các cảm biến. Tuy nhiên, chất lượng dự báo của các thuật toán ML phụ thuộc mạnh vào tính đầy đủ và độ chính xác của dữ liệu huấn luyện, đồng thời đòi hỏi kỹ thuật tiền xử lý dữ liệu hiệu quả.

1.4. Đánh giá hiệu suất online

Việc đánh giá hiệu suất vận hành hệ thống điện mặt trời theo thời gian thực là yếu tố quan trọng nhằm đảm bảo hiệu quả khai thác năng lượng và tối ưu hóa công tác bảo trì. Quá trình này yêu cầu giám sát liên tục các thông số điện như công suất, dòng điện, điện áp, kết hợp với các yếu tố môi trường gồm bức xạ mặt trời, nhiệt độ môi trường, nhiệt độ

tế bào PV và độ ẩm. Dữ liệu được thu thập, đồng bộ và xử lý thông qua hệ thống giám sát dựa trên nền tảng IoT.

Hệ thống đánh giá hiệu suất online có ưu thế đặc biệt khi triển khai tại các khu vực vùng sâu, vùng xa – nơi việc kiểm tra và bảo trì thủ công thường gặp nhiều khó khăn và tốn kém. Nhờ tích hợp kết nối mạng 4G và nền tảng trực quan (Grafana), người vận hành có thể theo dõi tình trạng hoạt động mọi lúc, mọi nơi mà không cần có mặt tại hiện trường. Điều này giúp giảm chi phí nhân lực, thời gian vận hành và tăng cường khả năng phản ứng khi xảy ra sự cố.

Ngoài chức năng giám sát, hệ thống còn tích hợp các mô hình học máy như SVR và XGBoost để dự đoán công suất đầu ra dựa trên dữ liệu môi trường. Việc so sánh liên tục giữa công suất thực tế và công suất dự đoán giúp phát hiện sớm các bất thường như suy giảm hiệu suất do bụi bẩn, nhiệt độ cao hoặc lỗi thiết bị. Đây là nền tảng phù hợp để triển khai bảo trì theo điều kiện (Condition-Based Maintenance – CBM), thay vì bảo trì định kỳ truyền thống.

Về mặt kiến trúc, hệ thống sử dụng phần cứng chi phí thấp (Raspberry Pi) cùng phần mềm mã nguồn mở (Node-RED, InfluxDB, Grafana), giúp giảm đáng kể chi phí đầu tư nhưng vẫn bảo đảm khả năng mở rộng và tùy chỉnh. Việc tích hợp dữ liệu điện và môi trường trên cùng nền tảng cho phép phân tích toàn diện và hỗ trợ ra quyết định kỹ thuật chính xác, kịp thời.

Tổng thể, hệ thống đánh giá hiệu suất online dựa trên IoT và học máy giúp giám sát hiệu quả và dự báo chính xác, mà còn đặc biệt phù hợp với các khu vực địa lý hạn chế. Giải pháp này góp phần nâng cao độ tin cậy, kéo dài tuổi thọ hệ thống và tối ưu chi phí vận hành cho các hệ thống điện mặt trời.

1.5. Kết luận

Chương 1 đã trình bày tổng quan về bối cảnh nghiên cứu trong lĩnh vực giám sát và đánh giá hiệu suất hệ thống điện mặt trời mái nhà. Qua đó, làm rõ tính cấp thiết của việc ứng dụng các công nghệ hiện đại như Internet of Things (IoT) và trí tuệ nhân tạo (AI) nhằm giải quyết các vấn đề về thu thập, giám sát và phân tích dữ liệu trong thời gian thực.

Chương 2: CƠ SỞ ĐỂ XÂY DỰNG MÔ HÌNH ƯỚC LƯỢNG CÔNG SUẤT VÀ HỆ THỐNG ĐÁNH GIÁ HIỆU SUẤT

2.1. Giới thiệu

Hệ thống điện mặt trời chịu ảnh hưởng mạnh mẽ bởi các yếu tố thời tiết như bức xạ mặt trời, nhiệt độ, bụi và độ ẩm. Việc dự báo công suất đầu ra là cần thiết để phục vụ giám sát hiệu suất, lên kế hoạch vận hành và bảo trì. Trong chương này, các mô hình dự báo hiệu suất điện mặt trời sẽ được trình bày, bao gồm mô hình truyền thống (dựa vào công thức vật lý và thống kê) và các mô hình học máy hiện đại. Bên cạnh đó, các chỉ số đánh giá và tiêu chí lựa chọn mô hình phù hợp cũng được đề cập để đảm bảo tính chính xác và khả năng áp dụng vào hệ thống thực tế.

2.2. Mô hình dự báo hiệu suất điện mặt trời

2.2.1. Mô hình vật lý

Mô hình vật lý là một hướng tiếp cận căn bản trong việc mô phỏng hành vi và dự báo công suất đầu ra của tế bào quang điện (PV), vốn dựa trên các định luật bán dẫn và mối quan hệ dòng–áp đặc trưng cho quá trình chuyển đổi năng lượng từ ánh sáng sang điện. Trong cấu trúc truyền thống, tế bào PV thường được biểu diễn thông qua các mạch tương đương điện học, nổi bật nhất là các biến thể của mô hình diode. Tùy theo mức độ chi tiết và mục tiêu mô phỏng, ba mô hình chính thường được sử dụng: mô hình một diode (SDM), hai diode (DDM), và ba diode (TDM) – mỗi loại đại diện cho một cấp độ chính xác và phức tạp khác nhau [27].

Ở mức cơ bản, mô hình một diode (SDM) giả định quá trình tái tổ hợp mang điện chỉ diễn ra trong vùng suy giảm của tiếp giáp p–n, và đặc tuyến dòng–áp được biểu diễn bởi năm tham số bao gồm: dòng quang điện I_{ph} , dòng bão hòa ngược I_0 , hệ số lý tưởng n , điện trở nối tiếp R_s và điện trở song song R_{sh} . Cấu trúc này phù hợp cho các phân tích lý tưởng hóa hoặc điều kiện chiếu sáng ổn định. Tuy nhiên, trong môi trường thực tế, nơi có nhiều yếu tố phi tuyến như ánh sáng yếu, sự không đồng đều giữa các cell, hay rò dòng ở vùng rìa, mô hình SDM không còn đủ khả năng mô phỏng chính xác [28].

Để khắc phục giới hạn trên, mô hình hai diode (DDM) được phát triển bằng cách bổ sung thêm một nhánh diode thứ hai, đại diện cho cơ chế tái tổ hợp trong vùng trung tính của vật liệu bán dẫn. Mô hình DDM mở rộng thành 7 tham số, cho phép mô phỏng tốt hơn đặc tuyến I–V trong toàn dải điều kiện vận hành, đặc biệt là ở vùng điện áp thấp nơi các cơ chế tái tổ hợp chiếm ưu thế. Trong thực nghiệm, DDM thường cho sai số RMSE thấp hơn SDM từ 15–30% nếu được tối ưu hóa bằng các thuật toán metaheuristic như Differential Evolution (DE), Particle Swarm Optimization (PSO) hoặc Artificial Bee Colony (ABC) [29].

Tuy nhiên, trong các nghiên cứu chuyên sâu hoặc khi cần mô hình hóa các hiệu ứng rò dòng cực nhỏ ở vùng biên tế bào – đặc biệt phổ biến trong các tấm pin cắt nhỏ (half-cut cell) hay cấu trúc liên kết nối phức tạp – mô hình ba diode (TDM) được xem là giải pháp chính xác nhất. TDM bổ sung thêm một diode nữa, đưa tổng số tham số cần xác định lên 9. Các nghiên cứu gần đây cho thấy TDM đặc biệt hiệu quả khi khớp dữ liệu thực nghiệm từ các mô-đun như RTC France hoặc Solarex MSX–60, với mức cải thiện RMSE từ 5–15% so với DDM nếu sử dụng thuật toán CLSHADE kết hợp hàm Lambert W để giải phương trình dòng điện phi tuyến [30].

Về mặt toán học, cả DDM và TDM đều có cấu trúc phương trình dạng siêu việt (transcendental) – không thể giải tích trực tiếp. Do đó, việc ước lượng tham số thường được triển khai thông qua phương pháp lặp kết hợp hàm Lambert W hoặc mô phỏng Monte Carlo để đảm bảo hội tụ chính xác. Mặc dù mang lại hiệu suất mô phỏng cao, song mô hình TDM cũng đi kèm với chi phí tính toán lớn, đòi hỏi cấu hình xử lý mạnh, đồng thời dễ bị nhiễu khi dữ liệu đầu vào không đầy đủ hoặc sai số cao [30].

Tóm lại, các mô hình diode truyền thống – từ SDM đến DDM và TDM – đóng vai trò nền tảng trong việc dự báo công suất PV theo hướng vật lý. Ở cấp độ ứng dụng, SDM có thể phù hợp với mô phỏng nhanh trong điều kiện tiêu chuẩn; DDM phù hợp hơn với hệ thống thực tế; trong khi TDM hướng tới nghiên cứu chuyên sâu hoặc tích hợp vào hệ thống điều khiển thông minh nơi yêu cầu độ chính xác cao trong mọi điều kiện chiếu sáng và vật lý tế bào. Dù vậy, để tăng khả năng thích ứng trong môi trường biến động – ví dụ như có bụi, râm bóng hay xuống cấp vật lý – các mô hình này cần được kết hợp với các thuật toán học máy hiện đại để tạo thành mô hình lai (hybrid), tận dụng ưu điểm từ cả hai thế giới mô phỏng vật lý và phân tích dữ liệu [31].

2.2.2. Mô hình học máy

Trong lĩnh vực dự báo công suất điện mặt trời, các mô hình học máy đang ngày càng được áp dụng rộng rãi nhờ khả năng xử lý linh hoạt, khai thác tự động mối quan hệ phi tuyến giữa nhiều biến đầu vào và đầu ra. Khác với các mô hình vật lý truyền thống, mô hình học máy không cần biết chi tiết các tham số thiết bị hay cấu trúc hệ thống, mà thay vào đó học trực tiếp từ dữ liệu vận hành trong quá khứ.

Dưới đây là phân tích chi tiết về các nhóm mô hình học máy phổ biến trong dự báo công suất PV:

a) Artificial Neural Networks (ANN)

Mạng nơron nhân tạo (Artificial Neural Network – ANN) là một trong những mô hình học máy được ứng dụng rộng rãi nhất trong dự báo công suất điện mặt trời do khả năng học và biểu diễn mối quan hệ phi tuyến phức tạp giữa các biến đầu vào (nhiệt độ, bức xạ mặt trời, độ ẩm, thời gian trong ngày...) và công suất đầu ra của hệ thống PV. Cấu trúc cơ bản của một ANN bao gồm ba tầng: tầng đầu vào (input layer), một hoặc nhiều tầng ẩn (hidden layers), và tầng đầu ra (output layer). Mỗi tầng bao gồm nhiều nút (neurons) được kết nối với các trọng số, và các trọng số này được tối ưu hóa trong quá trình huấn luyện để giảm sai số giữa giá trị đầu ra dự đoán và thực tế.

Ưu điểm lớn nhất của ANN là khả năng tự động học từ dữ liệu mà không cần mô tả tường minh quy luật vật lý, đồng thời có thể mô hình hóa các quá trình phi tuyến, bất định và có nhiễu – điều rất phổ biến trong dữ liệu năng lượng mặt trời thực tế. Trong nhiều nghiên cứu, ANN đã cho kết quả vượt trội so với các mô hình tuyến tính hoặc thống kê truyền thống. Ví dụ, nghiên cứu của Essam et al. (2022) [32] cho thấy mô hình ANN đạt hệ số xác định $R^2 = 0.9988$, RMSE = 0.88 W/kWp, MAE = 0.47 W/kWp khi dự báo công suất PV tại Florida, Mỹ – vượt trội hơn so với Random Forest và Decision Tree trong cùng điều kiện

Một điểm mạnh khác của ANN là tính thích nghi với dữ liệu theo thời gian, cho phép cập nhật trọng số định kỳ khi hệ thống PV có dấu hiệu suy giảm hoặc khi điều kiện môi trường thay đổi. Tuy nhiên, ANN cũng tồn tại một số hạn chế. Thứ nhất, quá trình huấn luyện yêu cầu lượng dữ liệu đủ lớn và đại diện để tránh hiện tượng quá khớp (overfitting). Thứ hai, lựa chọn kiến trúc mạng (số tầng ẩn, số neuron/tầng, hàm kích hoạt, learning

rate...) đóng vai trò then chốt trong chất lượng dự báo, nhưng hiện nay chưa có quy trình chung để xác định tối ưu các siêu tham số này.

Để nâng cao hiệu quả, nhiều nghiên cứu đã kết hợp ANN với các phương pháp xử lý dữ liệu như chuẩn hóa đầu vào, giảm chiều dữ liệu bằng PCA, hoặc tối ưu hóa siêu tham số bằng các thuật toán metaheuristic như PSO (Particle Swarm Optimization), GA (Genetic Algorithm). Ngoài ra, các biến thể sâu hơn như Deep ANN, CNN-ANN hoặc hybrid với mô hình thống kê cũng đang được nghiên cứu và triển khai trong các hệ thống PV quy mô lớn.

b) Support Vector Regression (SVR)

Support Vector Regression (SVR) là một biến thể của thuật toán máy vector hỗ trợ (SVM – Support Vector Machine), được thiết kế cho các bài toán hồi quy thay vì phân loại. Trong bối cảnh dự báo công suất điện mặt trời, SVR đã chứng minh hiệu quả cao trong việc xử lý dữ liệu phi tuyến, có nhiễu và khối lượng giới hạn – vốn là đặc trưng phổ biến của dữ liệu môi trường và sản lượng thực tế của hệ thống PV.

Một trong những đặc điểm quan trọng của SVR là việc sử dụng các hàm kernel để ánh xạ dữ liệu từ không gian đầu vào (input space) sang không gian đặc trưng có chiều cao hơn (feature space), nơi mối quan hệ giữa biến đầu vào và đầu ra có thể trở nên tuyến tính. Các kernel phổ biến bao gồm hàm tuyến tính, hàm Gaussian RBF (radial basis function), và polynomial. Trong dự báo công suất PV, kernel RBF thường được sử dụng rộng rãi vì có khả năng nắm bắt các mối quan hệ phi tuyến mạnh giữa các yếu tố như bức xạ, nhiệt độ, thời gian và công suất phát.

SVR đặc biệt hiệu quả trong bài toán dự báo ngắn hạn (short-term forecasting), nơi dữ liệu đầu vào thay đổi theo giờ hoặc phút. Ví dụ, trong nghiên cứu của Ramadhan et al. (2021) [33], mô hình SVR đã được sử dụng để dự báo công suất PV với độ chính xác cao, và được đánh giá là ổn định hơn so với ANN khi kích thước dữ liệu huấn luyện bị giới hạn hoặc phân bố không đồng đều. Ngoài ra, SVR thường có thời gian huấn luyện ngắn hơn và ít bị ảnh hưởng bởi local minima, nhờ tính chất convex của bài toán tối ưu.

Tuy nhiên, việc ứng dụng SVR cũng đặt ra thách thức trong việc lựa chọn siêu tham số: epsilon (ϵ), hệ số điều chỉnh lỗi (C), và tham số kernel (γ đối với RBF). Các tham số này ảnh hưởng trực tiếp đến độ chính xác và khả năng khái quát của mô hình. Để giải

quyết, nhiều nghiên cứu đã kết hợp SVR với các thuật toán tối ưu hóa như Particle Swarm Optimization (PSO), Grid Search, hay Bayesian Optimization để tự động điều chỉnh bộ tham số này.

c) *Random Forest (RF)*

Random Forest (RF) là một thuật toán học máy thuộc nhóm mô hình tổ hợp (ensemble learning), hoạt động theo nguyên lý bagging – tức huấn luyện nhiều mô hình con (cây quyết định – Decision Trees) độc lập trên các tập dữ liệu được chọn ngẫu nhiên từ dữ liệu gốc, sau đó kết hợp kết quả đầu ra bằng trung bình (đối với hồi quy) hoặc biểu quyết số đông (đối với phân loại). Trong dự báo công suất điện mặt trời, RF đã được chứng minh là một trong những phương pháp đáng tin cậy, đặc biệt khi dữ liệu đầu vào có tính nhiễu cao hoặc chứa nhiều đặc trưng không tuyến tính.

Ưu điểm nổi bật của RF là khả năng chống overfitting tốt hơn so với cây quyết định đơn lẻ, nhờ vào quá trình khử tương quan giữa các cây và sử dụng kỹ thuật bootstrap sampling. RF có khả năng xử lý tốt cả dữ liệu rời rạc lẫn liên tục, không yêu cầu chuẩn hóa dữ liệu, và ít nhạy cảm với các đặc trưng không liên quan. Đây là lý do vì sao RF thường được lựa chọn trong các ứng dụng năng lượng tái tạo, nơi dữ liệu môi trường (như nhiệt độ, độ ẩm, gió, bức xạ...) thay đổi nhanh và không tuân theo phân phối tuyến tính.

Trong dự báo công suất PV, RF có thể xử lý hàng chục đặc trưng đầu vào và vẫn duy trì hiệu năng tốt mà không cần cấu trúc mô hình phức tạp như ANN hay LSTM. Theo nghiên cứu của Essam et al. (2022) [32], RF cho kết quả ổn định và độ chính xác cao trên tập dữ liệu NREL tại Florida với RMSE thấp hơn 1 W/kWp và hệ số $R^2 > 0.99$, chỉ đứng sau ANN về hiệu năng tổng thể trong cùng điều kiện dự báo công suất PV đơn mô-đun

Ngoài ra, RF còn cung cấp một ưu điểm mà các mô hình như ANN hay SVR không có: khả năng đo lường tầm quan trọng của từng biến đầu vào (feature importance). Điều này đặc biệt hữu ích trong việc xác định yếu tố môi trường nào (bức xạ toàn phần, nhiệt độ tấm pin, độ ẩm, v.v.) có ảnh hưởng lớn nhất đến công suất PV tại từng thời điểm hoặc từng khu vực cụ thể. Khả năng giải thích (interpretability) này làm cho RF trở nên phù hợp với các ứng dụng cần minh bạch trong ngành năng lượng và quản lý hệ thống lưới điện.

Tuy nhiên, RF cũng có một số hạn chế. Do số lượng cây trong rừng có thể lớn (100–500 cây), mô hình có thể tiêu tốn tài nguyên tính toán đáng kể khi áp dụng trong thời gian

thực. Hơn nữa, mặc dù RF hoạt động tốt với dữ liệu không chuẩn hóa, nhưng hiệu năng có thể suy giảm khi dữ liệu có độ mất cân bằng cao hoặc nhiều nhiễu không liên quan. RF cũng không khai thác được cấu trúc chuỗi thời gian như các mô hình LSTM hay CNN.

d) *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) là một thuật toán học máy thuộc nhóm mô hình tổ hợp (ensemble learning) sử dụng kỹ thuật tăng cường (boosting). Khác với Random Forest (bagging), trong đó các cây quyết định được huấn luyện song song và độc lập, boosting xây dựng các cây một cách tuần tự – mỗi cây mới được huấn luyện để khắc phục sai số (residual) của các cây trước đó. XGBoost là phiên bản cải tiến của thuật toán gradient boosting truyền thống, được tối ưu hóa cả về tốc độ tính toán và khả năng chính xác, trở thành một trong những mô hình mạnh mẽ và phổ biến nhất hiện nay trong các bài toán hồi quy – bao gồm cả dự báo công suất điện mặt trời.

Trong bối cảnh hệ thống PV, công suất đầu ra phụ thuộc vào nhiều yếu tố phi tuyến như bức xạ mặt trời, nhiệt độ, độ ẩm, thời gian trong ngày, điều kiện che bóng và tính chất vật liệu mô-đun. Những mối quan hệ phức tạp như vậy thường không thể được biểu diễn hiệu quả bằng các mô hình tuyến tính hay cây đơn lẻ. XGBoost tận dụng hệ thống các cây yếu (shallow trees) để kết hợp lại thành một mô hình mạnh hơn, liên tục cập nhật hướng dẫn điều chỉnh sai số trong từng vòng lặp, từ đó cải thiện đáng kể độ chính xác dự báo.

XGBoost tích hợp các kỹ thuật quan trọng như:

- Regularization (L1/L2): giúp kiểm soát độ phức tạp mô hình, tránh overfitting;
- Parallel computing: huấn luyện cây nhanh hơn so với boosting truyền thống;
- Sparsity-aware: xử lý tốt dữ liệu bị thiếu hoặc phân bố không đều;
- Pruning và shrinkage: hỗ trợ loại bỏ các nhánh không cần thiết và giảm ảnh hưởng của mỗi cây.

Theo Essam et al. (2022) [32], XGBoost đã cho kết quả dự báo công suất PV chính xác vượt trội so với các mô hình học máy truyền thống như SVR và RF trong một số trường hợp, đặc biệt khi lượng dữ liệu lớn và có nhiễu. Trong nghiên cứu tại Florida (dữ liệu NREL), XGBoost đạt được RMSE < 1.0 W/kWp, hệ số xác định $R^2 > 0.99$, chứng minh khả năng mô hình hóa phi tuyến và tối ưu hóa sai số liên tục của thuật toán

Tuy nhiên, một số thách thức khi triển khai XGBoost là yêu cầu lựa chọn siêu tham số phù hợp (learning rate, max depth, subsample, colsample_bytree, v.v.), vì các tham số này ảnh hưởng lớn đến hiệu suất mô hình. Việc tuning cần sử dụng các kỹ thuật như Grid Search, Random Search hoặc Bayesian Optimization để đạt được cấu hình tối ưu.

Xét về toàn diện, XGBoost là mô hình phù hợp cho các hệ thống PV quy mô lớn hoặc các bài toán dự báo có tính biến thiên cao, yêu cầu độ chính xác và độ ổn định vượt trội. Khi được tối ưu hóa đúng cách và kết hợp với tiền xử lý dữ liệu hợp lý, XGBoost không chỉ đạt hiệu suất dự báo cao mà còn hoạt động tốt trong các điều kiện thực tế có nhiều và nhiều biến đầu vào đồng thời. Do đó, nó là một trong những lựa chọn hàng đầu trong các hệ thống giám sát và dự báo công suất điện mặt trời hiện đại.

2.2.3. Đánh giá và lựa chọn mô hình

Như đã trình bày ở Chương 1, trong đề án này chúng tôi sẽ lựa chọn các mô hình học máy để dự báo công suất lý tưởng của hệ thống điện mặt trời. Việc lựa chọn mô hình học máy cần được cân nhắc dựa trên đặc điểm cụ thể của tập dữ liệu và mục tiêu ứng dụng, thay vì chỉ dựa vào độ chính xác thống kê. Với dữ liệu quy mô nhỏ hoặc vừa, mô hình như SVR thường hiệu quả do khả năng khái quát tốt và ít tham số. Ngược lại, với dữ liệu lớn và nhiều đặc trưng, mô hình như XGBoost có ưu thế trong việc mô hình hóa các mối quan hệ phi tuyến sâu và kiểm soát overfitting.

Ngoài ra, nếu dữ liệu có tính chuỗi thời gian rõ rệt, các mô hình như LSTM sẽ phát huy hiệu quả nhờ khả năng ghi nhớ dài hạn. Trong khi đó, nếu bài toán yêu cầu mô hình có thể giải thích được (interpretability), các mô hình như XGBoost sẽ được ưu tiên hơn các mô hình học sâu.

Theo Yang et al. (2022) [34], không phải mô hình nào có sai số thấp cũng mô phỏng đúng hành vi hệ thống. Trong nghiên cứu của họ, dù cả SVR, ANN và RF đều đạt R^2 cao, chỉ SVR và ANN có khả năng học được quy luật nội tại, trong khi Random Forest thiếu năng lực này. Do đó, tác giả khuyến nghị rằng việc lựa chọn mô hình nên dựa trên khả năng mô hình hóa bản chất dữ liệu, không chỉ dựa vào RMSE hay MAE

Trong nghiên cứu này, SVR-RBF và XGBoost được lựa chọn để huấn luyện và so sánh, nhằm đánh giá hiệu quả dự báo công suất PV theo thời gian thực trên cơ sở dữ liệu đo đạc thực tế.

2.3. Giải pháp thu thập dữ liệu và giám sát

2.3.1. Giải pháp thu thập và giám sát cục bộ

Hệ thống thu thập và giám sát cục bộ cho các mảng điện mặt trời mái nhà chủ yếu dựa vào việc tận dụng khả năng ghi nhận dữ liệu tích hợp sẵn trong biến tần hoặc trong các bộ ghi dữ liệu (SmartLogger) để thu thập các thông số điện cơ bản bao gồm điện áp đầu vào/đầu ra, dòng điện, công suất và trạng thái vận hành mà không mở rộng thêm các cảm biến môi trường như bức xạ, nhiệt độ hay độ ẩm [35]. Về cơ bản, SmartLogger kết nối trực tiếp với biến tần qua giao thức RS-485, luân chuyển gói dữ liệu theo chu kỳ cố định (thường 1–5 phút) về bộ nhớ cục bộ hoặc giao diện web nội bộ, cho phép người vận hành theo dõi biểu đồ điện áp-dòng-công suất và nhận cảnh báo sự cố ngay tại chỗ mà không cần kết nối Internet [35]. Ở giải pháp đơn giản hơn, như trong báo cáo về hệ thống thu thập dữ liệu dựa trên vi điều khiển Arduino, dữ liệu công suất được ghi vào thẻ SD và có thể xuất qua cổng USB để phân tích hậu kỳ; tuy đề xuất này có thêm cảm biến nhiệt độ-độ ẩm (DHT11) và cảm biến ánh sáng (photoresistor), nhưng thực tế ứng dụng cục bộ ở quy mô mái nhà thường không triển khai môi trường lừa-vật-lý mà chỉ tập trung vào thông số điện năng để giữ chi phí ở mức tối thiểu [36].

Ưu điểm nổi bật của hệ thống giám sát cục bộ nằm ở tính chi phí và độ tin cậy cao: bởi không phụ thuộc vào đường truyền Internet, hệ thống vẫn hoạt động ổn định ngay cả khi mất kết nối bên ngoài, đồng thời hạn chế chi phí phát sinh cho SIM 4G hoặc dịch vụ đám mây [38]. Việc sử dụng SmartLogger tích hợp sẵn trong biến tần giúp loại bỏ nhu cầu đầu tư thêm phần cứng phức tạp, dễ dàng tích hợp vào hạ tầng điện hiện hữu và cung cấp giao diện trực quan ngay trên máy tính hoặc thiết bị di động trong mạng LAN nội bộ. Các nhà khai thác có thể cấu hình mức ngưỡng cảnh báo, lưu trữ lịch sử vận hành và xuất báo cáo hàng ngày/tuần/tháng mà không cần kiến thức chuyên sâu về lập trình hoặc quản lý cơ sở dữ liệu phức tạp [37].

Tuy nhiên, hệ thống này cũng tồn tại nhiều hạn chế đáng lưu ý. Trước hết, thiếu hụt thông tin về điều kiện môi trường khiến việc đánh giá hiệu suất tương quan giữa công suất thực và bức xạ mặt trời hoặc nhiệt độ mô-đun trở nên không chính xác, làm giới hạn khả năng phân tích nguyên nhân sụt giảm hiệu suất và ứng dụng các mô hình dự báo năng suất hay khuyến nghị bảo trì dự báo. Tiếp theo, lưu trữ dữ liệu cục bộ thường bị giới hạn bởi dung lượng EEPROM hoặc thẻ SD, dễ gặp rủi ro mất dữ liệu do hỏng thẻ, tràn bộ nhớ hoặc

sự cố nguồn, đồng thời thiếu các cơ chế sao lưu tự động lên hệ thống bên ngoài để khôi phục thông tin khi cần thiết. Hơn nữa, việc nâng cấp phần mềm firmware hay cập nhật giao diện người dùng đòi hỏi thao tác thủ công tại hiện trường, làm gia tăng chi phí nhân công và thời gian bảo trì, đặc biệt trong các dự án phân tán nhiều địa điểm [38].

Bên cạnh đó, khả năng mở rộng hệ thống khi số lượng biến tần tăng lên hoặc mô hình kinh doanh yêu cầu tích hợp thêm chức năng điều khiển tải tự động (Demand Response) vẫn còn hạn chế: SmartLogger và các bộ ghi dữ liệu cơ bản không hỗ trợ các giao thức truyền thông công nghiệp như MQTT hay OPC UA một cách linh hoạt, nên khó đáp ứng nhu cầu kết nối đa điểm và tích hợp với các nền tảng SCADA cấp cao hơn [37]. Cuối cùng, về mặt an ninh, dữ liệu lưu trữ cục bộ thiếu các lớp bảo mật nâng cao mã hóa khi ghi, xác thực truy cập và phân quyền chi tiết khiến hệ thống dễ bị tấn công nội bộ hoặc truy cập trái phép nếu không thiết lập đúng chính sách mạng LAN [38].

Tóm lại, hệ thống giám sát cục bộ cho điện mặt trời mái nhà, dựa trên SmartLogger hay các bộ thu thập dữ liệu đơn giản, mang lại lợi ích lớn về kinh tế và độ tin cậy trong vận hành cơ bản. Tuy nhiên, để nâng cao khả năng phân tích hiệu suất, dự báo và tự động hóa bảo trì, đồng thời đảm bảo an ninh dữ liệu và khả năng mở rộng, cần cân nhắc kết hợp thêm các giải pháp IoT nâng cao bao gồm bổ sung cảm biến môi trường, giao thức truyền thông hiện đại và nền tảng đám mây nhằm khắc phục những hạn chế cố hữu của phương thức giám sát cục bộ truyền thống.

2.3.2. Giải pháp thu thập và giám sát từ xa

Hệ thống thu thập và giám sát từ xa cho các mảng điện mặt trời (remote PV monitoring) là một giải pháp dựa trên nền tảng IoT kết hợp với đám mây và các giao thức truyền thông hiện đại, cho phép quan sát, phân tích và điều khiển hoạt động của hệ thống PV từ bất cứ đâu thông qua Internet. Về phần cứng, bên cạnh việc tận dụng khả năng ghi nhận điện áp, dòng điện và công suất từ biến tần hoặc SmartLogger, người ta thường bổ sung các mô-đun xử lý dữ liệu như Raspberry Pi, Arduino hoặc các bộ vi điều khiển chuyên dụng (BeagleBone, PLC) để thu thập, tiền xử lý sơ bộ và đóng gói dữ liệu trước khi truyền lên máy chủ trung tâm [38].

Ở tầng truyền thông, hệ thống từ xa sử dụng đa dạng giao thức không dây lẫn có dây: từ Wi-Fi, Ethernet cho kết nối băng thông cao trong nội bộ, đến ZigBee, LoRa và Bluetooth

cho các mạng cảm biến công suất thấp; đồng thời, đối với các vị trí xa hoặc không có hạ tầng LAN, mạng di động 3G/4G/5G được ứng dụng để đưa dữ liệu lên nền tảng đám mây [Shaheer Ansari, 2021]. Tầng phần mềm đám mây thường bao gồm các dịch vụ lưu trữ (database), xử lý thời gian thực (stream processing) và kho cảng dữ liệu lớn (data lake), cho phép thực hiện phân tích mô tả, chẩn đoán, dự báo và khuyến nghị từ việc giám sát biểu đồ điện áp-dòng-công suất, phát hiện sụt giảm bất thường cho đến tối ưu hóa lịch trình vận hành và bảo trì dự báo [35].

Giao diện người dùng (web dashboard hoặc ứng dụng di động) kết nối trực tiếp với máy chủ đám mây, cung cấp báo cáo tổng quan, cảnh báo SMS/Email và khả năng điều khiển từ xa qua relay hoặc lệnh API. Ví dụ, trong hệ thống sử dụng Raspberry Pi không dây, dữ liệu bức xạ, nhiệt độ tấm và độ ẩm môi trường còn có thể được tích hợp để cải thiện độ chính xác mô hình dự báo công suất, đồng thời hỗ trợ so sánh hiệu suất thực tế với đầu ra lý thuyết [37].

Ưu điểm chính của giải pháp giám sát từ xa nằm ở khả năng quan sát đa điểm và mở rộng quy mô linh hoạt: một trung tâm điều khiển có thể theo dõi hàng trăm đến hàng nghìn hệ PV phân tán địa lý, tận dụng sức mạnh tính toán đám mây để thực hiện các thuật toán học máy và phân tích nâng cao, đồng thời giảm thiểu chi phí cài đặt tại mỗi site nhờ tái sử dụng phần cứng phổ thông [38]. Môi trường đám mây cũng đảm bảo sao lưu tự động, phục hồi dữ liệu khi thiết bị cục bộ gặp sự cố, và hỗ trợ cập nhật firmware từ xa (OTA), giúp duy trì độ tin cậy vận hành liên tục.

Tuy nhiên, hệ thống từ xa cũng chịu những thách thức về tính sẵn sàng mạng và độ trễ truyền tải: phụ thuộc vào chất lượng đường truyền Internet hay mạng di động, có thể gặp gián đoạn trong trường hợp mất sóng hoặc băng thông bị giới hạn; đồng thời, yêu cầu chi phí duy trì dịch vụ đám mây, SIM di động và bản quyền phần mềm phân tích có thể gia tăng tổng chi phí sở hữu (TCO) [38]. Vấn đề bảo mật và quyền riêng tư dữ liệu cũng cần được đảm bảo qua các cơ chế mã hóa đầu cuối (TLS/SSL), xác thực đa yếu tố và phân quyền truy cập, nếu không dễ dẫn đến nguy cơ tấn công hoặc truy cập trái phép

Tóm lại, hệ thống thu thập và giám sát từ xa kết hợp IoT, đám mây và đa giao thức truyền thông cung cấp giải pháp toàn diện, linh hoạt và mở rộng cho việc quản lý hiệu suất PV ở mọi quy mô. Để đạt hiệu quả tối ưu, cần cân bằng giữa chi phí kết nối, yêu cầu phân

tích nâng cao và mức độ tin cậy mạng, đồng thời xây dựng chiến lược bảo mật chặt chẽ để bảo vệ dữ liệu vận hành thiết bị.

2.4. Kết luận

Chương 2 đã cung cấp nền tảng lý thuyết và thực tiễn cần thiết để xây dựng một hệ thống dự báo công suất và đánh giá hiệu suất cho điện mặt trời mái nhà. Thông qua việc phân tích các mô hình vật lý từ đơn giản (SDM) đến phức tạp (TDM), chương này đã làm rõ vai trò của các tiếp cận truyền thống trong việc mô hình hóa đặc tính dòng-áp và hiệu suất tế bào quang điện, đồng thời chỉ ra các hạn chế cố hữu về khả năng thích ứng trong điều kiện vận hành thực tế biến động.

Song song đó, chương cũng trình bày toàn diện các phương pháp học máy hiện đại như ANN, SVR, Random Forest và XGBoost – mỗi phương pháp đều có những ưu điểm riêng trong việc học các mối quan hệ phi tuyến từ dữ liệu thực tế và thích nghi với môi trường vận hành thay đổi. Việc đánh giá lựa chọn mô hình được đề cập không chỉ dừng ở các chỉ số sai số (RMSE, MAE) mà còn mở rộng đến khả năng giải thích, tính ổn định, hiệu năng tính toán và sự phù hợp với quy mô ứng dụng.

Cuối cùng, chương đã tổng hợp và phân tích hai hướng triển khai hệ thống thu thập dữ liệu: giám sát cục bộ truyền thống và giám sát từ xa dựa trên nền tảng IoT – mỗi giải pháp đều có điểm mạnh và thách thức riêng, phụ thuộc vào yêu cầu về chi phí, mức độ tự động hóa, khả năng mở rộng và độ tin cậy kết nối.

Những nội dung trong chương này không chỉ cung cấp cơ sở lý thuyết cho các chương tiếp theo, mà còn đóng vai trò then chốt trong việc lựa chọn kiến trúc hệ thống phù hợp và xây dựng mô hình dự báo chính xác, bền vững trong các điều kiện vận hành thực tế tại Việt Nam và các khu vực có điều kiện môi trường tương tự.

Chương 3: XÂY DỰNG MÔ HÌNH HỌC MÁY ƯỚC LƯỢNG HIỆU SUẤT

3.1. Giới thiệu

Chương này trình bày quá trình xây dựng và đánh giá hai mô hình học máy dự báo sản lượng điện của hệ thống quang điện. Cụ thể, hai mô hình được triển khai gồm: Support Vector Regression (SVR) với hàm kernel RBF và XGBoost - một thuật toán boosting hiện đại dựa trên cây quyết định. Đây là hai phương pháp hồi quy có khả năng mô hình hóa quan hệ phi tuyến, phù hợp với đặc điểm biến động theo thời gian và chịu ảnh hưởng từ các yếu tố thời tiết của hệ thống năng lượng mặt trời.

Các mô hình được huấn luyện trên tập dữ liệu lịch sử thu thập từ hệ thống điện mặt trời mái nhà của một nhà máy tại huyện Hòa Vang (Đà Nẵng). Chuỗi đo bao phủ từ 02 tháng 03 đến 23 tháng 04 năm 2025 (≈ 52 ngày), với tần suất 5 phút/điểm, tạo thành 19 878 điểm dữ liệu liên tục trong cả chu kỳ 24 giờ. Bộ dữ liệu chứa các biến đầu vào then chốt chi phối hiệu suất PV, gồm: bức xạ mặt trời (irradiance), nhiệt độ môi trường, độ ẩm tương đối và công suất tức thời (input_power); thêm cột giờ trong ngày được suy xuất để đặc trưng hóa tính mùa vụ ngày–đêm. Sau giai đoạn tiền xử lý — loại bỏ giá trị ngoại lệ, nội suy/loại bỏ điểm khuyết và chuẩn hóa đặc trưng — toàn bộ dữ liệu của 51 ngày đầu được dùng làm tập huấn luyện. Ngày cuối cùng (23/04/2025) được sử dụng để đánh giá khả năng dự báo ngoài mẫu được học trong điều kiện thời tiết thực tế. Cách chia này bảo đảm phép kiểm định diễn ra trên dữ liệu hoàn toàn chưa từng xuất hiện trong quá trình huấn luyện, phản ánh đúng kịch bản vận hành thời gian thực của hệ thống giám sát.

Các mô hình sẽ được đánh giá theo ba chỉ số phổ biến: MSE, RMSE, và R^2 . Ngoài ra, hiệu suất mô hình còn được phân tích trực quan thông qua biểu đồ phân tán, biểu đồ chuỗi thời gian và biểu đồ sai số theo giờ. Việc kết hợp đánh giá định lượng và định tính giúp làm rõ năng lực dự báo, độ ổn định và mức độ sai lệch tại từng thời điểm trong ngày. Kết quả từ chương này sẽ cung cấp cơ sở định hướng cho việc lựa chọn mô hình phù hợp để tích hợp vào hệ thống giám sát và dự báo công suất ở các bước tiếp theo.

3.2. Mô hình học máy (đề xuất)

3.2.1. Mô hình SVR-Linear

Support Vector Regression (SVR) là một phương pháp học máy được phát triển bởi Smola và Schölkopf [39] dựa trên Support Vector Machine (SVM) của Vladimir N. Vapnik [40]. SVR là một biến thể của SVM, khác với Support Vector Machine (SVM) – vốn tập trung vào phân loại, SVR hướng tới xây dựng một mô hình hồi quy chính xác nhờ khả năng ánh xạ dữ liệu đầu vào phức tạp thành các giá trị đầu ra liên tục, đồng thời duy trì độ chính xác cao ngay cả với các tập dữ liệu phi tuyến tính. SVR tập trung vào xây dựng một mô hình hồi quy chính xác để dự đoán giá trị liên tục. SVR đạt được điều này bằng cách sử dụng một ống ε (ε -tube), trong đó sai số dự đoán giữa giá trị thực tế y và giá trị dự đoán $f(x)$ được giới hạn trong một ngưỡng cho phép ε (tương đương với độ chính xác xấp xỉ của hàm) [45].

$$|y_i - f(x_i)|_{\varepsilon} = \max(0, |y_i - f(x_i)| - \varepsilon) \quad (3.1)$$

Mô hình SVR bắt đầu từ một hàm dự đoán tuyến tính trong không gian đặc trưng:

$$f(x) = \langle \omega, x \rangle + b \quad (3.2)$$

hoặc trong không gian đặc trưng cao hơn:

$$f(x) = \langle \omega, \phi(x) \rangle + b \quad (3.3)$$

Trong đó:

- ω : vector trọng số (weight vector) trong không gian đặc trưng,
- $\phi(x)$: ánh xạ phi tuyến đưa dữ liệu đầu vào x vào không gian đặc trưng cao hơn
- b : hệ số chặn (bias term), giúp điều chỉnh vị trí của siêu phẳng phân tách (separating hyperplane) trong bài toán phân loại.

Để xây dựng mô hình, SVR giải quyết một bài toán tối ưu hóa lồi (convex optimization) nhằm tối thiểu hóa độ phức tạp của mô hình và sai số dự đoán:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^n |y_i - f(x_i)|_{\varepsilon} \quad (3.4)$$

với các ràng buộc:

- $y_i - \langle \omega, x_i \rangle - b \leq \varepsilon + \xi_i^*$

- $\langle \omega, x_i \rangle + b - y_i \leq \varepsilon + \xi_i$
- $\xi_i, \xi_i^* \geq 0$

trong đó:

- C : là tham số điều chỉnh (regularization parameter), kiểm soát sự đánh đổi giữa độ phức tạp của mô hình ($\frac{1}{2} \|\omega\|^2$) và sai số dự đoán ($\xi_i + \xi_i^*$).
- ε : là ngưỡng sai số, xác định độ rộng của ống ε (ε - tube), trong đó sai số $|y_i - f(x_i)|$ không bị phạt
- ξ_i, ξ_i^* : các biến slack (slack variables) dương (vượt quá ngưỡng trên của ống) và âm (vượt quá ngưỡng dưới của ống) của điểm dữ liệu thứ i , cho phép một số điểm dữ liệu nằm ngoài ống ε với mức phạt, đảm bảo mô hình xử lý tốt nhiều hoặc dữ liệu ngoại lai.

Kernels là khái niệm cốt lõi giúp SVR xử lý hiệu quả các bài toán có mối quan hệ phi tuyến giữa dữ liệu đầu vào và đầu ra. Kernel hoạt động bằng cách ánh xạ dữ liệu từ không gian đầu vào ban sang một không gian đặc trưng cao hơn - nơi các mối quan hệ phi tuyến có thể được biểu diễn tuyến tính. Trong SVR, thay vì tính toán trực tiếp ánh xạ mà có thể rất phức tạp trong không gian cao chiều - kernel sử dụng kernel trick để tính tích vô hướng giữa các vector trong không gian đặc trưng thông qua một hàm kernel đơn giản [45]:

$$k(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle \quad (3.5)$$

Các loại kernel phổ biến trong SVR:

- Kernel tuyến tính:

$$k(x_i, x_j) = x_i \cdot x_j \quad (3.6)$$

- Kernel đa thức (Polynomial):

$$k(x, x') = \langle x \cdot x' \rangle^d \quad (3.7)$$

- Kernel Radial Basis Function (RBF) (kernel Gaussian):

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) \quad (3.8)$$

3.2.2. Mô hình XGBoosting

XGBoost (Extreme Gradient Boosting), được phát triển bởi Tianqi Chen và Carlos Guestrin (2016) [41], là một thuật toán học máy tiên tiến dựa trên Gradient Tree Boosting, được thiết kế để giải quyết các bài toán hồi quy và phân loại với hiệu suất cao. Thuật toán này tập trung vào việc tối ưu hóa một mục tiêu học tập điều chuẩn hóa, tận dụng gradient bậc hai để tăng cường độ chính xác và hiệu quả tính toán. Với khả năng xử lý dữ liệu lớn, giảm quá khớp và tối ưu hóa hiệu suất, XGBoost là một công cụ mạnh mẽ trong học máy, phù hợp cho nhiều loại bài toán mà không bị giới hạn bởi các mục đích cụ thể như dự báo thời gian.

XGBoost xây dựng một tập hợp các cây quyết định theo cách cộng gộp, trong đó mỗi cây cải thiện giá trị đầu ra từ các cây trước đó. Giá trị dự đoán $\hat{y}_i^{(t)}$ cho mẫu x_i tại bước t được tính như sau [45]:

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (3.9)$$

trong đó:

- $f_k(x_i)$: hàm dự đoán của cây thứ k

XGBoost tối ưu hàm mục tiêu sau tại bước t :

$$G^{(t)} = \sum_{i=1}^n l(\hat{y}_i^t, y_i) + \sum_{k=1}^t \Omega(f_k) \quad (3.10)$$

trong đó:

- n : số lượng mẫu trong tập dữ liệu
- y_i : giá trị thực tế
- $l(\hat{y}_i, y_i)$: hàm mất mát
- $\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \|w\|^2$: là hàm điều chuẩn (regularization term) cho cây f_k
 - T : số lá trong cây – phản ánh độ phức tạp của cây. Một cây quyết định được chia thành các nút (nodes) và lá là các nút cuối cùng không tiếp tục phân chia.
 - w : là một vector chứa các giá trị dự đoán (weights) tương ứng với mỗi lá trong cây quyết định. Mỗi lá j (từ 1 đến T) có một giá trị w_j được tính trong quá trình huấn luyện để tối ưu hóa hàm mục tiêu.
 - γ : hệ số phạt độ phức tạp
 - λ : hệ số phạt L2-norm của w , giúp giảm nguy cơ quá khớp

Tại bước t , cây f_t được thêm vào để giảm thiểu xấp xỉ Taylor bậc hai của hàm mục tiêu:

$$G^{(t)} = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right] + \Omega(f_t) \quad (3.11)$$

trong đó:

- $g_i = \partial_{y_i^{(t-1)}} L(y_i, \hat{y}_i^{(t-1)})$: gradient bậc một.
- $h_i = \partial_{y_i^{(t-1)}}^2 L(y_i, \hat{y}_i^{(t-1)})$: gradient bậc hai.

3.3. Tiêu chí đánh giá: RMSE, MAE, R²

Trong đề án này, chúng tôi sử dụng 3 chỉ số phổ biến để đánh giá các mô hình học máy [42][43][45].

RMSE (Root Mean Square Error - Sai số bình phương trung bình) là một chỉ số đo lường trung bình của sai số bình phương giữa giá trị dự đoán và giá trị thực tế, sau đó lấy căn bậc hai của kết quả. Nó nhấn mạnh các sai số lớn hơn (outliers) do việc bình phương các sai số, làm cho nó trở thành một chỉ số nhạy cảm với các giá trị ngoại lệ.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (H_{i,m} - H_{i,c})^2} \quad (3.12)$$

trong đó:

- n : số lượng quan sát
- $H_{i,c}$: giá trị dự đoán
- A : giá trị thực tế

RMSE cung cấp một thước đo tổng quát về độ lệch giữa giá trị dự đoán và thực tế, nhưng nó phóng đại các sai số lớn do việc bình phương. Giá trị *RMSE* càng nhỏ, mô hình càng chính xác, với giá trị lý tưởng là 0. *RMSE* thường được sử dụng để đánh giá hiệu suất ngắn hạn của mô hình và cung cấp cái nhìn toàn diện về phân phối sai số. 1

MAE (Mean Absolute Error - Sai số tuyệt đối trung bình) là một chỉ số đo lường trung bình của sai số tuyệt đối giữa giá trị dự đoán (calculated/estimated values) và giá trị thực tế (measured/actual values). Nó cung cấp một cách đơn giản để đánh giá mức độ chính xác của mô hình dự đoán mà không bị ảnh hưởng bởi dấu của sai số (dương hay âm).

$$MAE = \frac{1}{n} \sum_{i=1}^n |H_{i,c} - H_{i,m}| \quad (3.13)$$

trong đó:

- n : số lượng quan sát
- $H_{i,c}$: giá trị dự đoán
- $H_{i,m}$: giá trị thực tế

Giá trị MAE càng nhỏ, mô hình càng chính xác. Giá trị lý tưởng là 0, nghĩa là không có sai số giữa giá trị dự đoán và thực tế. Không bị ảnh hưởng bởi các giá trị ngoại lệ (outliers) vì nó chỉ lấy giá trị tuyệt đối, không nâng bậc (squaring) như $RMSE$. Phù hợp để so sánh hiệu suất giữa các mô hình khác nhau trên cùng một tập dữ liệu.

R^2 (Coefficient of Determination - Hệ số xác định) là một chỉ số thống kê đo lường mức độ mà mô hình dự đoán có thể giải thích sự biến thiên của dữ liệu thực tế. Nó cho biết phần trăm biến thiên trong giá trị thực tế được mô hình dự đoán giải thích.

$$R^2 = 1 - \frac{\sum_{i=1}^n (H_{i,m} - H_{i,c})}{\sum_{i=1}^n (H_{i,m} - H_{m,ave})^2} n \quad (3.14)$$

trong đó:

- n : số lượng quan sát
- $H_{i,m}$: giá trị thực tế
- $H_{m,ave}$: trung bình của giá trị thực tế
- $H_{i,c}$: giá trị dự đoán

R^2 có giá trị từ 0 đến 1. Giá trị R^2 càng gần 1, mô hình càng tốt trong việc giải thích dữ liệu thực tế. Giá trị $R^2 = 0$ cho thấy mô hình không giải thích được sự biến thiên của dữ liệu, trong khi $R^2 = 1$ chỉ ra dự đoán hoàn hảo. Kết quả thực nghiệm và thảo luận

3.4. Huấn luyện và đánh giá mô hình

3.4.1. Kết quả huấn luyện mô hình học máy

a) Mô hình SVR-RBF

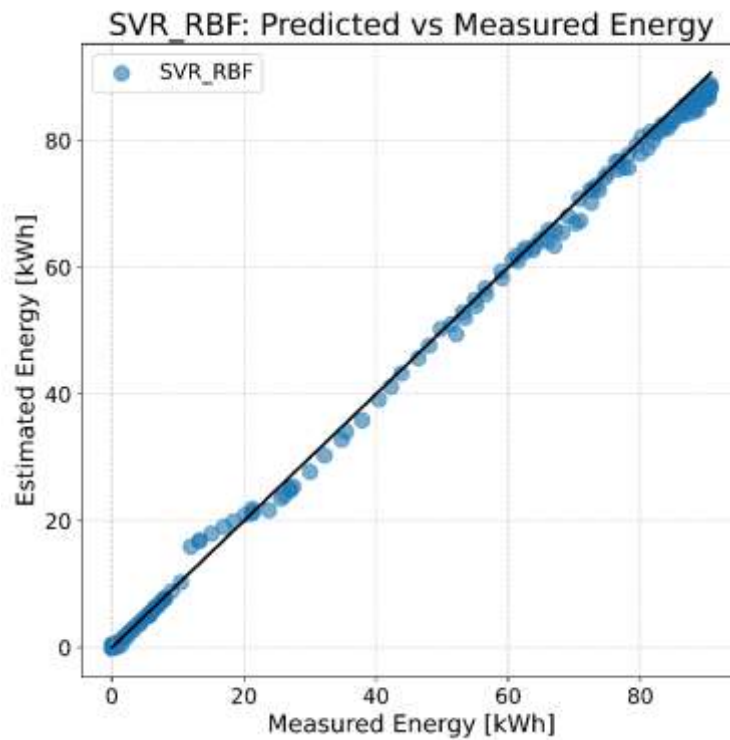
Mô hình Support Vector Regression (SVR) sử dụng hàm kernel RBF được áp dụng nhằm dự báo sản lượng điện của hệ thống quang điện. Sau quá trình huấn luyện và tinh chỉnh, mô hình đạt được các chỉ số đánh giá chất lượng tốt, gồm: sai số bình phương trung bình (MSE) là **1.5266**, sai số căn phương bình phương trung bình (RMSE) đạt **1.2356 kWh**, và hệ số xác định (R^2) ở mức **0.9988**. Các chỉ số này cho thấy mô hình có khả năng khái quát tốt và tái hiện được mối quan hệ phi tuyến giữa đầu vào và đầu ra với độ chính xác cao.

Hiệu suất tổng thể của mô hình được thể hiện rõ trong Hình 3.1, là biểu đồ phân tán giữa sản lượng điện thực tế và giá trị dự báo. Các điểm dữ liệu phân bố tập trung sát đường chéo lý tưởng $y=x$, cho thấy mức độ tương đồng cao giữa hai tập giá trị. Đây là dấu hiệu cho thấy mô hình học được tốt xu hướng tổng thể, đặc biệt trong các khoảng giá trị công suất cao – nơi sai số có xu hướng lớn hơn ở nhiều mô hình khác.

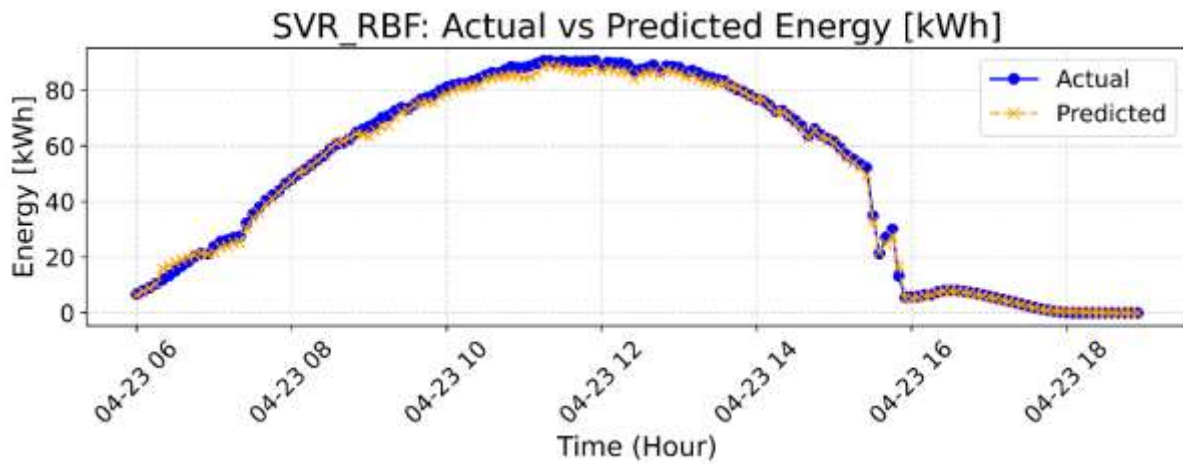
Chi tiết theo chuỗi thời gian được trình bày trong Hình 3.2, với trục hoành là thời gian (ngày 23/04) và trục tung là sản lượng điện theo đơn vị kWh. Mô hình SVR-RBF bám sát gần hoàn toàn với đường thực tế từ lúc mặt trời bắt đầu chiếu sáng đến khi kết thúc ngày. Trong khoảng thời gian từ 08:00 đến 15:00 – khi sản lượng biến động mạnh và đạt đỉnh – mô hình vẫn duy trì được khả năng dự báo ổn định và chính xác.

Phân tích sâu hơn về sai số tại từng thời điểm được thể hiện trong Hình 3.3, là biểu đồ thể hiện độ lệch (chênh lệch có dấu) giữa sản lượng dự báo và giá trị thực tế. Sai số chủ yếu dao động trong khoảng ± 2 kWh, tuy nhiên có một số điểm đột biến nhỏ vào đầu giờ sáng và cuối buổi chiều, với đỉnh lệch lớn nhất xấp xỉ ± 4 kWh. Những thời điểm này trùng với giai đoạn chuyển tiếp bức xạ mạnh – vốn là thách thức với nhiều mô hình dự báo. Dù vậy, phần lớn thời gian trong ngày, sai số dao động xung quanh giá trị 0 và không xuất hiện hiện tượng sai lệch tích lũy.

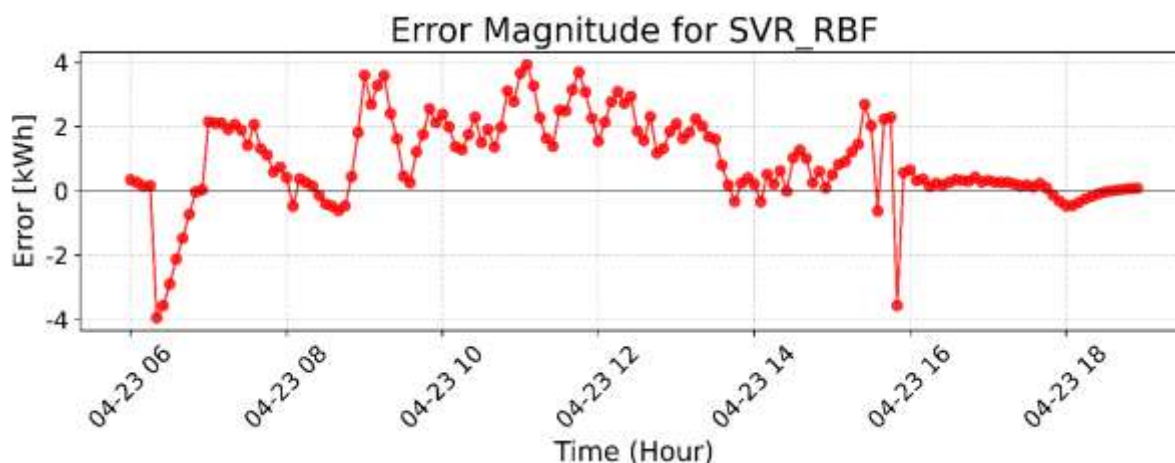
Nghiên cứu và triển khai hệ thống thu thập dữ liệu của điện mặt trời mái nhà dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất



Hình 3.1. Biểu đồ phân tán giữa sản lượng thực tế và sản lượng dự báo của mô hình SVR-RBF



Hình 3.2. So sánh sản lượng điện đo được và dự báo theo thời gian trong ngày 23/04



Hình 3.3. Độ lệch dự báo giữa dự báo và thực tế theo thời gian ($Error = Prediction - Actual$)

Tóm lại, mô hình SVR-RBF chứng minh được hiệu suất cao thông qua các chỉ số định lượng cũng như biểu đồ trực quan. Mô hình không chỉ có khả năng tái hiện tốt xu hướng sản lượng điện theo thời gian mà còn duy trì độ chính xác cao ở hầu hết các khoảng thời gian trong ngày. Điều này cho thấy SVR-RBF là một lựa chọn đáng tin cậy cho bài toán dự báo sản lượng ngắn hạn trong hệ thống điện mặt trời.

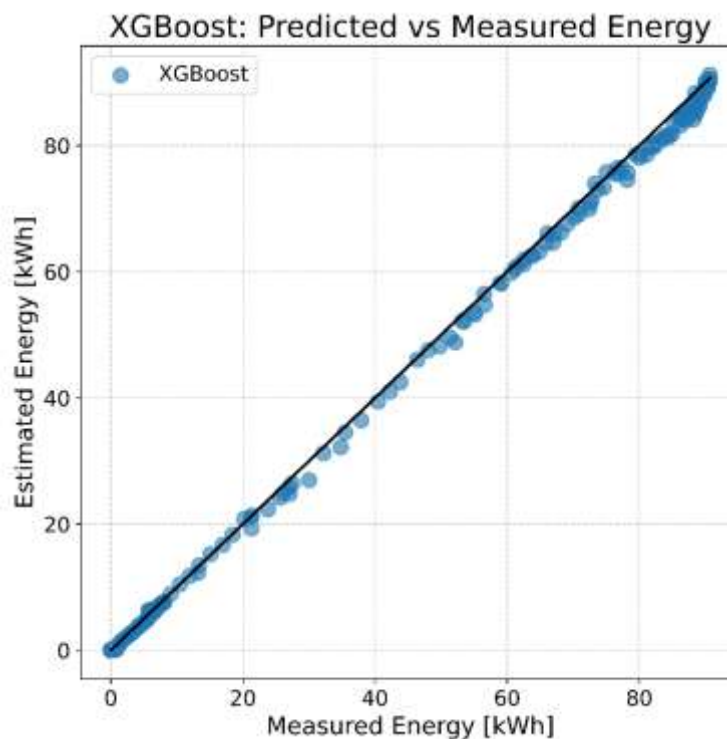
b) Mô hình XGBoosting

Thuật toán XGBoost, một biến thể mạnh mẽ của phương pháp boosting dựa trên cây quyết định, được triển khai nhằm dự báo sản lượng điện đầu ra của hệ thống quang điện. Với đặc điểm nổi bật là khả năng học sâu mối quan hệ phi tuyến qua quá trình huấn luyện lặp lại và điều chỉnh trọng số lỗi, mô hình đã thể hiện hiệu suất dự báo rất tốt. Cụ thể, giá trị sai số bình phương trung bình (MSE) đạt 1.2456, sai số căn phương bình phương trung bình (RMSE) là 1.1161 kWh, trong khi hệ số xác định (R^2) đạt mức 0.9990. Đây đều là những chỉ số thể hiện mô hình có độ chính xác cao, với sai số thấp và khả năng giải thích phương sai dữ liệu rất mạnh.

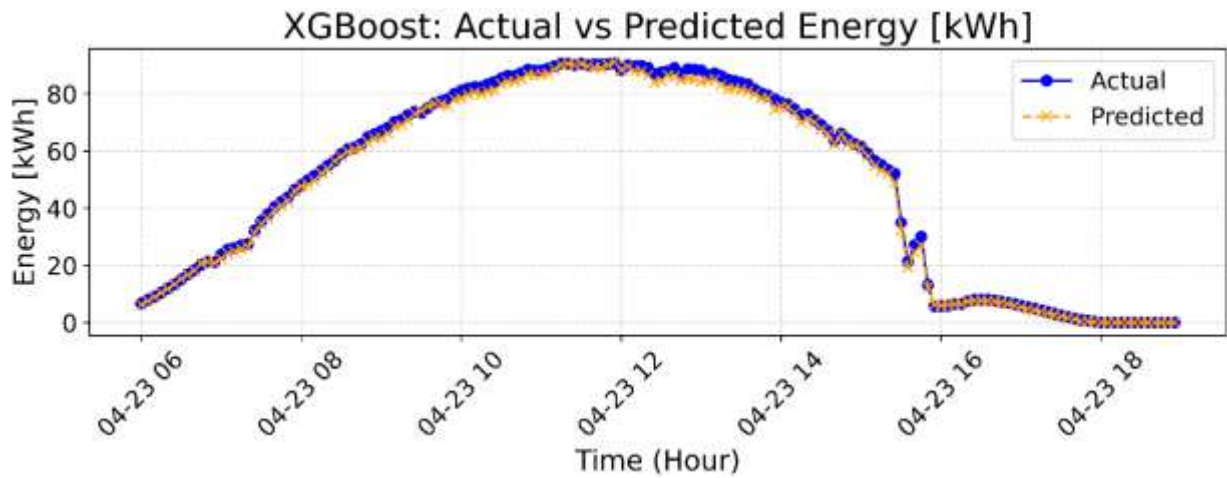
Biểu đồ phân tán thể hiện trong **Hình 3.4** mô tả mối quan hệ giữa sản lượng đo được và sản lượng dự báo. Các điểm dữ liệu nằm sát đường chéo $y=x$, phân bố đồng đều từ vùng công suất thấp đến vùng cao. Điều này cho thấy mô hình không chỉ duy trì độ chính xác ở vùng dữ liệu tập trung mà còn thể hiện tính ổn định trên toàn miền giá trị – điều thường không dễ đạt được ở các mô hình hồi quy phi tuyến.

Chi tiết hơn về biến thiên theo thời gian được thể hiện trong **Hình 2**, với sản lượng điện dự báo (nét đứt màu cam) và thực tế (đường liền màu xanh) được hiển thị theo giờ trong ngày. Mô hình XGBoost bám sát gần như hoàn toàn đường cong thực tế trong toàn bộ chu kỳ chiếu sáng. Đặc biệt, trong khoảng thời gian từ 10:00 đến 14:00 – giai đoạn mà công suất đạt đỉnh và thay đổi nhanh – mô hình vẫn duy trì sai số nhỏ và không có hiện tượng dao động ngẫu nhiên hay vượt đỉnh.

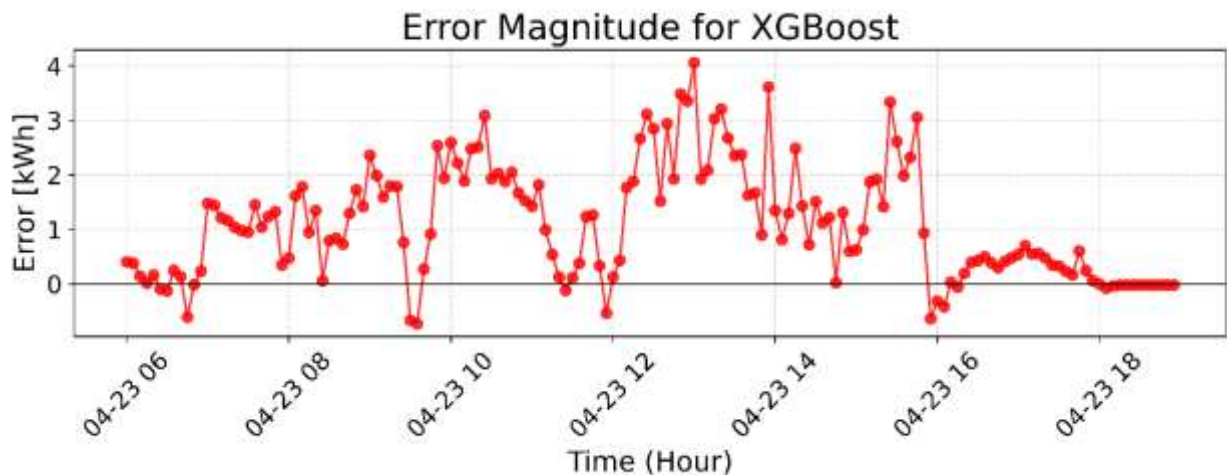
Hình 3 cung cấp thông tin chi tiết về sai số dự báo theo thời gian. Sai số dao động chủ yếu trong khoảng từ 0 đến 2 kWh, với một số đỉnh cục bộ không vượt quá 4 kWh. So với SVR-RBF, mô hình này cho thấy sai số ổn định hơn trong phần lớn khung giờ ban ngày. Ngoài ra, mức sai lệch gần như tiệm cận 0 trong giai đoạn cuối ngày cho thấy mô hình không mắc lỗi tích lũy và kiểm soát tốt trong các giai đoạn công suất thấp.



Hình 3.4. Biểu đồ phân tán giữa sản lượng điện thực tế và dự báo từ mô hình XGBoost



Hình 3.5. So sánh diễn biến công suất đo được và dự báo trong ngày 23/04



Hình 3.6. Độ lệch giữa giá trị dự báo và thực tế theo thời gian ($Error = Prediction - Actual$)

Tổng kết lại, XGBoost thể hiện hiệu suất vượt trội trên cả ba chỉ số đánh giá. Mô hình không chỉ dự báo chính xác mà còn ổn định trong các điều kiện vận hành thay đổi nhanh. Với khả năng thích ứng cao và mức sai số nhỏ, XGBoost là một ứng viên mạnh mẽ cho các hệ thống dự báo công suất ngắn hạn, đặc biệt trong môi trường năng lượng tái tạo có nhiều biến động khó lường.

3.5. Kết luận

Trong chương này, hai mô hình học máy là SVR-RBF và XGBoost đã được xây dựng, huấn luyện và đánh giá nhằm dự báo sản lượng điện của hệ thống quang điện theo chuỗi

thời gian. Cả hai mô hình đều thể hiện năng lực dự báo tốt, với sai số thấp và hệ số xác định R^2 tiệm cận 1, cho thấy khả năng mô hình hóa chính xác mối quan hệ giữa các biến đầu vào và đầu ra năng lượng.

Cụ thể, SVR-RBF cho thấy khả năng theo dõi biến thiên sản lượng theo thời gian rất sát với dữ liệu thực tế, đặc biệt trong các khung giờ cao điểm. Tuy nhiên, sai số tại một số thời điểm chuyển tiếp (sáng sớm và chiều muộn) vẫn còn dao động nhẹ. Trong khi đó, XGBoost thể hiện hiệu suất vượt trội hơn ở cả ba tiêu chí đánh giá, đồng thời duy trì độ ổn định cao hơn trong toàn bộ chu kỳ vận hành. Mô hình này cũng kiểm soát tốt sai số tại các thời điểm biên và không xuất hiện hiện tượng lệch tích lũy.

Từ kết quả phân tích, XGBoost được đánh giá là lựa chọn tối ưu hơn trong bối cảnh dự báo sản lượng ngắn hạn của hệ thống điện mặt trời, đặc biệt dưới điều kiện thời tiết thay đổi nhanh. Những đánh giá này đóng vai trò quan trọng để lựa chọn mô hình phù hợp cho triển khai thực tế trong các hệ thống giám sát và điều hành năng lượng tái tạo.

Chương 4: XÂY DỰNG HỆ THỐNG IOT VÀ TÍCH HỢP MÔ HÌNH ƯỚC LƯỢNG

4.1. Giới thiệu

Chương này tập trung vào việc trình bày chi tiết quá trình thiết kế, xây dựng và triển khai hệ thống Internet vạn vật (IoT) phục vụ cho việc thu thập dữ liệu và giám sát hiệu suất của hệ thống điện mặt trời mái nhà. Bên cạnh đó, chương cũng sẽ mô tả cách thức tích hợp các mô hình học máy đã được nghiên cứu và huấn luyện ở chương trước vào kiến trúc IoT này để thực hiện dự báo công suất đầu ra theo thời gian thực.

4.2. Kiến trúc hệ IoT

4.2.1. Kiến trúc phần cứng

a) Kiến trúc tổng thể

Trong các hệ thống điện mặt trời công suất lớn, thiết bị biến tần thường đi kèm với một thành phần phụ trợ gọi là SmartLogger. Thiết bị này có nhiệm vụ thu thập dữ liệu, thực hiện giao tiếp với biến tần thông qua các giao thức nối tiếp như RS-485. SmartLogger được kết nối với Internet bằng phương thức có dây hoặc không dây, phổ biến nhất là qua bộ định tuyến 4G sử dụng SIM di động. Cấu hình này được định danh là hệ thống con (Subsystem 1 – Sub1). Với kiến trúc truyền thông như vậy, người dùng có thể giám sát các thông số điện và trạng thái vận hành của hệ thống thông qua ứng dụng chuyên dụng do nhà sản xuất biến tần cung cấp.

Nhằm mở rộng năng lực giám sát dựa trên hạ tầng thu thập dữ liệu sẵn có, nghiên cứu này đề xuất tích hợp thêm một hệ thống con bổ sung – gọi là Sub2 – chuyên thu thập dữ liệu môi trường theo thời gian thực. Phân hệ này sẽ ghi nhận các thông số như bức xạ mặt trời, nhiệt độ môi trường, độ ẩm không khí và nhiệt độ bề mặt tấm pin. Thành phần xử lý trung tâm của Sub2 là một máy tính nhúng Raspberry Pi, được kết nối với bộ định tuyến 4G qua giao thức TCP/IP sử dụng cổng Ethernet. Như vậy, Sub1 và Sub2 cùng chia sẻ một kết nối mạng duy nhất thông qua bộ định tuyến 4G, trong khi toàn bộ các thiết bị của hai

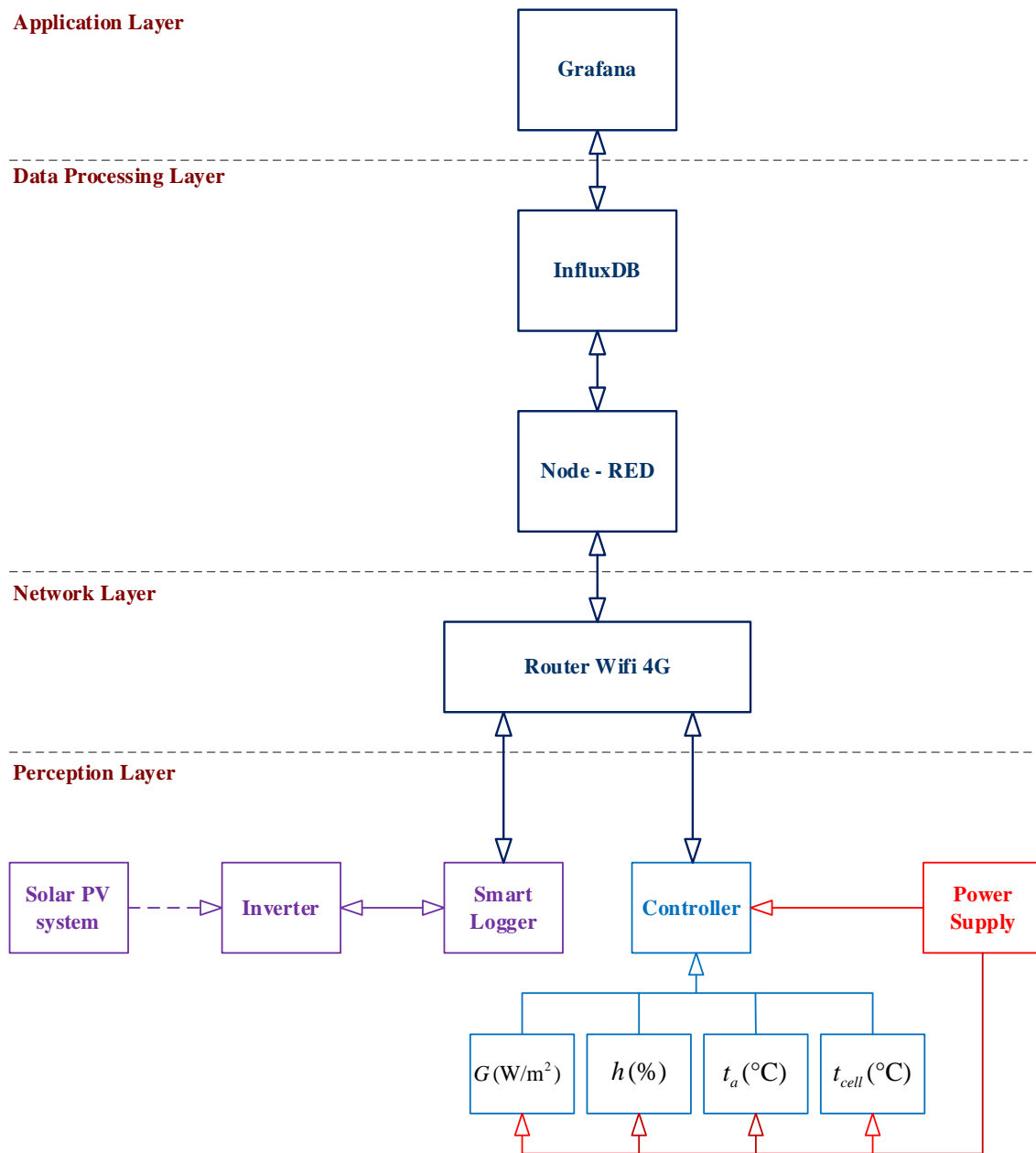
phân hệ – ngoại trừ bộ định tuyến – cấu thành lớp cảm nhận (perception layer) trong kiến trúc hệ thống IoT.

Raspberry Pi trong cấu trúc này được định cấu hình như một máy chủ cục bộ, đảm nhiệm đồng thời hai chức năng chính: (i) thu thập dữ liệu từ các cảm biến thời tiết, và (ii) truy xuất thông tin điện năng từ SmartLogger. Cả hai luồng dữ liệu – dữ liệu môi trường và dữ liệu điện – được xử lý và lưu trữ tại chỗ, sau đó được truyền đến nền tảng Node-RED bằng giao thức MQTT – một chuẩn giao tiếp nhẹ và phổ biến trong các ứng dụng IoT. Các nghiên cứu trước đã chỉ ra rằng MQTT là giải pháp phù hợp để truyền dữ liệu cảm biến theo thời gian thực với độ trễ thấp và tài nguyên yêu cầu tối thiểu. Do đó, hệ thống đề xuất áp dụng cùng phương pháp nhằm đảm bảo tính kết nối ổn định và hiệu quả truyền thông.

Bên cạnh đó, để hạn chế tác động tiêu cực của nhiễu điện từ và đảm bảo độ chính xác khi truyền dữ liệu trên khoảng cách dài, nghiên cứu lựa chọn sử dụng cảm biến thông minh có giao tiếp số (serial digital sensors) thay cho các loại cảm biến analog truyền thống. Điều này cho phép tăng độ chính xác và độ tin cậy khi tích hợp vào hệ thống.

Tổng thể, kiến trúc hệ thống được xây dựng theo mô hình phân lớp IoT, bao gồm: lớp cảm nhận (perception layer), lớp mạng (network layer), lớp xử lý dữ liệu (data processing layer) và lớp ứng dụng (application layer). Sơ đồ khối mô tả tổng thể giải pháp giám sát điện mặt trời dựa trên nền tảng IoT được minh họa tại Hình 4.1.

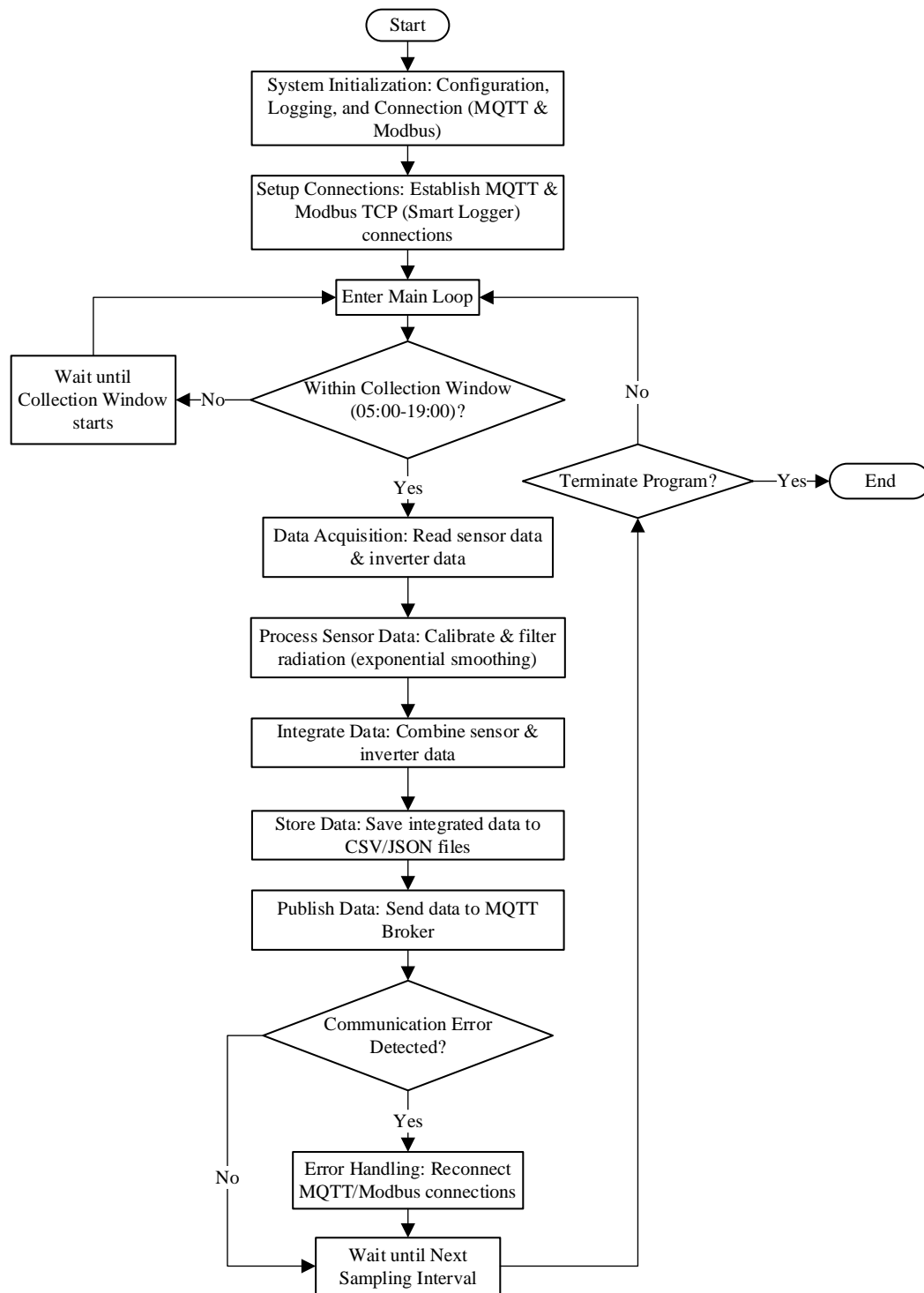
Nghiên cứu và triển khai hệ thống thu thập dữ liệu của điện mặt trời mái nhà dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất



Hình 4.1. Sơ đồ khối của giải pháp IoT được đề xuất [44]

Hình mô tả quy trình thu thập và xử lý dữ liệu trong hệ thống này được mô tả trong Hình 4.2

Nghiên cứu và triển khai hệ thống thu thập dữ liệu của điện mặt trời mái nhà dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất



Hình 4.2. Lưu đồ thuật toán của giải pháp dựa trên IoT được đề xuất [44]

b) Phần cứng sử dụng

Dựa trên kiến trúc đề xuất, chúng tôi triển khai các thiết bị phần cứng tương ứng. kiến trúc hệ thống với các thiết bị phần cứng được mô tả trong Hình 4.5.

● Bộ điều khiển Raspberry Pi 4B

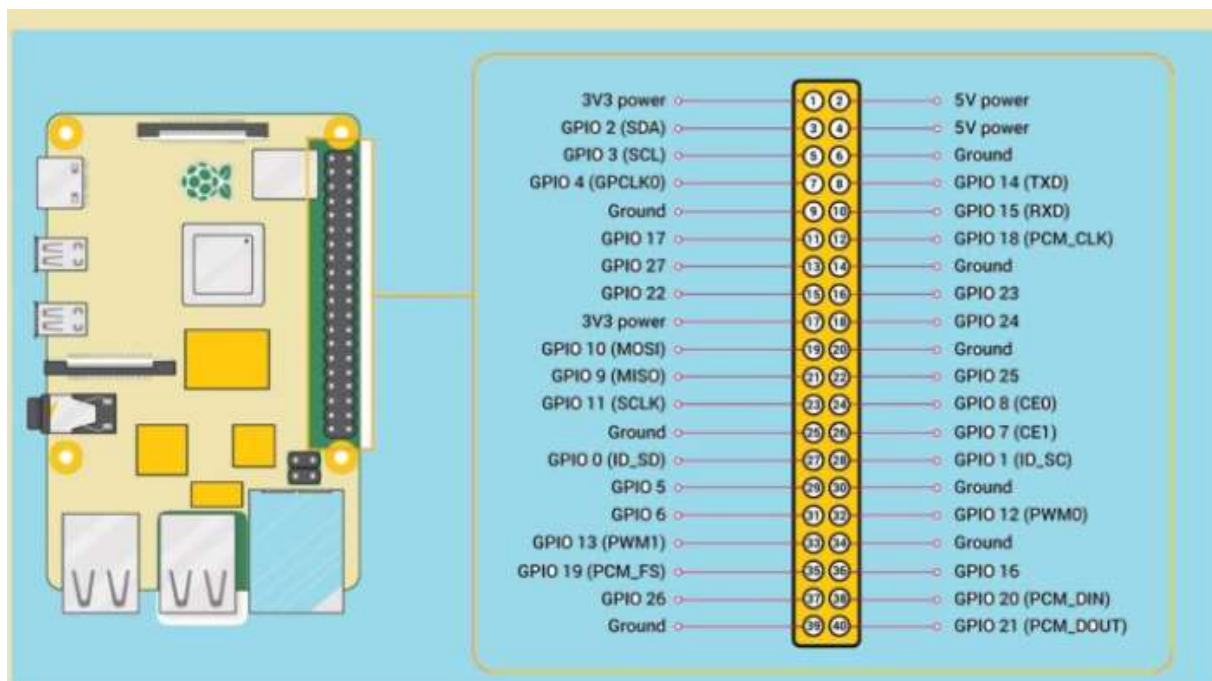
Raspberry Pi 4 Model B (Pi 4B) là bo mạch máy tính đơn (single-board computer) thế hệ mới của Raspberry Pi Foundation. So với các phiên bản trước, Pi 4B nâng cấp mạnh mẽ về CPU, GPU và khả năng I/O, đồng thời hỗ trợ tối đa 8 GB RAM LPDDR4.



Hình 4.3. Raspberry Pi 4B

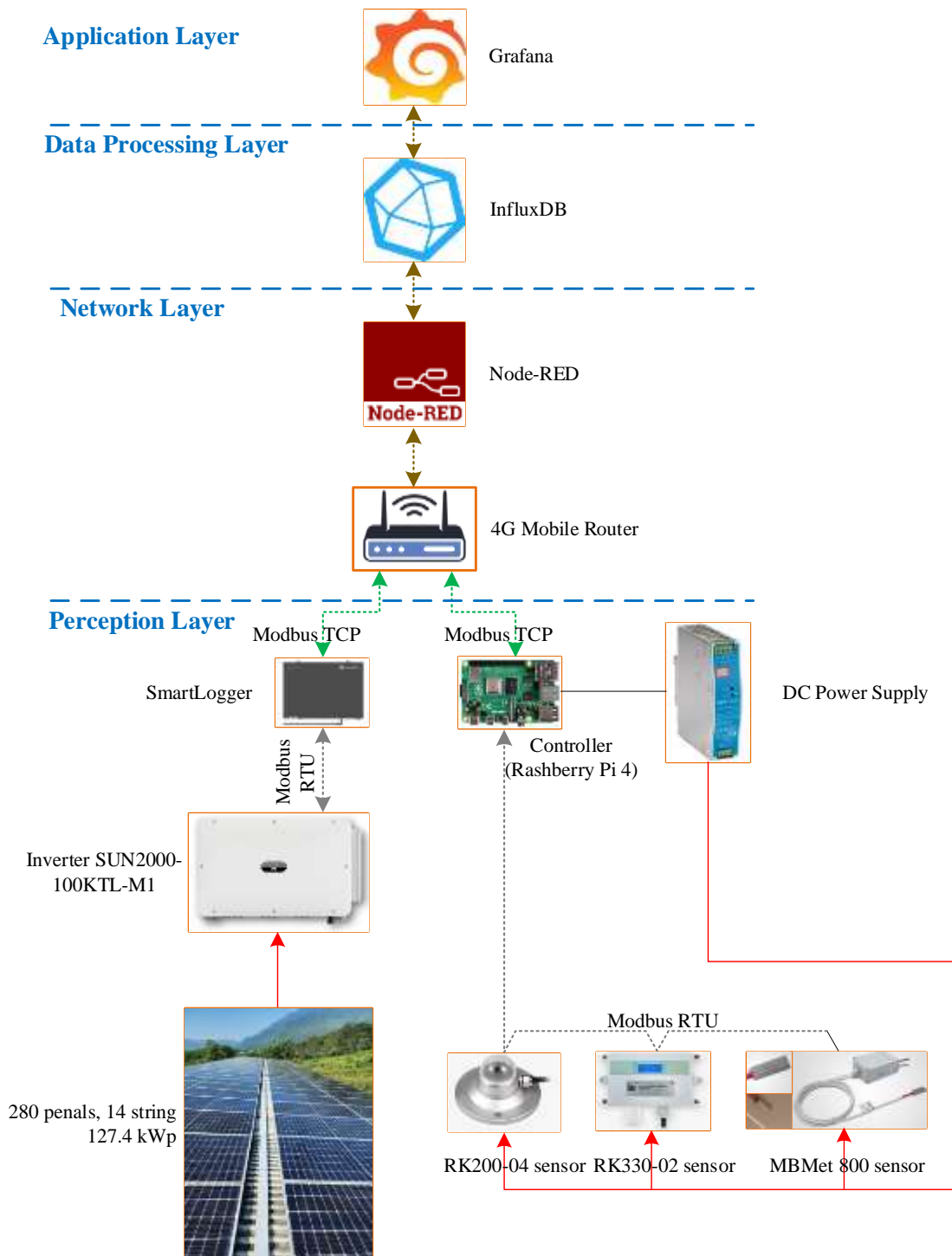
❖ Cấu trúc chân

Pi 4B tích hợp một header 40 chân chuẩn (2 × 20) hoàn toàn tương thích với các mẫu Pi trước. Các chân này hỗ trợ đa dạng chức năng: GPIO số, giao tiếp I²C/SPI/UART, tín hiệu PWM, nguồn và mass (GND)



Hình 4.4. Giới thiệu chân Raspberry Pi4

Nghiên cứu và triển khai hệ thống thu thập dữ liệu của điện mặt trời mái nhà dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất



Hình 4.5. Cấu hình thực nghiệm của kiến trúc IoT được đề xuất [44]

- ❖ Power & GND: Chân 1 (3V3), 2 (5V), 4 (5V), 6 (GND), 9 (GND), 14 (GND), 20 (GND), 25 (GND), 30 (GND), 34 (GND), 39 (GND).

Trong đó, 3 chân cấp 3,3 V (Chân 1), 2 chân cấp 5 V (Chân 2, 4) và 8 chân GND phân tán để cấp nguồn và tản nhiệt.

Bảng 4.1. GPIO và pin của Raspberry Pi 4B

GPIO0 (BCM 17), Pin 11	GPIO1 (BCM 18), Pin 12
GPIO2 (BCM 27), Pin 13	GPIO3 (BCM 22), Pin 15
GPIO4 (BCM 23), Pin 16	GPIO5 (BCM 24), Pin 18
GPIO6 (BCM 25), Pin 22	GPIO7 (BCM 4), Pin 7
GPIO8 (BCM 2), Pin 3	GPIO9 (BCM 3), Pin 5
GPIO10 (BCM 8), Pin 24	GPIO12 (BCM 10), Pin 23
GPIO11 (BCM 7), Pin 26	GPIO14 (BCM 11, TXD0), Pin 8
GPIO13 (BCM 9), Pin 21	GPIO16 (BCM 12), Pin 32
GPIO15 (BCM 10, RXD0), Pin 10	GPIO18 (BCM 19), Pin 35
GPIO17 (BCM 13), Pin 33	GPIO20 (BCM 26), Pin 38
GPIO19 (BCM 16), Pin 36	GPIO22 (BCM 21), Pin 40
GPIO21 (BCM 20), Pin 38	

- Giao tiếp I2C

I2C₀: SDA (BCM 2, Pin 3), SCL (BCM 3, Pin 5)

I2C₁ (dành cho camera): SDA¹ (BCM 0, Pin 27), SCL¹ (BCM 1, Pin 28)

- Giao tiếp SPI

- ❖ SPI0

- CE0 (BCM 8, Pin 24)
- CE1 (BCM 7, Pin 26)
- MISO (BCM 9, Pin 21)

- MOSI (BCM 10, Pin 19)
- SCLK (BCM 11, Pin 23)
- ❖ SPI1: nhiều kênh hơn, bản chất sử dụng GPIO 16–21 khi cấu hình alternates functions

Bảng 4.2. Thông số Raspberry Pi 4B

Hạng mục	Thông số
Hệ xử lý (SoC)	Broadcom BCM2711, Quad-core ARM Cortex-A72 64-bit @ 1,5 GHz
Bộ nhớ	LPDDR4 SDRAM 1 GB / 2 GB / 4 GB / 8 GB (1 x LPDDR4, 3200 MT/s)
Đồ họa	GPU VideoCore VI 3D, hỗ trợ H.265 4Kp60, H.264 1080p60
Kết nối mạng	Gigabit Ethernet (RJ45, qua USB 2.0 đến ≈ 300 Mbps) / Wi-Fi 802.11 b/g/n/ac (2,4 GHz & 5 GHz) / Bluetooth 5.0 BLE
Cổng USB	2 × USB 3.0; 2 × USB 2.0
Video & Âm thanh	2 × micro-HDMI (4Kp60 dual-display) / 3,5 mm TRRS A/V (analog audio + composite video)
Lưu trữ	Thẻ microSD (máy chủ boot & lưu trữ OS) / SDIO qua GPIO (phụ trợ)
GPIO Header	40 pins (2 × 20): 26 × GPIO số, 2 × I ² C, 5 × SPI, 1 × UART, 2 × PWM, 3 × 3,3 V, 2 × 5 V, 8 × GND
Nguồn cấp	USB-C 5 V = 3 A (khuyến nghị) / 5 V = 2,5 A (khi không sử dụng USB 3.0 nhiều)
Tiêu thụ điện	< 7,6 W (idle) – < 15 W (tải nặng & USB 3.0 full load)
Wi-Fi & Bluetooth	Chuẩn 802.11 b/g/n/ac (dual-band) / Bluetooth 5.0 (BLE)

Kết nối camera & màn hình	1 × MIPI CSI 2-lane (camera), 1 × MIPI DSI 2-lane (display)
Âm thanh & Video composite	Analog stereo + composite video out qua jack 4-ring TRRS
Nhiệt độ hoạt động	0 °C ... +50 °C (khuyến nghị); có thể hoạt động đến +85 °C tùy làm mát; bo mạch tự throttle CPU nếu > 85 °C)
Độ ẩm	10 % ... 90 % RH (không ngưng tụ)
Cơ khí	Kích thước: 88 mm × 58 mm × 19,5 mm; Trọng lượng: ≈ 46 g
Chứng nhận & Tiêu chuẩn	CE, FCC, RoHS, WEEE

● Inverter

SUN2000-100KTL-M1 là bộ biến tần (inverter) dạng String của Huawei, công suất 100 kW, được thiết kế đặc biệt cho các hệ thống điện mặt trời thương mại và công nghiệp. Thiết bị hỗ trợ 10 MPPT độc lập, cho phép tối ưu hoá thu năng lượng từ các chuỗi tấm pin khác nhau. Với hiệu suất tối đa đạt 98,8 % và khả năng chịu điều kiện môi trường khắc nghiệt (IP66), SUN2000-100KTL-M1 đảm bảo vận hành ổn định, bảo vệ an toàn và dễ dàng giám sát thông qua RS-485, MBUS hoặc kết nối không dây qua Bluetooth/WLAN.



Hình 4.6. Inverter SUN2000-100KTL-M1

Bảng 4.3. Thông số kỹ thuật của Inverter SUN2000-100KTL-M1

Thông số kỹ thuật chính	Giá trị
Công suất định mức (AC)	100 000 W (ở 380 V / 400 V / 480 V, 40 °C)
Công suất cực đại (AC, $\cos \varphi=1$)	110 000 W
Hiệu suất tối đa	98,8 % @ 480 V; 98,6 % @ 380 V / 400 V
Hiệu suất châu Âu	98,6 % @ 480 V; 98,4 % @ 380 V / 400 V
Số MPPT	10 MPPT độc lập
Điện áp khởi động MPPT	200 V
Phạm vi điện áp hoạt động MPPT	200 V ~ 1 000 V
Điện áp đầu vào danh định	600 V @ 380 V / 400 V; 720 V @ 480 V
Điện áp đầu vào tối đa	1 100 V
Dòng ngắn mạch tối đa mỗi MPPT	40 A
Dòng đầu vào tối đa mỗi MPPT	26 A
Số chuỗi đầu vào (strings)	20
Điện áp đầu ra danh định	380 V / 400 V / 480 V (3P + PE); 220 V / 230 V (3P + N + PE)

Tần số lưới	50 Hz / 60 Hz
Dòng đầu ra danh định	152,0 A @ 380 V; 144,4 A @ 400 V; 120,3 A @ 480 V
Dòng đầu ra tối đa	168,8 A @ 380 V; 160,4 A @ 400 V; 133,7 A @ 480 V
Độ sét hài tổng (THD)	< 3 %
Điều chỉnh hệ số công suất (cos φ)	0,8 (đấu ngược) ... 0,8 (đấu xuôi)
Chỉ số bảo vệ (IP)	IP66
Nhiệt độ hoạt động	-25 °C ... 60 °C
Độ ẩm tương đối	0 ... 100 %
Độ cao vận hành tối đa	5 000 m (giảm công suất khi trên 4 000 m)
Phương pháp làm mát	Quạt cưỡng bức (Forced Air Cooling)
Cấp độ cách điện và bảo vệ đảo ngược cực	Bảo vệ chống đảo ngược cực DC, chống quá dòng AC, chống đảo lưới
Giám sát chuỗi pin	Có cảnh báo lỗi chuỗi (String Fault Monitoring)
Thiết bị chống sét	SPD Loại II cho cả DC và AC
Giao tiếp	LED hiển thị trạng thái; RS-485; MBUS (cần biến áp cách ly); USB; WLAN/Bluetooth + App Huawei SmartLogger
Kích thước (R x C x D)	1 035 mm × 700 mm × 365 mm
Trọng lượng (bao gồm tấm gắn)	85 kg
Tiêu chuẩn	IP66, không dùng biến áp (Transformerless)

● Smartlogger 3000A

SmartLogger3000A là bộ thu thập và quản lý dữ liệu (gateway) của Huawei, dùng trong hệ thống điện mặt trời. Thiết bị cho phép kết nối tối đa 80 thiết bị (inverter, cảm biến, bộ giám sát...) qua các giao thức Modbus-RTU, Modbus-TCP và IEC 60870. SmartLogger3000A hỗ trợ cả kết nối có dây (Ethernet, RS-485, MBUS) và không dây (2G/3G/4G, WLAN), với khả năng giám sát, quản lý từ xa qua web hoặc app. Thiết kế chịu được điều kiện khắc nghiệt (IP20 bên trong tủ, hoạt động -40 °C đến 60 °C) và có tùy chọn gắn tường, thanh DIN hoặc để bàn



Hình 4.7. Smartlogger 3000A

Bảng 4.4. Thông số kỹ thuật của Smartlogger 3000A

Thông số cơ bản	Giá trị
Số lượng thiết bị kết nối tối đa	80 thiết bị
Giao tiếp WAN/LAN	1 cổng WAN, 1 cổng LAN (10/100/1000 Mbps)
Cổng RS-485	3 cổng (1 200–115 200 bps, 1 000 m)
Cổng MBUS	1 cổng (max AC 800 V, 1 000 m)
Kết nối 2G/3G/4G	LTE, HSPA+, GSM/GPRS/EDGE (đa băng tần)
Ngõ vào/ra số tương tự và số	DI x 4, DO x 2 (12 V/100 mA), AI x 4
Giao thức hỗ trợ	Modbus-TCP, IEC 60870-5-104, Modbus-RTU, IEC 60870-5-103, DL/T 645

Giao diện tương tác	LED (RUN, ALM, 4G), web nhúng, USB 2.0, WLAN (cho cài đặt)
Nhiệt độ hoạt động	-40 °C ... 60 °C
Độ ẩm tương đối	5 % ... 95 % (không ngưng tụ)
Độ cao tối đa	4 000 m
Nguồn nuôi	AC 100–240 V (50/60 Hz); DC 12 V/24 V
Công suất tiêu thụ	Thường 8 W, tối đa 15 W
Kích thước (RxCxD)	225 × 160 × 43,6 mm
Trọng lượng	2 kg
Cấp bảo vệ	IP20
Lắp đặt	Gắn tường, thanh DIN, để bàn

● **Gateway router wifi 4G**

Gateway router WiFi 4G của Huawei là thiết bị kết hợp chức năng định tuyến (router) và điểm truy cập không dây (access point), đồng thời hỗ trợ kết nối Internet qua mạng di động 4G LTE. Thiết bị thường được trang bị khe cắm SIM để hút sóng LTE, phát WiFi chuẩn 802.11ac hoặc 802.11n, kèm cổng Ethernet RJ45 để kết nối trực tiếp với máy tính hoặc switch. Với khả năng tự động chuyển đổi giữa đường truyền 4G và WAN có dây (nếu có), thiết bị đảm bảo kết nối Internet ổn định cho văn phòng nhỏ, cửa hàng hoặc các công trình di động.



Bảng 4.5. Thông số kỹ thuật Router wifi 4G

Hình 4.8. Gateway router WiFi 4G

Thông số cơ bản	Giá trị điển hình
-----------------	-------------------

Chuẩn mạng di động	LTE Cat 6 (tốc độ tải xuống đến 300 Mbps, tải lên 50 Mbps)
Băng tần LTE hỗ trợ	B1/B3/B7/B8/B20/B28 (tùy thị trường)
Khe SIM	1 khe SIM (Nano-SIM)
Chuẩn WiFi	IEEE 802.11 ac/n/g/b (dual-band 2.4 GHz & 5 GHz)
Tốc độ WiFi tối đa (điều kiện lý thuyết)	300 Mbps @ 2.4 GHz; 867 Mbps @ 5 GHz
Ăng-ten tích hợp / ngoài	2 × ăngten LTE (RSRP cải thiện) + 2 × ăngten WiFi
Cổng Ethernet LAN	4 cổng RJ45 10/100/1000 Mbps
Cổng WAN (tùy chọn)	1 cổng RJ45 10/100/1000 Mbps
Cổng USB (tùy model)	1 cổng USB 2.0 (chia sẻ lưu trữ / 3G fallback)
Chức năng định tuyến (Routing)	NAT, DHCP server, DHCP client, DMZ, Port Forwarding
Bảo mật	WEP, WPA-PSK, WPA2-PSK, SPI Firewall, DoS Protection
QoS / Traffic Management	Lưu lượng theo QoS, giới hạn băng thông, giám sát truy cập
ID mạng ảo (VPN)	IPsec, PPTP, L2TP (tùy model)
Nhiệt độ hoạt động	-20 °C ... 55 °C
Độ ẩm tương đối	5 % ... 95 % (không ngưng tụ)
Nguồn điện	12 V DC / 1,5 A (Adapter kèm theo)
Kích thước (RxCxD)	~ 180 mm × 110 mm × 95 mm

Trọng lượng	~ 350 g
--------------------	---------

● **Cảm biến nhiệt độ và độ ẩm (RK330-02)**

RK330-02 là cảm biến đo nhiệt độ và độ ẩm không khí treo tường (Wall-mounted Atmospheric Temperature & Humidity Sensor) do Hunan Rika Electronic Technology sản xuất. Thiết bị sử dụng cảm biến nhiệt độ–độ ẩm kỹ thuật số tích hợp sẵn, được bảo vệ bên trong mái che chống nước và tia tử ngoại, giúp đo lường chính xác trong các ứng dụng giám sát môi trường, nông nghiệp, lâm nghiệp, phòng thí nghiệm, kho bãi, nhà kính, và phòng viễn thông. RK330-02 có thể xuất tín hiệu qua RS-485 với giao thức Modbus RTU, tiêu thụ điện năng thấp (< 20 mA), hoạt động trong dải nhiệt độ rộng (-40 °C đến +70 °C) và đạt cấp bảo vệ IP65 chống bụi-nước, phù hợp lắp đặt ngoài trời hoặc trong môi trường ẩm ướt.



Hình 4.9. RK330-02

Bảng 4.6. Thông số kỹ thuật của RK330-02

Thông số kỹ thuật chính	Giá trị
Phạm vi đo nhiệt độ	-30 °C ... +70 °C
Sai số nhiệt độ	± 0,3 °C @ 25 °C
Lặp lại nhiệt độ (Repeatability)	± 0,1 °C
Phạm vi đo độ ẩm	0 %RH ... 100 %RH
Sai số độ ẩm	± 3 %RH (10 %–90 %RH)
Lặp lại độ ẩm (Repeatability)	± 0,5 %RH

Ổn định lâu dài (Long-term Stability)	< 0,04 °C/năm (nhiệt độ); < 0,5 %RH/năm (độ ẩm)
Thời gian phản hồi (Response Time)	≤ 1 giây
Nguồn cung cấp (Supply)	5 VDC; 12–24 VDC
Ngõ ra (Output Signal)	RS-485 (Modbus RTU)
Tùy chọn hiển thị	LCD (tùy chọn)
Tiêu thụ điện (Current Consumption)	< 20 mA
Nhiệt độ hoạt động	–40 °C ... +70 °C
Cấp bảo vệ (Ingress Protection)	IP65
Kích thước (Dimension)	117 × 86 × 41 mm
Trọng lượng (Weight)	170 g (không đóng gói)
Vật liệu vỏ (Shelter material)	ABS

● **Cảm biến bức xạ mặt trời (RK200-04)**

Cảm biến bức xạ mặt trời RK200-04 của Hunan Rika Electronic Technology dựa trên nguyên lý tế bào silicon, thiết kế không có bộ phận chuyển động nên không cần bảo trì và có thể hoạt động ổn định ở độ cao lớn. Thiết bị đo được bức xạ trong dải bước sóng 300–1100 nm, với độ nhạy cao, tiêu thụ điện năng thấp và tuổi thọ dài. RK200-04 có thể xuất tín hiệu analog (0–5 V hoặc 4–20 mA) hoặc giao tiếp số RS-485, cho phép sử dụng để giám sát bức xạ trực tiếp, bức xạ phản xạ trên mặt nghiêng (góc gắn ngang), hoặc bức xạ tán xạ nếu có thêm bộ che.



Hình 4.10. RK200-04

Bảng 4.7. Thông số kỹ thuật của RK200-04

Thông số cơ bản	Giá trị
Dải bước sóng đo	300–1 100 nm
Nguồn cung cấp	5 VDC hoặc 12–24 VDC
Phạm vi đo	0–1 500 W/m ²
Độ phân giải	1 W/m ²
Ngõ ra	0–5 VDC; 4–20 mA; RS-485
Thời gian phản hồi	≤ 5 s
Sai số hiệu chỉnh cosine	≤ ± 10 % (góc tới 10 °)
Độ phi tuyến tính	≤ ± 3 %
Ảnh hưởng nhiệt độ	± 0,08 %/ °C
Ổn định hằng năm	≤ ± 2 %/năm
Nhiệt độ hoạt động	– 40 °C ... + 80 °C
Mức bảo vệ	IP65
Trọng lượng (không đóng gói)	≈ 420 g
Vật liệu vỏ	Hợp kim nhôm

● Bộ nguồn Meanwell EDR-75-24

Meanwell EDR-75-24 là bộ nguồn công nghiệp gắn thanh DIN (DIN-rail) thuộc dòng EDR của hãng Meanwell, cho công suất 75 W và điện áp ngõ ra 24 V DC (tương đương 3,2 A). Thiết kế mỏng gọn, độ rộng chỉ 40 mm, giúp tiết kiệm không gian lắp đặt trong tủ điện. EDR-75-24 có dải điện áp đầu vào rộng (88 – 264 VAC hoặc 125 – 370 VDC), tích hợp các chức năng bảo vệ (ngắn mạch, quá dòng, quá áp, quá nhiệt) và đạt hiệu suất chuyển đổi cao (đến 89 %), phù hợp cấp nguồn ổn định cho PLC, cảm biến, relay và các thiết bị tự động hóa trong môi trường công nghiệp.



Hình 4.11. Bộ nguồn Meanwell EDR-75-24

Bảng 4.8. Thông số kỹ thuật của Meanwell EDR-75-24

Thông số chính	Giá trị
Công suất định mức	75 W
Điện áp ngõ ra danh định	24 V DC
Dòng ngõ ra định mức	3,2 A
Dải điện áp đầu vào	88 – 264 VAC / 125 – 370 VDC
Hiệu suất chuyển đổi (Typ.)	88 % – 89 %
Đầu ra phụ tải không tải (Max.)	< 1 W
Bảo vệ	Ngắn mạch, quá dòng (hiccup), quá áp (shutdown), quá nhiệt (shutdown)
Nhiệt độ hoạt động	-30 °C ... +70 °C (derating từ +50 °C)

Độ ẩm tương đối	20 % ... 90 % RH (không ngưng tụ)
Kích thước (W × H × D)	40 × 122 × 127 mm
Trọng lượng	~ 0,6 kg
Cấp bảo vệ (IP Rating)	IP20
Giấy chứng nhận	UL 508, UL 60950-1, EN 62368-1, CE, RoHS

4.2.2. Phương thức truyền thông và lưu trữ

a) Modbus

Modbus là một chuẩn giao thức truyền thông công nghiệp mở, phát triển lần đầu bởi *Modicon* (nay thuộc *Schneider Electric*) năm 1979 và chính thức thành chuẩn mở vào 1996. Chuẩn này cho phép trao đổi dữ liệu ổn định, truyền tải đi xa, miễn phí bản quyền và được các nhà sản xuất tự do tích hợp vào thiết bị của họ.

- Nguyên tắc hoạt động:

Modbus RTU hoạt động dựa trên nguyên tắc Master – Slave tức là một bên nhận (Master) và một bên truyền tín hiệu (Slave) thông qua địa chỉ thanh ghi. Phương thức truyền của Modbus RTU bằng đường truyền vật lý RS232 hoặc RS485, Modbus TCP/IP thì truyền trên địa chỉ IP thông qua Internet.

Trong công nghiệp - tự động hoá có 03 loại Modbus thông dụng Modbus RTU, Modbus ASCII, Modbus TCP/IP. Sự khác nhau giữa các loại Modbus:

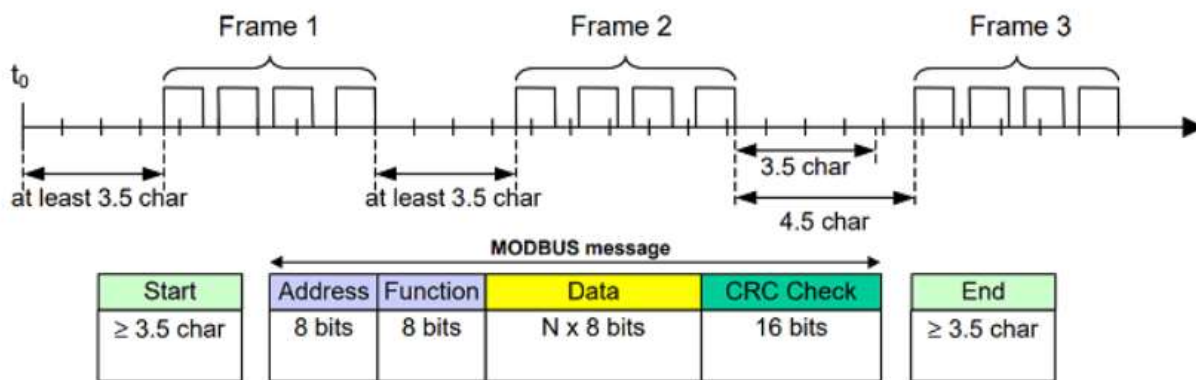
Ở đây, Hệ thống chỉ sử dụng Modbus RTU và Modbus TCP:

- Modbus RTU: mã hoá dạng nhị phân với 1byte dữ liệu và một byte truyền thông có tốc độ truyền 9600 – 57600 baud.
- Modbus ASCII: mã hoá dạng hexadecimal – 4bit, cần 2byte truyền thông cho một byte thông tin.
- Modbus TCP/IP Modbus TCP: là modbus RTU qua internet hay còn gọi là Modbus IP tương ứng với một địa chỉ là một IP.

❖ **Modbus RTU**

Modbus RTU đóng gói PDU (Function + Data) dưới dạng nhị phân, bổ sung CRC-16 để kiểm soát lỗi.

- Cấu trúc khung



Hình 4.12. Mô tả truyền dữ liệu của Modbus RTU

- **Address:** ID Slave (1–247).
- **Function:** mã lệnh (ví dụ 0x03 = đọc Holding Registers).
- **Data:** tham số địa chỉ, byte count, giá trị.
- **Crc-16:** đa thức 0xA001, tính trên Address→Data

Giữa hai khung phải có **inter-frame gap** ≥ 3.5 char-time để phân biệt khung.

❖ Ví dụ:

Slave id	Function code	Address_H	Address_L	Quantity_H	Quantity_L	CRC_L	CRC_H
0x01	0x03	0x00	0x00	0x00	0x01	0X84	0x0A

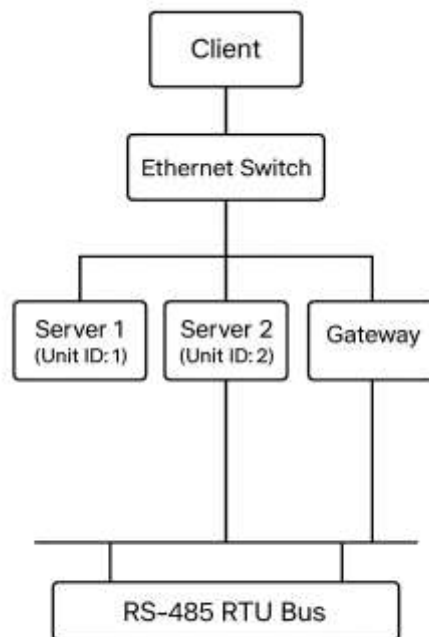
Hình 4.13. Host scan order trên 1 thiết bị cảm biến sử dụng RS485

❖ **Modbus TCP**

Modbus TCP/IP (Modbus TCP) là biến thể Modbus đóng gói PDU (Function + Data) bên trong gói TCP/IP, vận hành theo mô hình client–server. Giao thức này sử dụng port TCP 502 để trao đổi yêu cầu–phản hồi, tận dụng khả năng truyền tải full-duplex của Ethernet. Với Modbus TCP, dữ liệu Modbus được đơn giản hóa và đóng gói trong một gói tin TCP/IP.

Trong Modbus TCP, không có khái niệm về SlaveID như trong Modbus RTU, bởi vì nó sử dụng địa chỉ IP để xác định các thiết bị mục tiêu. Máy chủ (hoặc thiết bị chứa dữ liệu) không cần phải biết về SlaveID mà chỉ cần biết địa chỉ IP của các thiết bị cụ thể.

➤ Cấu trúc kết nối điển hình:



Hình 4.14. Cấu trúc kết nối điển hình

Mỗi server lắng nghe trên port 502; client khởi tạo TCP connection đến IP:502.

MBAP Header (7 byte) đi trước PDU, bao gồm Transaction ID, Protocol ID (=0), Length, Unit ID.

➤ Cơ chế truyền thông:

1. Client mở TCP socket tới server.
2. Gói “Modbus Application Protocol” (MBAP + PDU) được gửi trong khung TCP.
3. Server phản hồi trên cùng kết nối với MBAP echo và PDU response.
4. Kết nối có thể để mở lâu (persistent) hoặc đóng sau mỗi request tùy cấu hình.

b) MQTT

MQTT (*Message Queuing Telemetry Transport*) là một giao thức truyền tin theo mô hình publish/subscribe, được thiết kế tối ưu cho các hệ thống IoT (Internet of Things) có băng thông hạn chế, tài nguyên hoặc năng lượng hạn hẹp.

MQTT hướng tới việc trao đổi dữ liệu nhẹ, độ trễ thấp và khả năng hoạt động ổn định trong môi trường mạng không đáng tin cậy.

❖ MQTT Broker (Máy chủ trung tâm)

- Chịu trách nhiệm nhận, lưu trữ và chuyển tiếp thông điệp giữa các client.
- Đảm bảo việc phân phối tin hiệu đến đúng người đăng ký (subscriber) dựa trên topic.
- Có thể cấu hình ở dạng standalone (cục bộ) hoặc cluster (phân tán) để tăng khả năng chịu lỗi và mở rộng.

❖ MQTT Client

- Đóng vai trò là bên publisher (xuất bản) thông điệp lên một topic cụ thể, hoặc bên subscriber (đăng ký) để nhận thông điệp từ các topic quan tâm.
- Một client có thể vừa publish vừa subscribe đồng thời, nếu cần trao đổi hai chiều
- Thường được tích hợp trong thiết bị nhúng, cảm biến, gateway hoặc ứng dụng backend

❖ Mô hình Publish/Subscribe

- Chủ thể Publish: Gửi thông điệp (message) lên một topic nhất định, không quan tâm trực tiếp đến đối tượng đích.
- Chủ thể Subscribe: Đăng ký nhận tin dựa trên các topic, có thể dùng ký tự đại diện (wildcard) như + hoặc # để đăng ký nhiều cấp hoặc nhiều nhánh topic.
- Broker: Là trung gian đứng giữa, giữ vai trò định tuyến nội dung; các publisher và subscriber không giao tiếp trực tiếp với nhau mà thông qua broker.

❖ Cấu trúc thông điệp

Message bao gồm:

- Topic (chuỗi ký tự): Xác định “đường dẫn” để phân phối; ví dụ sensor/temperature/room1.

- Payload (dữ liệu nhị phân hoặc text): Thông tin thực tế (giá trị cảm biến, lệnh điều khiển) có thể là JSON, XML, hoặc định dạng nhị phân.
- QoS (Quality of Service): Chỉ định mức độ đảm bảo giao nhận:
 - 1) QoS 0: “At most once” – gửi một lần, không đảm bảo nhận; không lưu trữ lại.
 - 2) QoS 1: “At least once” – đảm bảo ít nhất một lần; có thể nhận trùng lặp.
 - 3) QoS 2: “Exactly once” – đảm bảo đúng một lần, dùng hai bước handshake, tiêu tốn băng thông và tài nguyên tính toán hơn.
- Retained Flag: Nếu bật, broker lưu lại thông điệp cuối cùng trên topic đó để gửi ngay cho bất kỳ subscriber mới nào.
- Last Will and Testament: Cơ chế cho phép client định nghĩa một thông điệp báo lỗi (will) sẽ được broker phát đi nếu client bất ngờ mất kết nối, phục vụ mục đích giám sát tình trạng thiết bị.

❖ Cơ chế kết nối và duy trì

- Connect: Client gửi yêu cầu kết nối (CONNECT) tới broker, kèm thông tin như Client ID, tùy chọn Username/Password và các tham số keep-alive.
- Keep-Alive: Broker và client trao đổi gói ping định kỳ (PINGREQ / PINGRESP) để đảm bảo kết nối vẫn sống; nếu broker không nhận tín hiệu nào quá thời gian keep-alive, broker xem client đã mất và sẽ gửi Last Will nếu được cấu hình.
- Disconnect: Khi client đóng kết nối một cách chính thức, gửi gói DISCONNECT để broker biết có thể xóa trạng thái giữ lại (session) nếu không cần phục hồi.

4.3. Tích hợp mô hình AI vào hệ thống IoT

Để tích hợp mô hình học máy (AI) vào hệ thống giám sát thời gian thực dựa trên nền tảng IoT, một yêu cầu quan trọng là khả năng triển khai mô hình đã huấn luyện sao cho có thể dự báo tức thời tại máy tính nhúng Raspberry Pi. Trong nghiên cứu này, mô hình được xây dựng bằng ngôn ngữ Python, và sau khi huấn luyện, toàn bộ pipeline học máy – bao gồm các bước tiền xử lý và mô hình dự báo được lưu trữ dưới định dạng *.pkl* (pickle).

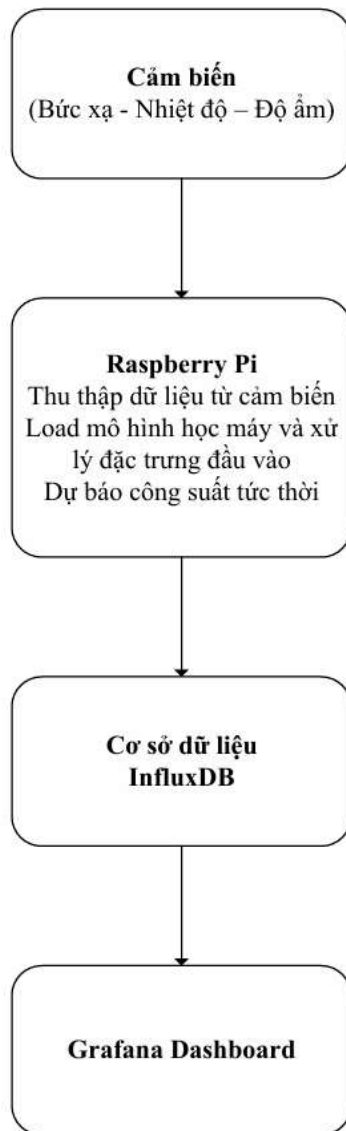
Việc lựa chọn định dạng *.pkl* thay vì các định dạng khác như *.json*, *.h5*, hoặc *.onnx* được cân nhắc kỹ lưỡng dựa trên đặc thù của hệ thống và yêu cầu triển khai. Cụ thể, *.pkl* cho phép tuần tự hóa toàn bộ đối tượng Python, bao gồm cả mô hình học máy và các bước xử lý đặc trưng đầu vào StandardScaler. Điều này giúp giảm đáng kể thời gian phát triển và đảm bảo tính nhất quán tuyệt đối giữa giai đoạn huấn luyện và suy luận (inference). Ngoài ra, *.pkl* có ưu điểm là nhẹ, tốc độ nạp nhanh và dễ dàng tích hợp vào các hệ thống giám sát IoT sử dụng Python – vốn là lựa chọn phổ biến cho các ứng dụng trên Raspberry Pi.

Sau khi được nạp trên thiết bị biên, mô hình *.pkl* sẽ nhận đầu vào từ các cảm biến môi trường (bức xạ mặt trời, nhiệt độ, độ ẩm) truyền về thông qua giao thức MQTT, thực hiện dự báo công suất tức thời, và xuất kết quả lên hệ cơ sở dữ liệu InfluxDB. Các giá trị đầu ra được trực quan hóa trên nền tảng Grafana, giúp người dùng có thể giám sát tình trạng hoạt động của hệ thống PV theo thời gian thực, đồng thời phát hiện bất thường hoặc suy giảm hiệu suất nếu có.

Để minh họa rõ hơn cách thức tích hợp mô hình học máy vào hệ thống giám sát điện mặt trời thời gian thực, Hình 4.15 dưới đây mô cơ chế hoạt động của mô hình AI tích hợp trong hệ thống IoT.

❖ Quy trình vận hành:

1. *Tầng cảm biến*: Các cảm biến môi trường như cảm biến bức xạ, nhiệt độ và độ ẩm được kết nối với Raspberry Pi thông qua giao thức MQTT.
2. *Tầng thu thập và xử lý dữ liệu*: Raspberry Pi đóng vai trò là máy chủ cục bộ (local gateway), sử dụng Node-RED và các đoạn script Python để thu thập dữ liệu từ cảm biến theo chu kỳ 5 phút. Sau khi dữ liệu được ghi nhận, nó được chuẩn hóa và chuyển thành định dạng phù hợp với mô hình dự báo.
3. *Tầng mô hình học máy*: Mô hình *.pkl* được nạp sẵn trên Raspberry Pi. Khi có dữ liệu đầu vào, mô hình thực hiện dự báo công suất đầu ra tương ứng. Việc sử dụng mô hình dưới dạng *.pkl* giúp đảm bảo tốc độ dự báo nhanh và độ chính xác cao.
4. *Tầng lưu trữ và trực quan hóa*: Kết quả dự báo và dữ liệu thực tế sẽ được ghi vào cơ sở dữ liệu thời gian thực InfluxDB. Dữ liệu này sau đó được hiển thị qua giao diện trực quan trên Grafana, cho phép người vận hành theo dõi biến động công suất, so sánh giữa giá trị đo và giá trị dự báo, cũng như phát hiện sớm các sai lệch hiệu suất do môi trường như bụi hoặc thời tiết bất thường.



Hình 4.15. Cơ chế dự báo của mô hình AI trong hệ thống IoT

4.4. Xây dựng giao diện giám sát theo thời gian thực

Giao diện giám sát giúp trực quan hóa các thông số vận hành theo thời gian thực, từ đó hỗ trợ phát hiện sớm các bất thường, góp phần nâng cao độ tin cậy và hiệu quả vận hành hệ thống.

Mặc dù các hãng sản xuất Inverter hiện nay đều cung cấp phần mềm giám sát từ xa, tuy nhiên các giải pháp này thường bị giới hạn về chức năng — vốn chủ yếu tập trung vào các thông số cơ bản như điện áp, dòng điện và công suất. Việc đánh giá hiệu suất

tổng thể hoặc tồn thất do suy giảm chất lượng tấm pin dựa trên các phần mềm này còn tồn tại nhiều hạn chế, dẫn đến quy trình bảo trì vẫn chủ yếu theo định kỳ, thiếu linh hoạt và chưa tối ưu về mặt kinh tế.

Bên cạnh đó, do các phần mềm hãng thường mang tính “đóng gói”, việc tích hợp thêm cảm biến môi trường hoặc mở rộng chức năng là không khả thi. Xuất phát từ những hạn chế này, chúng tôi đề xuất một hệ thống giám sát độc lập kết hợp ba nền tảng mã nguồn mở: Node-RED, InfluxDB, và Grafana:

- Node-RED là nền tảng lập trình trực quan theo mô hình luồng dữ liệu (flow-based), cho phép kết nối và xử lý dữ liệu IoT thông qua giao diện kéo-thả. Với khả năng tích hợp nhiều giao thức (MQTT, HTTP, Modbus...), Node-RED giúp xử lý dữ liệu một cách linh hoạt và nhanh chóng.
- InfluxDB là cơ sở dữ liệu hướng chuỗi thời gian (time-series), tối ưu cho việc ghi nhận và truy vấn dữ liệu gắn mốc thời gian. Cơ chế lưu trữ theo cột và ngôn ngữ truy vấn chuyên biệt cho phép xử lý hiệu quả các tác vụ phân tích xu hướng và phát hiện bất thường.
- Grafana là công cụ trực quan hóa dữ liệu thời gian thực, hỗ trợ đa dạng nguồn dữ liệu. Người dùng có thể xây dựng dashboard tùy biến, biểu diễn dữ liệu qua biểu đồ, bảng số liệu và thiết lập cảnh báo theo điều kiện cụ thể.

Với hệ thống đề xuất, dữ liệu từ cảm biến và các thanh ghi Inverter được lưu trữ cục bộ trên SSD và đồng thời truyền qua giao thức MQTT đến Node-RED. Tại đây, dữ liệu được xử lý và ghi vào InfluxDB. Cuối cùng, Grafana truy xuất dữ liệu từ InfluxDB để trình bày dưới dạng Dashboard thời gian thực. Giao diện này hỗ trợ người vận hành theo dõi toàn diện các thông số môi trường và kỹ thuật, từ đó tối ưu lịch bảo trì, nâng cao hiệu suất, và cải thiện khả năng phản ứng trước sự cố hệ thống.

4.5. Kết luận

Trong chương này, hệ thống giám sát và thu thập dữ liệu thời gian thực cho điện mặt trời mái nhà dựa trên nền tảng IoT và phần mềm mã nguồn mở đã được triển khai. Hệ thống này được xây dựng dựa trên sự kết hợp của nhiều thành phần, bao gồm phần cứng nhúng như Raspberry Pi, bộ thu thập dữ liệu inverter SmartLogger, các cảm biến môi trường (bức xạ, nhiệt độ, độ ẩm), cùng với các công cụ phần mềm như Node-RED, InfluxDB và Grafana.

Chương 5: KẾT QUẢ TRIỂN KHAI VÀ PHÂN TÍCH

5.1. Giới thiệu

Trong chương này, chúng tôi sẽ trình bày chi tiết về quá trình triển khai và kết quả hoạt động thực nghiệm của hệ thống giám sát và thu thập dữ liệu cho điện mặt trời mái nhà, được xây dựng dựa trên nền tảng Internet vạn vật (IoT) và tích hợp các mô hình trí tuệ nhân tạo (AI). Cụ thể, chương sẽ giới thiệu giao diện giám sát trực quan được phát triển trên nền tảng Grafana, nơi các thông số vận hành và môi trường quan trọng của hệ thống điện mặt trời được tổng hợp và hiển thị theo thời gian thực.

5.2. Hệ thống thực nghiệm

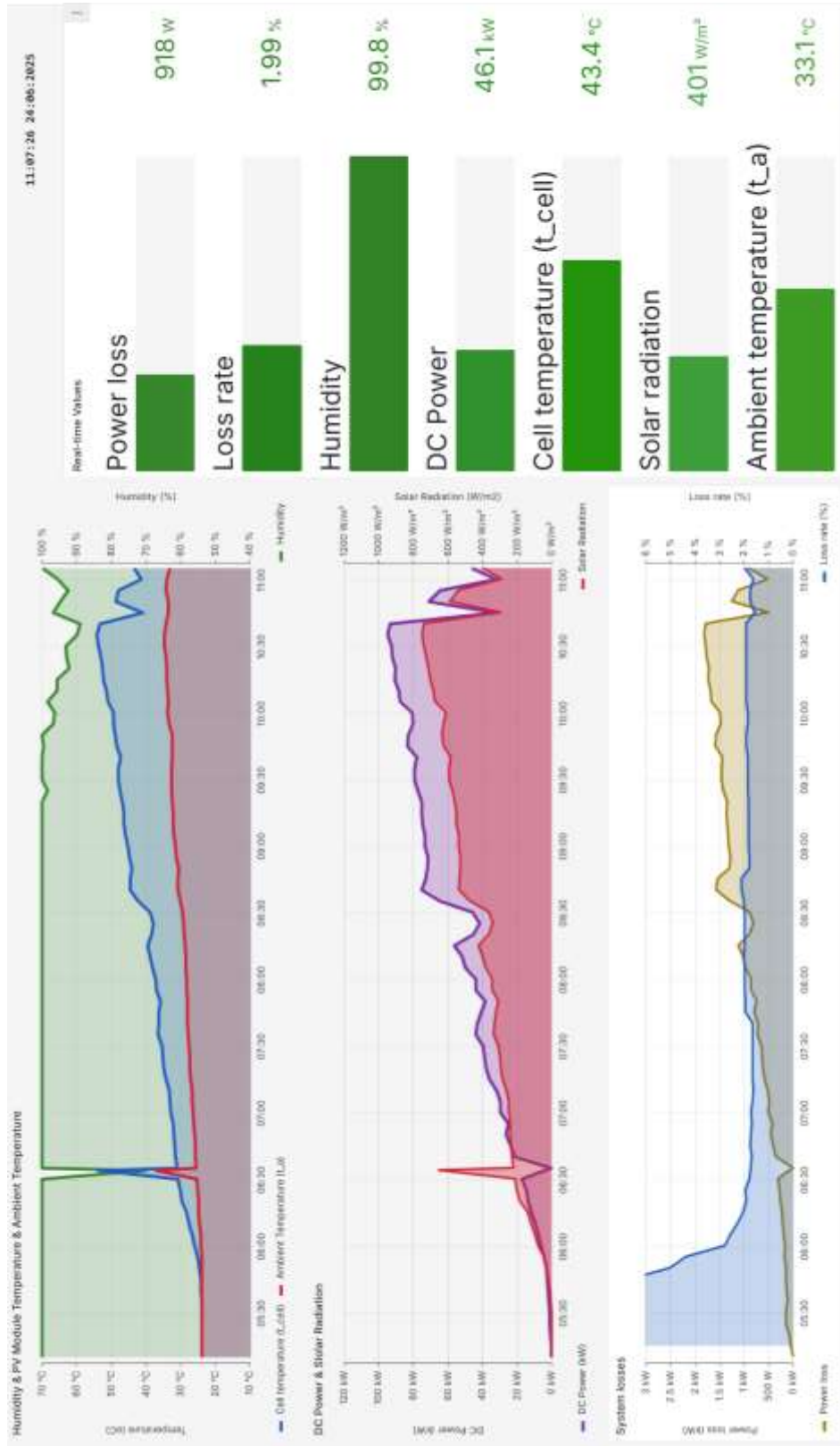
Hệ thống giám sát được triển khai sẽ ghi nhận và hiển thị dữ liệu thời gian thực từ hệ thống điện mặt trời trên nền tảng Grafana, đồng bộ dữ liệu từ Node-RED thông qua MQTT và lưu trữ trên InfluxDB.

Hình 5.1 minh họa kết quả trực quan hóa các thông số môi trường và điện năng, gồm ba biểu đồ chính:

- Biểu đồ (a): Độ ẩm, nhiệt độ môi trường và nhiệt độ tế bào PV. Biểu đồ này mô tả sự biến thiên trong ngày của ba đại lượng: nhiệt độ tế bào quang điện (T_{cell}), nhiệt độ môi trường ($T_{ambient}$), và độ ẩm tương đối (Humidity).
- Biểu đồ (b): Bức xạ mặt trời và công suất DC. Biểu đồ này thể hiện mối quan hệ giữa bức xạ mặt trời (W/m^2) và công suất DC đầu vào inverter (kW).
- Biểu đồ (c): Tổn thất công suất và tỷ lệ tổn thất. Biểu đồ này minh họa sự chênh lệch giữa công suất đầu vào DC và công suất đầu ra AC thông qua hai đại lượng công suất tổn thất (Power loss) và tỷ lệ tổn thất (Loss rate).

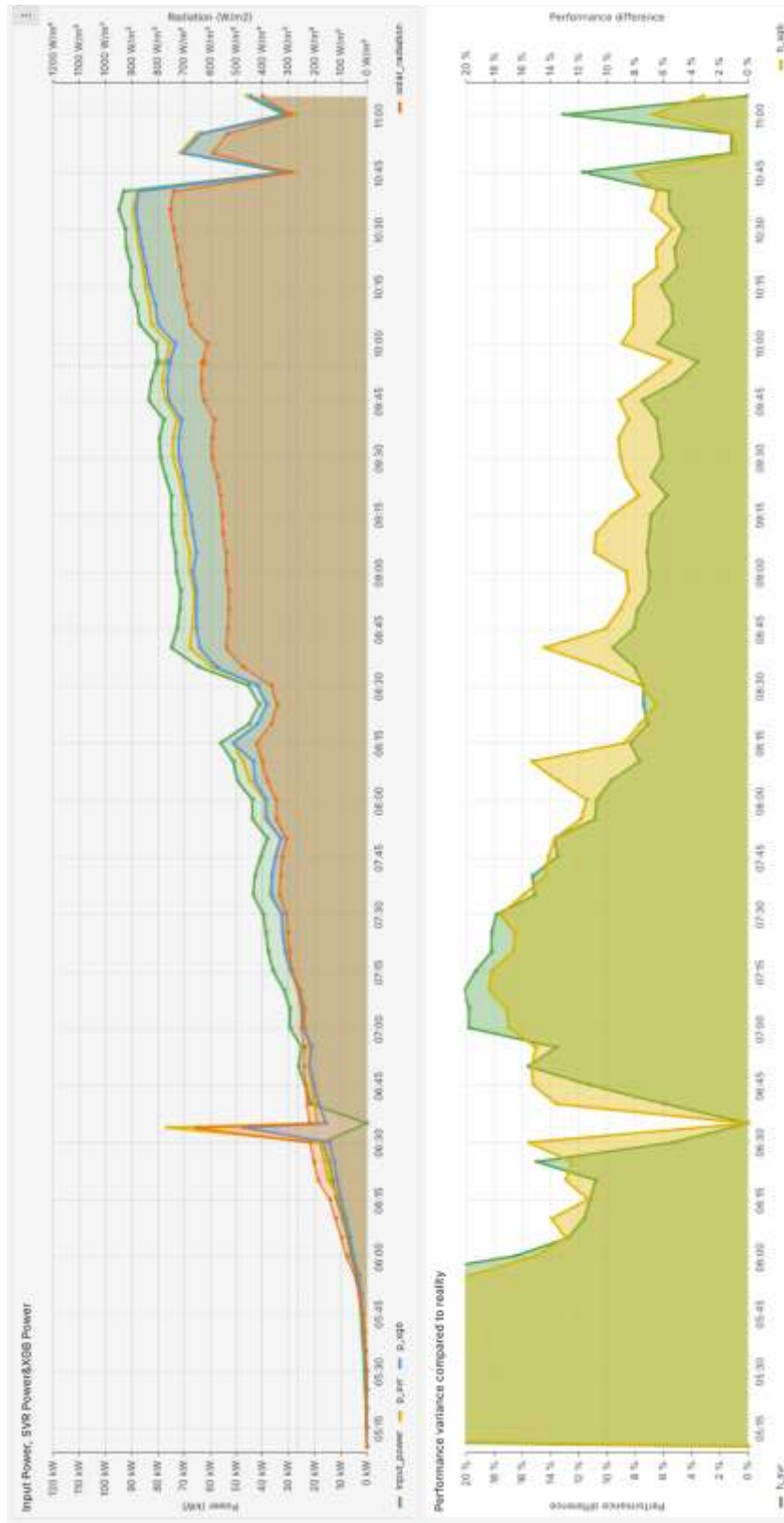
Hình 5.2 trình bày kết quả so sánh giữa công suất thực tế đo được từ inverter và công suất dự đoán từ hai mô hình học máy (SVR và XGBoost) trong cùng điều kiện vận hành. Biểu đồ gồm hai phần chính:

- Biểu đồ trên thể hiện công suất đầu vào DC đo từ hệ thống (đường xám nhạt), công suất dự đoán bởi mô hình SVR (màu xanh lá) và XGBoost (màu vàng), cùng với đường cong bức xạ mặt trời (màu cam, trục phải).
- Biểu đồ dưới minh họa sự chênh lệch dự báo của 2 mô hình so với thực tế.



Hình 5.1. Giao diện Dashboard của hệ thống

Nghiên cứu và triển khai hệ thống thu thập dữ liệu của điện mặt trời mái nhà dựa trên nền tảng IoT và AI phục vụ cho việc đánh giá hiệu suất



Hình 5.2. Công suất dự báo của mô hình AI hiển thị trên Dashboard

Sinh viên thực hiện: Ngô Minh Hiến
Đỗ Đoàn Tuấn Hùng
Trịnh Hoàng Khoa

Hướng dẫn: TS. Nguyễn Kim Ánh
KS. Lê Minh Hoàng Hạc

Hình 5.3 minh họa một ví dụ về giao diện trực quan hiển thị các thông số điện vận hành của hệ thống điện mặt trời mái nhà (PV) trong thời gian thực. Các thông số được theo dõi và cập nhật liên tục bao gồm:

- Công suất đầu ra AC (Output Power)
- Dòng điện AC ba pha (AC Current: L1, L2, L3)
- Điện áp AC ba pha (AC Voltage: L1, L2, L3)
- Tần số lưới (Frequency) và Nhiệt độ biến tần (Inverter Temperature)
- Công suất phản kháng (Reactive Power)
- Nhiệt độ biến tần (Inverter Temperature)

Voltage of Each String (V)																			
Timestamp	PV 2	PV 4	PV 5	PV 6	PV 7	PV 8	PV 10	PV 12	PV 13	PV 14	PV 16	PV 19	PV 19	PV 20					
2025-05-12 17:05:02	736.70	730.30	725.80	725.80	729.70	754.30	719.90	719.90	743.90	742.50	735.60	735.60	735.60	738.60					
2025-05-12 17:10:02	742.50	736.30	724.00	724.00	735.50	730.10	725.90	725.90	743.90	736.50	735.60	735.60	732.60	732.60					
2025-05-12 17:15:02	742.60	718.30	733.90	733.90	725.10	736.10	725.90	725.90	731.90	736.50	741.60	741.60	738.60	738.60					
2025-05-12 17:20:02	747.20	736.30	733.90	733.90	736.30	730.10	729.60	729.60	732.30	731.20	746.20	746.20	739.00	739.00					
2025-05-12 17:25:02	747.20	731.20	727.90	727.90	736.30	730.10	734.20	734.20	743.90	748.50	742.90	742.90	744.30	744.30					

Current Per String (A)																			
Timestamp	PV 2	PV 4	PV 5	PV 6	PV 7	PV 8	PV 10	PV 12	PV 13	PV 14	PV 16	PV 19	PV 19	PV 20					
2025-05-12 17:05:02	0.760	0.520	0.660	0.670	0.440	0.580	0.790	0.540	0.630	0.760	0.660	0.620	0.610	0.600					
2025-05-12 17:10:02	0.560	0.520	0.660	0.470	0.440	0.380	0.590	0.540	0.630	0.560	0.680	0.620	0.810	0.600					
2025-05-12 17:15:02	0.560	0.520	0.660	0.470	0.440	0.380	0.590	0.540	0.630	0.580	0.660	0.620	0.810	0.600					
2025-05-12 17:20:02	0.560	0.520	0.660	0.470	0.440	0.380	0.590	0.540	0.630	0.560	0.660	0.620	0.810	0.600					
2025-05-12 17:25:02	0.560	0.520	0.660	0.470	0.440	0.380	0.590	0.540	0.630	0.560	0.660	0.620	0.810	0.600					

Monitoring: AC Output & Grid Voltage												
Timestamp	AC Power (kW)	Q (Var)	Vab (V)	Vbc (V)	Vca (V)	Inv Temp (oC)	Ia (A)	Va (V)	Ib (A)	Vb (V)	Ic (A)	Vc (V)
2025-05-12 17:05:02	6.11	0.00400	403	405	405	54.7	8.94	232	8.77	234	8.82	234
2025-05-12 17:10:02	5.52	0.00400	403	404	404	53.9	8.13	232	7.83	234	7.98	234
2025-05-12 17:15:02	5.14	0.0100	402	403	404	53.5	7.51	232	7.52	233	7.57	234
2025-05-12 17:20:02	5.47	-0.0260	402	403	403	53.1	7.95	232	7.73	233	7.78	234
2025-05-12 17:25:02	5.38	-0.00300	402	403	403	52.8	7.95	232	7.73	233	7.78	234

Hình 5.3. Dữ liệu điện thời gian thực thu thập từ hệ thống PV, bao gồm cả nhiệt độ biến tần

5.3. Đánh giá kết quả

Kết quả triển khai cho thấy hệ thống giám sát thời gian thực dựa trên kiến trúc IoT (Raspberry Pi – SmartLogger – MQTT – Node-RED – InfluxDB – Grafana) vận hành ổn định, đáp ứng tốt các yêu cầu thu thập, xử lý, lưu trữ và trực quan hóa dữ liệu từ hệ thống điện mặt trời. Chuỗi dữ liệu từ cảm biến môi trường và các thông số điện (công suất, điện áp, dòng điện, tần số, nhiệt độ...) được ghi nhận đầy đủ, đồng bộ và hiển thị liên tục mà không mất gói dữ liệu.

Hệ thống cho phép giám sát cả ở cấp chuỗi PV lẫn toàn hệ thống, hỗ trợ đánh giá hiệu suất theo thời gian thực. Việc tích hợp mô hình học máy để so sánh giữa công suất dự đoán và thực tế giúp phát hiện sớm các biểu hiện suy giảm hiệu suất do bụi bẩn, quá nhiệt hoặc lỗi thiết bị tiềm ẩn. Ngoài ra, khả năng trích xuất dữ liệu phục vụ nghiên cứu, huấn luyện mô hình hoặc đánh giá suy hao theo mùa cũng được bảo đảm.

5.4. Thảo luận kết quả và tính ứng dụng thực tế

Hệ thống đã chứng minh hiệu quả trong việc giám sát và thu thập dữ liệu thời gian thực cho các hệ thống PV mái nhà. Điểm nổi bật là khả năng truy cập dữ liệu từ xa thông qua nền tảng web, phù hợp với các khu vực không thuận tiện cho việc bố trí nhân lực vận hành thường xuyên.

Tích hợp các mô hình học máy như SVR và XGBoost không chỉ hỗ trợ giám sát mà còn nâng cao năng lực dự báo, giúp phát hiện bất thường mà không cần thiết bị đo chuyên dụng. Điều này góp phần tối ưu hóa lịch bảo trì và nâng cao độ tin cậy trong vận hành.

Tuy nhiên, một số thách thức vẫn còn tồn tại. Hệ thống phụ thuộc vào kết nối mạng (4G), nên việc truy cập từ xa có thể bị gián đoạn khi mất tín hiệu. Ngoài ra, hiệu quả của mô hình học máy phụ thuộc lớn vào chất lượng và độ bao phủ của dữ liệu huấn luyện, đòi hỏi thu thập dữ liệu dài hạn và đa dạng hơn.

KẾT LUẬN

❖ Kết luận chung

Đồ án đã tập trung nghiên cứu, thiết kế và triển khai thực nghiệm hệ thống giám sát hiệu suất cho điện mặt trời mái nhà, kết hợp công nghệ IoT và mô hình học máy nhằm nâng cao khả năng giám sát và dự báo công suất.

Hệ thống được xây dựng trên nền tảng Raspberry Pi, SmartLogger, Node-RED, InfluxDB và Grafana, cho thấy khả năng vận hành ổn định, chi phí thấp và dễ mở rộng. Toàn bộ chuỗi dữ liệu từ cảm biến môi trường và thông số điện (công suất, dòng điện, điện áp, nhiệt độ, độ ẩm, bức xạ) được thu thập liên tục, lưu trữ có tổ chức và hiển thị trực quan theo thời gian thực.

Việc tích hợp các mô hình học máy như SVR và XGBoost giúp dự đoán công suất đầu ra dựa trên điều kiện môi trường. Trong đó, XGBoost cho kết quả ổn định và chính xác hơn, hỗ trợ đánh giá hiệu suất, phát hiện suy giảm bất thường và lập kế hoạch bảo trì hợp lý.

Hệ thống còn hỗ trợ giám sát từ xa thông qua nền tảng web, đặc biệt phù hợp với các khu vực vùng sâu, vùng xa, nơi việc bố trí nhân sự vận hành thường xuyên là khó khăn. Nhờ đó, có thể giảm chi phí quản lý và tăng khả năng phản ứng khi xảy ra sự cố.

❖ Hướng phát triển

Dựa trên kết quả triển khai thực nghiệm, một số định hướng phát triển tiềm năng nhằm nâng cao hiệu quả hệ thống bao gồm:

- 1) Phát triển cảnh báo thông minh và bảo trì theo tình trạng (CBM): Xây dựng cơ chế cảnh báo khi phát hiện sai lệch công suất bất thường hoặc thông số môi trường vượt ngưỡng, nhằm hỗ trợ chủ động trong vận hành và tối ưu hóa lịch bảo trì.
- 2) Mở rộng hệ thống giám sát đa điểm: Hướng tới xây dựng nền tảng giám sát tập trung cho nhiều hệ thống PV phân tán, cho phép đánh giá hiệu suất toàn cục và vận hành đồng bộ thông qua phân tích học máy theo cụm.

TÀI LIỆU THAM KHẢO

- [1] IRENA (2024), World Energy Transitions Outlook 2024: 1.5°C Pathway, International Renewable Energy Agency, Abu Dhabi
- [2] Mei Xu, Pu Xie, and Bai-Chen Xie, "Study of China's optimal solar photovoltaic power development path to 2050," Resources Policy, vol. 65, 101541, Mar. 2020. [Online]. Available: <https://doi.org/10.1016/j.resourpol.2019.101541>
- [3] Mohammed H. Qais, Hany M. Hasanien, and Saad Alghuwainem, "Identification of electrical parameters for three-diode photovoltaic model using analytical and sunflower optimization algorithm," Applied Energy, vol. 250, pp. 109–117, Sep. 2019. [Online]. Available: <https://doi.org/10.1016/j.apenergy.2019.05.013>
- [4] İpek Çetinkaya, "Parameter extraction of single, double, and triple-diode photovoltaic models using the Weighted Leader Search Algorithm," Global Challenges, vol. 8, no. 5, 2300355, Apr. 2024. [Online]. Available: <https://doi.org/10.1002/gch2.202300355>
- [5] Y. Shen, M. Fouladirad, and A. Grall, "Impact of dust and temperature on photovoltaic panel performance: A model-based approach to determine optimal cleaning frequency," Heliyon, vol. 10, no. 16, e35390, Aug. 2024. [Online]. Available: <https://doi.org/10.1016/j.heliyon.2024.e35390>
- [6] J. F. Gaviria, G. Narváez, C. Guillen, L. F. Giraldo, and M. Bressan, "Machine learning in photovoltaic systems: A review," Renewable Energy, vol. 196, pp. 298–318, Aug. 2022. [Online]. Available: <https://doi.org/10.1016/j.renene.2022.06.105>
- [7] G. Nofuentes, B. García-Domingo, J.V. Muñoz, and F. Chenlo, "Analysis of the dependence of the spectral factor of some PV technologies on the solar spectrum distribution," Applied Energy, vol. 113, pp. 302–309, 2014.
- [8] King, D. L., Kratochvil J. A. and Boyson, W. E., "Measuring the solar spectral and angle of incidence effects on photovoltaics modules and solar irradiance sensors," IEEE Photovoltaic Specialists Conference, Anaheim, California (1997).

- [9] H. A. Zondag, “Flat-plate PV-thermal collectors and systems: A review,” *Renewable and Sustainable Energy Reviews*, vol. 12, no. 4, pp. 891–959, May 2008. [Online]. Available: <https://doi.org/10.1016/j.rser.2005.12.012>
- [10] S. Dubey, J. N. Sarvaiya, and B. Seshadri, “Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world – A review,” *Energy Procedia*, vol. 33, pp. 311–321, 2013. [Online]. Available: <https://doi.org/10.1016/j.egypro.2013.05.072>
- [11] K. Kanuga, “Degradation of polyester film exposed to accelerated indoor damp heat aging,” in *Proc. 37th IEEE Photovoltaic Specialists Conference (PVSC)*, Seattle, WA, USA, Jun. 2011, pp. 000096–000100. [Online]. Available: <https://doi.org/10.1109/PVSC.2011.6185853>
- [12] G. Oreski, B. Ottersböck, and A. Omazic, “Degradation processes and mechanisms of encapsulants,” in *Durability and Reliability of Polymers and Other Materials in Photovoltaic Modules*, *Plastics Design Library*, 1st ed., T. D. Dang, Ed. William Andrew, 2019, pp. 135–152. [Online]. Available: <https://doi.org/10.1016/B978-0-12-811545-9.00006-9>
- [13] P. Rajput, O. S. Sastry, and G. N. Tiwari, “Effect of irradiance, temperature exposure and an Arrhenius approach to estimating weathering acceleration factor of glass, EVA and Tedlar in a composite climate of India,” *Solar Energy*, vol. 144, pp. 267–277, Mar. 2017. [Online]. Available: <https://doi.org/10.1016/j.solener.2017.01.027>
- [14] M. M. Fouad, L. A. Shihata, and E. I. Morgan, “An integrated review of factors influencing the performance of photovoltaic panels,” *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1499–1511, Dec. 2017. [Online]. Available: <https://doi.org/10.1016/j.rser.2017.05.141>
- [15] A. Hussain, A. Batra, and R. Pachauri, “An experimental study on effect of dust on power loss in solar photovoltaic module,” *Renewables: Wind, Water, and Solar*, vol. 4, Art. no. 9, Dec. 2017. [Online]. Available: <https://doi.org/10.1186/s40807-017-0042-z>

- [16] S. A. M. Said and H. M. Walwil, "Fundamental studies on dust fouling effects on PV module performance," *Solar Energy*, vol. 107, pp. 328–337, Sep. 2014. [Online]. Available: <https://doi.org/10.1016/j.solener.2014.05.048>
- [17] E. Ayora, M. Munji, K. Kaberere, and T. Bundi, "Performance analysis of 600 kWp grid-tied rooftop solar photovoltaic systems at Strathmore University in Kenya," *Results in Engineering*, vol. 19, 101302, Sep. 2023. [Online]. Available: <https://doi.org/10.1016/j.rineng.2023.101302>
- [18] Z. Liu, Y. Gong, Z. Wang, Y. Li, and D. Liu, "Revealing the intrinsic mechanisms of hot and cold spots within a locally shaded photovoltaic module based on micro-electrical characteristics," *Energies*, vol. 17, no. 17, Art. no. 4462, Sep. 2024. [Online]. Available: <https://doi.org/10.3390/en17174462>
- [19] M. S. El-Shobokshy and F. M. Hussein, "Effect of dust with different physical properties on the performance of photovoltaic cells," *Solar Energy*, vol. 51, no. 6, pp. 505–511, Dec. 1993. [Online]. Available: [https://doi.org/10.1016/0038-092X\(93\)90135-B](https://doi.org/10.1016/0038-092X(93)90135-B)
- [20] Jamil, W. J., Abdul Rahman, H., & Shaari, S. (2015). Effect of dust physical on the performance of poly-crystalline PV module based on indoor experiment. Proceedings of the 2015 Conference on Emerging Energy and Process Technology (CONCEPT2015), 15–16 December, A'Famosa Resort, Melaka, Malaysia
- [21] P. Wang, M. Kong, L. Wang and L. Ni, "The Effect of the Superhydrophobic Film on the Generation Efficiency of Photovoltaic Modules Affected by Salt-Containing Dust Deposition," in *IEEE Journal of Photovoltaics*, vol. 9, no. 6, pp. 1727-1732, Nov. 2019, doi: 10.1109/JPHOTOV.2019.2930909.
- [22] Chen, Y., Wang, D., Liu, Y. et al. Research on Photovoltaic Performance Reduction due to Dust Deposition: Modelling and Experimental Approach. *J. Therm. Sci.* 28, 1186–1194 (2019). <https://doi.org/10.1007/s11630-019-1177-6>
- [23] W. Al-Kouz, S. Al-Dahidi, B. Hammad, and M. Al-Abed, "Modeling and analysis framework for investigating the impact of dust and temperature on PV systems' performance and optimum cleaning frequency," *Applied Sciences*, vol. 9, no. 7, Art. no. 1397, Apr. 2019. [Online]. Available: <https://doi.org/10.3390/app9071397>

[24] F. Carigiet, C. J. Brabec, and F. P. Baumgartner, “Long-term power degradation analysis of crystalline silicon PV modules using indoor and outdoor measurement techniques,” *Renewable and Sustainable Energy Reviews*, vol. 144, Art. no. 111005, Jul. 2021. [Online]. Available: <https://doi.org/10.1016/j.rser.2021.111005>

[25] V. Sharma and S. S. Chandel, “Performance and degradation analysis for long term reliability of solar photovoltaic systems: A review,” *Renewable and Sustainable Energy Reviews*, vol. 27, pp. 753–767, Nov. 2013. [Online]. Available: <https://doi.org/10.1016/j.rser.2013.07.046>

[26] A. Omazic, G. Oreski, M. Halwachs, G. C. Eder, C. Hirschl, L. Neumaier, G. Pinter, and M. Erceg, “Relation between degradation of polymeric components in crystalline silicon PV module and climatic conditions: A literature review,” *Solar Energy Materials and Solar Cells*, vol. 192, pp. 123–133, Apr. 2019. [Online]. Available: <https://doi.org/10.1016/j.solmat.2018.12.027>

[27] A. Rezaee Jordehi, “Parameter estimation of solar photovoltaic (PV) cells: A review,” *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 354–371, Aug. 2016. [Online]. Available: <https://doi.org/10.1016/j.rser.2016.03.049>

[28] R. Venkateswari and N. Rajasekar, “Review on parameter estimation techniques of solar photovoltaic systems,” *International Transactions on Electrical Energy Systems*, vol. 31, no. 11, Art. no. e13113, Nov. 2021. [Online]. Available: <https://doi.org/10.1002/2050-7038.13113>

[29] R. Venkateswari and N. Rajasekar, “Review on parameter estimation techniques of solar photovoltaic systems,” *International Transactions on Electrical Energy Systems*, vol. 31, no. 11, Art. no. e13113, Nov. 2021. [Online]. Available: <https://doi.org/10.1002/2050-7038.13113> [30] M. Calasan, S. H. E. Abdel Aleem, and A. F. Zobaa, “A new approach for parameters estimation of double and triple diode models of photovoltaic cells based on iterative Lambert W function,” *Solar Energy*, vol. 218, pp. 392–412, Apr. 2021. [Online]. Available: <https://doi.org/10.1016/j.solener.2021.02.038>

[31] P. Gupta and R. Singh, “PV power forecasting based on data-driven models: A review,” *International Journal of Sustainable Engineering*, vol. 14, no. 6, pp. 1733–1755, Oct. 2021. [Online]. Available: <https://doi.org/10.1080/19397038.2021.1986590>

- [32] Y. Essam, A. N. Ahmed, R. Ramli, K.-W. Chau, M. S. I. Ibrahim, and M. Sherif, “Investigating photovoltaic solar power output forecasting using machine learning algorithms,” *Engineering Applications of Computational Fluid Mechanics*, vol. 16, no. 1, pp. 2002–2034, Oct. 2022. [Online]. Available: <https://doi.org/10.1080/19942060.2022.2126528>
- [33] Gupta, A.K., Singh, R.K. A review of the state of the art in solar photovoltaic output power forecasting using data-driven models. *Electr Eng* 107, 4727–4770 (2025). <https://doi.org/10.1007/s00202-024-02759-0>
- [34] R. Yang, T. Xie, and Z. Liu, “The application of machine learning methods to predict the power output of internal combustion engines,” *Energies*, vol. 15, no. 9, Art. no. 3242, Apr. 2022. [Online]. Available: <https://doi.org/10.3390/en15093242>
- [35] M. Uzair, S. Al-Kafrawi, K. Al-Janadi, and I. Al-Bulushi, “A low-cost, real-time rooftop IoT-based photovoltaic (PV) system for energy management and home automation,” *Energy Engineering*, vol. 119, no. 1, pp. 83–101, Nov. 2021. [Online]. Available: <https://doi.org/10.32604/EE.2022.016411>
- [36] M. A. Abdoukader, A. I. Kassim, and Y. A. Guedi, Design of a data acquisition and real-time monitoring system for PV module, Technical Report, University of Djibouti, Jun. 2020. [Online]. Available: <https://doi.org/10.13140/RG.2.2.20672.05121>
- [37] J. M. Paredes-Parra, A. Mateo-Aroca, G. Silvente-Niñirola, M. C. Bueso, and Á. Molina-García, “PV module monitoring system based on low-cost solutions: Wireless Raspberry application and assessment,” *Energies*, vol. 11, no. 11, Art. no. 3051, Nov. 2018. [Online]. Available: <https://doi.org/10.3390/en11113051>
- [38] S. Ansari, A. Ayob, M. S. H. Lipu, M. H. M. Saad, and A. Hussain, “A review of monitoring technologies for solar PV systems using data processing modules and transmission protocols: Progress, challenges and prospects,” *Sustainability*, vol. 13, no. 15, Art. no. 8120, Jul. 2021. [Online]. Available: <https://doi.org/10.3390/su13158120>
- [39] Smola, A.J., Schölkopf, B. A tutorial on support vector regression. *Statistics and Computing* 14, 199–222 (2004). <https://doi.org/10.1023/B:STCO.0000035301.49549.88>

- [40] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. New York, NY, USA: Springer-Verlag, 2000. doi: 10.1007/978-1-4757-3264-1.
- [41] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [42] M. Jobayer, M. A. H. Shaikat, M. N. Rashid, and M. R. Hasan, "A systematic review on predicting PV system parameters using machine learning," *Heliyon*, vol. 9, p. e16815, 2023. doi: 10.1016/j.heliyon.2023.e16815.
- [43] S. G. Gouda, Z. Hussein, S. Luo, and Q. Yuan, "Model selection for accurate daily global solar radiation prediction in China," *J. Clean. Prod.*, vol. 221, pp. 132-144, Jun. 2019. doi: 10.1016/j.jclepro.2019.02.211.
- [44] Tran, T.A.A., Ngo, M.H., Nguyen, K.A., " An IoT- and Mobile Network-Based Architecture for Monitoring of a Solar PV System", In proceeding of The 10th International Conference on Applying New Technology in Green Buildings, Danang, Vietnam, 25-26 July, 2025. IEEE.
- [45] Nguyen, K. A., Do, D. T. H., Trinh, H. K., Nguyen, N. K., & Doan, N. B., "Machine learning-based estimation of power output in solar photovoltaic systems under real-world conditions," in *Proceedings of The Technology Innovation for Sustainable Development 2025*, Da Nang, Vietnam, July 04–05, 2025