

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KỸ THUẬT MÁY TÍNH

NGUYỄN TIẾN TOÀN

KHÓA LUẬN TỐT NGHIỆP
HỆ THỐNG GIÁM SÁT VÀ DỰ BÁO LƯỢNG MƯA CÁC
TỈNH BẮC TRUNG BỘ

Monitoring And Rainfall Prediction System For The North Central
Provinces

CỬ NHÂN NGÀNH KỸ THUẬT MÁY TÍNH

TP. HỒ CHÍ MINH, NĂM 2024

ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA KỸ THUẬT MÁY TÍNH

NGUYỄN TIẾN TOÀN - 21521548

KHÓA LUẬN TỐT NGHIỆP
HỆ THỐNG GIÁM SÁT VÀ DỰ BÁO LƯỢNG MƯA CÁC
TỈNH BẮC TRUNG BỘ

Monitoring And Rainfall Prediction System For The North Central
Provinces

CỬ NHÂN NGÀNH KỸ THUẬT MÁY TÍNH

GIẢNG VIÊN HƯỚNG DẪN
TS. ĐOÀN DUY

TP. HỒ CHÍ MINH, NĂM 2024

THÔNG TIN HỘI ĐỒNG CHẤM KHÓA LUẬN TỐT NGHIỆP

Hội đồng chấm khóa luận tốt nghiệp, thành lập theo Quyết định số 07/QĐ-ĐHCNTT ngày 06 tháng 01 năm 2025 của Hiệu trưởng Trường Đại học Công nghệ Thông tin.

LỜI CẢM ƠN

Lời đầu tiên, em muôn bày tỏ lòng biết ơn sâu sắc đến quý thầy cô và các bạn sinh viên khoa Kỹ thuật máy tính nói riêng và trường đại học Công nghệ thông tin - Đại học Quốc gia Thành phố Hồ Chí Minh nói chung đã tạo điều kiện, hỗ trợ nhóm em thực hiện nghiên cứu đồ án này.

Trong suốt quá trình học tập và nghiên cứu, em đã nhận được sự hỗ trợ tận tình từ phí nhà trường và đoàn khoa thông qua nhiều hình thức khác nhau, từ đó có thêm cho mình những kiến thức và kinh nghiệm, góc nhìn là những hành trang quý báu đối với em sau này.

Đặc biệt nhất, em xin gửi lời cảm ơn chân thành của nhóm em đối với người anh, người thầy là TS. Đoàn Duy là người hướng dẫn trong suốt quá trình thực hiện nghiên cứu. Những kiến thức chuyên môn, những lời chia sẻ từ anh mà nhóm em may mắn có cơ hội nhận được đã, đang và sẽ luôn là những bài học quý giá trong quá trình phát triển bản thân em sau này.

Một lần nữa, em xin gửi lời chân thành cảm ơn đến quý thầy cô, giảng viên của khoa Kỹ thuật máy tính nói riêng và ban giám hiệu, công nhân viên chức công đoàn nhà trường, các bạn sinh viên thuộc trường Đại học Công nghệ thông tin đã cho em có cơ hội được trải nghiệm môi trường học tập, nghiên cứu và phát triển bản thân trong suốt quãng đời sinh viên của mình.

Mặc dù rất nỗ lực để luận văn được hoàn thiện nhất, tuy nhiên khó có thể tránh khỏi hạn chế và thiếu sót. Kính mong nhận được sự thông cảm và góp ý từ thầy cô và các bạn.

*Sinh viên thực hiện
Nguyễn Tiến Toàn
Khoa Kỹ thuật máy tính – Lớp KTMT 2021*

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI	2
1.1 Đặt vấn đề	2
1.2 Lý do chọn đề tài	3
1.3 Đối tượng và phạm vi thực hiện	4
1.3.1 Đối tượng nghiên cứu	4
1.3.2 Phạm vi thực hiện	4
1.3.3 Mục tiêu nghiên cứu	4
1.4 Phương pháp thực hiện	6
CHƯƠNG 2. TỔNG QUAN ĐỀ TÀI	7
2.1 Tình hình nghiên cứu quốc tế	7
2.2 Tình hình nghiên cứu trong nước	8
CHƯƠNG 3. CƠ SỞ LÝ THUYẾT	9
3.1 Thiết bị đầu cuối	9
3.2 Máy học và Học sâu	13
3.2.1 Lý thuyết thống kê	13
3.2.2 Mô hình Random Forest	15
3.2.3 Mô hình Long Short Term Memory	16
3.2.4 Các thang đo đánh giá hiệu suất	18
3.3 Nguồn dữ liệu	21
3.3.1 Mã nguồn mở OPEN-METEO	21
3.3.2 Hệ thống các trạm đo mưa VRAIN	23
3.3.3 Bộ dữ liệu ứng dụng cho mô hình	27

CHƯƠNG 4. PHƯƠNG PHÁP NGHIÊN CỨU VÀ THỰC HIỆN	31
4.1 Tổng quan hệ thống	31
4.2 Quy trình hoạt động thiết bị đầu cuối	31
4.2.1 Kiến trúc phân lớp thiết bị đầu cuối	31
4.2.2 Kịch bản hoạt động	33
4.3 Hệ thống máy chủ	42
4.3.1 Mô đun thu thập dữ liệu và trích chọn đặc trưng	42
4.3.2 Tiền xử lý dữ liệu	45
4.3.3 Mô-đun dự báo lượng mưa	47
4.3.4 Hệ thống quản trị cơ sở dữ liệu	51
CHƯƠNG 5. ĐÁNH GIÁ KẾT QUẢ THỰC NGHIỆM	52
5.1 Vận hành thiết bị đầu cuối	52
5.1.1 Đóng gói thiết bị	52
5.1.2 Hiển thị dữ liệu	53
5.2 Giao tiếp với hệ thống máy chủ	55
5.3 Đánh giá kết quả huấn luyện	57
5.3.1 Mô hình dự báo theo dữ liệu IFS	57
5.3.2 Mô hình dự báo theo dữ liệu IFS kết hợp Vrain	58
5.3.3 Đánh giá mô hình trên tập Test bộ dữ liệu IFS	59
5.3.4 Đánh giá mô hình trên tập Test bộ dữ liệu kết hợp Vrain	61
5.3.5 Giao diện người dùng	63
CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	65
6.1 Kết luận	65
6.2 Khó khăn gặp phải	65
6.3 Hướng phát triển	67
TÀI LIỆU THAM KHẢO	69
PHỤ LỤC	72

DANH MỤC HÌNH ẢNH

Hình 3.1.	Sơ đồ kết nối các thành phần thiết bị đầu cuối	9
Hình 3.2.	Khung truyền dữ liệu giao thức I2C	11
Hình 3.3.	Mô hình kết nối MQTT	12
Hình 3.4.	Trục quan dữ liệu khí tượng thuỷ văn tại trạm Triệu Ái, Quảng Trị	14
Hình 3.5.	Kiến trúc mô hình LSTM	17
Hình 3.6.	Điểm dữ liệu theo mô hình IFS ở tỉnh Hà Tĩnh	22
Hình 3.7.	Điểm dữ liệu theo mô hình IFS ở tỉnh Quảng Bình	23
Hình 3.8.	Điểm dữ liệu theo mô hình IFS ở tỉnh Quảng Trị	23
Hình 3.9.	Thiết bị cảm biến đo mưa của Vrain	24
Hình 3.10.	Bộ xử lý bên trong cảm biến đo mưa của Vrain	25
Hình 3.11.	Vị trí các trạm đo mưa Vrain tỉnh Hà Tĩnh	26
Hình 3.12.	Vị trí các trạm đo mưa Vrain tại tỉnh Quảng Bình	26
Hình 3.13.	Vị trí các trạm đo mưa Vrain tại tỉnh Hà Tĩnh	27
Hình 3.14.	Cặp điểm dữ liệu tỉnh Hà Tĩnh	28
Hình 3.15.	Cặp điểm dữ liệu tỉnh Quảng Bình	28
Hình 3.16.	Cặp điểm dữ liệu tỉnh Quảng Trị	29
Hình 4.1.	Kiến trúc hệ thống	32
Hình 4.2.	Sơ đồ chuyển trạng thái hoạt động	35
Hình 4.3.	Nửa trên biểu đồ tuần tự trạng thái khởi tạo	36
Hình 4.4.	Nửa dưới biểu đồ tuần tự trạng thái khởi tạo	36
Hình 4.5.	Biểu đồ tuần tự trạng thái mở đầu	37
Hình 4.6.	Phần trên biểu đồ tuần tự trạng thái hiển thị dữ liệu	38

Hình 4.7.	Nửa dưới biểu đồ tuần tự trạng thái hiển thị dữ liệu	40
Hình 4.8.	Biểu đồ tuần tự trạng thái tắt màn hình	41
Hình 4.9.	Ma trận hệ số tương quan của dữ liệu tại Triệu Ái	44
 Hình 5.1.	Bên ngoài thiết bị đầu cuối	52
Hình 5.2.	Bên trong thiết bị đầu cuối	53
Hình 5.3.	Màn hình hiển thị dữ liệu cảm biến	53
Hình 5.4.	Màn hình hiển thị dữ liệu lượng mưa dự báo tỉnh Hà Tĩnh . . .	54
Hình 5.5.	Giao diện dữ liệu lượng mưa dự báo tỉnh Quảng Bình	54
Hình 5.6.	Giao diện dữ liệu lượng mưa dự báo tỉnh Quảng Trị	55
Hình 5.7.	Token và chủ đề theo dõi trên nền tảng flespi	55
Hình 5.8.	Dữ liệu cảm biến được gửi đến nền tảng flespi	56
Hình 5.9.	Dữ liệu cập nhật giữa hệ thống máy chủ và thiết bị đầu cuối .	56
Hình 5.10.	Đánh giá quá trình huấn luyện trên tập Validation	57
Hình 5.11.	Đánh giá quá trình huấn luyện trên tập Validation	58
Hình 5.12.	lượng mưa dự đoán và thực tế theo bộ dữ liệu IFS	59
Hình 5.13.	lượng mưa dự đoán và thực tế trong tháng 1 năm 2025	60
Hình 5.14.	Lượng mưa dự đoán và thực tế theo bộ dữ liệu IFS và Vrain .	61
Hình 5.15.	Lượng mưa dự đoán và thực tế trong tháng 1 năm 2025	62
Hình 5.16.	Giao diện web trạm Kỳ Thượng - Hà Tĩnh	64

DANH MỤC BẢNG

Bảng 3.1. Bảng khoảng cách giữa các cặp điểm thu thập	29
Bảng 4.1. Tóm tắt mô hình	47

DANH MỤC TỪ VIỆT TẮT

ACK	Acknowledged
ADC	Analog_to_Digital Converter
Adam	Adaptive Moment Estimation
AI	Artificial Intelligence
ANN	Artificial Neural Network
DA	Data Assimilation
DIY	Do It Yourself
ECMWF	The European Centre for Medium-Range Weather Forecasts
GPIO	General Purpose Input/Output
GNSS	Global Navigation Satellite System
GWR	Gradient-weighted Class Activation
I2C	Inter-Integrated Circuit
IFS	Integrated Forecasting System
IoT	Internet of Things
LSTM	Long short-term memory
MAE	Mean Absolute Error
MQTT	Message Queuing Telemetry Transport
MSE	Mean Squared Error
NACK	Not Acknowledged
RNN	Recurrent neural network
RMSE	Root Mean Square Error
RTOS	Real Time Operating System
SANN	Seasonal Artificial Neural Network
SPI	Serial Peripheral Interface
UART	Universal Asynchronous Receiver/Transmitter
Wifi	Wireless Fidelity

TÓM TẮT KHÓA LUẬN

Với sự phát triển không ngừng của công nghệ thông tin, đặc biệt trong lĩnh vực trí tuệ nhân tạo và điện toán đám mây, đã mang đến nhiều giải pháp giúp giải quyết nhiều vấn đề của con người với hiệu năng cao và tiết kiệm chi phí. Đặc biệt là trong lĩnh vực khí tượng thuỷ văn, với bài toán dự báo lượng mưa. Với khu vực các tỉnh Bắc Trung Bộ, nơi thường xuyên phải hứng chịu những cơn mưa có cường độ lớn hàng năm từ những cơn bão hoặc áp thấp nhiệt đới do vị trí địa lý và địa hình của khu vực [1], gây ra những thiên tai như lũ quét, xói mòn, sạt lở đất và thiệt hại về tài sản người và của [2]. Vì thế, việc dự đoán được lượng mưa không chỉ góp phần nâng cao khả năng phòng chống thiên tai mà còn góp phần nghiên cứu xu thế khí tượng thuỷ văn trong tương lai, trong bối cảnh biến đổi khí hậu đang diễn biến ngày càng phức tạp [3].

Khoa luận này nghiên cứu và xây dựng một hệ thống áp dụng kiến trúc IoT (Internet of Things) nhằm thu thập dữ liệu khí tượng thuỷ văn theo thời gian thực và mô hình học sâu (Deep Learning) với vai trò tính toán lượng mưa dự đoán theo giờ mà một khu vực có thể hứng chịu. Dữ liệu lượng mưa đều được liên tục cập nhật và lưu trữ trên hệ thống máy chủ và cũng được hiển thị đồng thời trên giao diện web và màn hình trên hệ thống phần cứng, đóng vai trò hỗ trợ người dùng quan sát lượng mưa dự đoán.

Với hệ thống thiết bị đầu cuối, nhóm sử dụng mô đun Goouuu-ESP32 có khả năng giao tiếp với máy chủ thông qua tín hiệu wifi, mô đun ESP32 này sẽ điều khiển các cảm biến đo đặc ba yếu tố khí tượng: nhiệt độ, độ ẩm, áp suất không khí. Với nền tảng công nghệ hiện thực các chức năng trên là ESP-IDF, một nền tảng mạnh mẽ, tận dụng tối đa các tính năng và khả năng của dòng chip ESP trong việc xây dựng các ứng dụng hệ thống IoT.

Về hệ thống máy chủ, nhóm sử dụng mô hình mạng nơ ron LSTM (Long Short Term Memory), là một mạng nơ-ron hồi tiếp được thường được áp dụng trong dự đoán dữ liệu chuỗi thời gian, phù hợp với bài toán dự báo lượng mưa dựa vào các yếu tố khí tượng thuỷ văn từ 24 giờ trước đó. Lượng mưa dự báo 24 giờ tiếp theo sẽ được hiển thị trên một giao diện web nhằm hỗ trợ người dùng quan sát, theo dõi.

Chương 1. GIỚI THIỆU ĐỀ TÀI

1.1 Đặt vấn đề

Trong bối cảnh biến đổi khí hậu đang diễn ra ngày càng phức tạp, với sự xuất hiện ngày càng nhiều những cơn bão nhiệt đới ở khu vực Biển Đông và Thái Bình Dương, mang theo lượng mưa lớn kỷ lục hàng năm vào khu vực đất liền các tỉnh Bắc Trung Bộ gây ra những thiên tai bên cạnh bão như sạt lở đất, lũ quét, ... gây ra thiệt hại không nhỏ về tài sản cũng như đe doạ đến sự an toàn của người dân. Vì vậy, bài toán dự báo lượng mưa luôn là một đề tài được nghiên cứu nhằm giảm thiểu thiệt hại hay xa hơn là có những biện pháp ứng phó lâu dài giúp người dân có thể tái sản xuất, kinh tế xã hội không bị đình trệ sau những trận mưa lớn.

Những hệ thống dự báo lượng mưa trên thế giới đều ứng dụng trí tuệ nhân tạo với những mô hình học máy, học sâu với mạng nơ ron nhằm đưa ra dự đoán về lượng mưa không chỉ trong ngắn hạn (vài giờ cho đến vài ngày) [4] mà còn trong dài hạn (từ nhiều tuần đến nhiều tháng) [5]. Các mô hình thường dự báo lượng mưa dựa trên những yếu tố khí tượng thuỷ văn: nhiệt độ, độ ẩm, áp suất không khí, tốc độ gió, ... được thu thập thông qua các cảm biến được đặt tại các trạm đo khí tượng thuỷ văn thế giới, hoặc là những dữ liệu được thu thập từ vệ tinh hoặc ra-đa, những dữ liệu này sau đó được tổng hợp tái phân tích trước khi mô hình ứng dụng để cho ra lượng mưa dự báo với quy mô diện tích lớn hơn.

Tuy vậy, nhưng bài toán dự báo lượng mưa vẫn đặt ra những thách thức về độ chính xác của các mô hình dự báo. Với những hệ thống dựa vào dữ liệu từ cảm biến, thì lượng mưa dự báo có thể mang tính cục bộ hoặc thiếu chính xác do sai số từ phần cứng [6]. Vì vậy đòi hỏi một mạng lưới các trạm đo khí tượng thuỷ văn nhằm mở rộng quy mô dự báo. Còn những mô hình ứng dụng vệ tinh hay ra đa thời tiết có hạn chế về độ chính xác dữ liệu thu thập do ảnh hưởng bởi các yếu tố khác như mây, sương mù hay những hiệu ứng tác động từ bầu khí quyển [7]. Bên cạnh những vấn đề về dữ liệu đầu vào mô hình, một yếu tố cũng ảnh hưởng không nhỏ chính là độ tin cậy dữ liệu khí tượng thuỷ văn lịch sử tại khu vực dự báo. Với các hiện tượng thời tiết cực đoan

ngày càng diễn biến phức tạp, với tần suất thường xuyên hơn và cường độ đột biến ảnh hưởng đến hiệu các mô hình dự báo dựa trên dữ liệu quá khứ [8].

Trong khuôn khổ khoá luận tốt nghiệp, nhóm đề xuất một mô hình dự báo tận dụng cả hai yếu tố từ dữ liệu cảm biến và những yếu tố khác từ những nguồn dữ liệu vệ tinh, ra đa nhằm có được sự đa dạng từ dữ liệu đầu vào. Kết hợp với thuật toán dự đoán dựa trên mạng nơ ron LSTM, giúp góp phần gia tăng độ chính xác về lượng mưa dự báo theo giờ cho 24 giờ tiếp theo.

1.2 Lý do chọn đề tài

Trong bối cảnh tình hình biến đổi khí hậu đang ngày càng diễn biến phức tạp, cụ thể là hiện tượng La Nina thường diễn ra ở nửa cuối năm gây ra những trận mưa lớn đột biến từ những vùng áp thấp hay thậm chí là những cơn bão [9]. Vì thế dự báo lượng mưa ở các tỉnh Bắc Trung Bộ nói riêng và cả nước Việt Nam nói chung vẫn là một đề tài luôn nóng hổi bởi tính cấp thiết của nó.

Việc hệ thống ứng dụng hai nguồn dữ liệu khí tượng thuỷ văn đầu vào không chỉ làm đa dạng nguồn dữ liệu, kết hợp với mạng nơ ron LSTM không chỉ giúp gia tăng độ chính xác mô hình mà còn có thể phát triển trong tương lai với sự cải tiến cả về mặt phần cứng và phần mềm.

Nhóm tin rằng đề tài không chỉ mang tính học thuật mà còn có tính ứng dụng thực tế. Ở Việt Nam, khu vực các tỉnh thành Bắc Trung Bộ là nơi thường xuyên chịu những trận mưa lớn hằng năm - Trung bình mỗi năm, Bắc Trung Bộ có khoảng 9-10 cơn bão, 3-4 cơn áp thấp nhiệt đới, 17-18 trận dông, tố, lốc, mưa đá [10]. Vì vậy, nếu có thể dự báo được những cơn mưa có cường độ đột biến ở khu vực này thì trong tương lai có thể ứng dụng ở các khu vực khác có lượng mưa ít khắc nghiệt hơn. Qua đó góp phần xây dựng mạng lưới thông tin dự báo khí tượng thuỷ văn cho cả nước một cách chính xác và nhanh chóng.

Hệ thống dự báo được ứng dụng mô hình học sâu và kiến trúc IoT góp phần thúc đẩy cả hai lĩnh vực trí tuệ nhân tạo và hệ thống nhúng trong việc giải quyết các vấn đề của người, qua đó có thể liên tục cải tiến sản phẩm để gia tăng hiệu năng hoặc giải quyết vấn đề tiết kiệm năng lượng. Ngoài ra, những cơ sở, nền tảng tri thức về lĩnh

vực khí tượng thuỷ văn được ứng dụng trong mô hình góp phần mở rộng kiến thức về mảng này, qua đó có thể có những ý tưởng sản phẩm liên quan đến như mô hình dự báo hạn hán, xâm ngập mặn.

1.3 Đối tượng và phạm vi thực hiện

1.3.1 Đối tượng nghiên cứu

Đối tượng nghiên cứu của nhóm bao gồm: Các nội dung lý thuyết, thực hành liên quan đến hệ thống nhúng và IoT, mô hình học sâu trên hệ thống máy chủ. Hiện thực chức năng cho mô đun nhúng, bo mạch chủ nhằm hiện thực các chức năng thu thập dữ liệu từ cảm biến và giao tiếp với máy chủ. Nghiên cứu và hiện thực các giải thuật dự báo lượng mưa, tối ưu những yếu tố chính ảnh hưởng đến lượng mưa. Hiện thực các giao diện người dùng trên màn hình trên thiết bị và giao diện web.

1.3.2 Phạm vi thực hiện

Về phạm vi thời gian, đề tài được nghiên cứu và phát triển trong thời gian diễn ra khóa luận tốt nghiệp.

Về phạm vi không gian, đề tài hướng đến việc dự báo lượng mưa cho 3 tỉnh Bắc Trung Bộ. Cụ thể là các tỉnh Hà Tĩnh, Quảng Bình, Quảng Trị. Đây là ba tỉnh đại diện cho khu vực nghiên cứu, cũng là những tỉnh chịu ảnh hưởng nặng nề từ những trận mưa lớn, kéo theo sạt lở, lũ quét, xói mòn.

Về phạm vi nội dung, đề tài sẽ giới hạn lại trong vấn đề vận hành hệ thống IoT bao gồm thiết bị phần cứng và máy chủ, kết nối thiết bị ngoại vi (ở đây là các cảm biến) từ đó thu thập dữ liệu đo đạc được rồi gửi lên máy chủ, sau đó xử lý và xuất ra lượng mưa dự báo trong 24 giờ tiếp theo ở mỗi tỉnh trong ba tỉnh thành trên.

1.3.3 Mục tiêu nghiên cứu

Nghiên cứu và hiện thực được một hệ thống dự báo lượng mưa bao gồm hệ thống phần cứng và máy chủ. Hệ thống sử thu thập dữ liệu thuỷ văn từ các cảm biến phần cứng: nhiệt độ, độ ẩm và áp suất không khí. Sau đó sẽ được tổng hợp với các yếu tố khác được thu thập từ dữ liệu ra đa dạng rồi từ đó sẽ đi qua mạng nơ ron LSTM tính toán được lượng mưa dự kiến trong 24 giờ tiếp theo. Lượng mưa dự báo sẽ được hiển

thị trên giao diện người dùng nhằm hỗ trợ công việc quan sát và đánh giá, cụ thể ở đây là màn hình phần cứng và giao diện web. Như vậy, hệ thống dự báo lượng mưa khu vực các tỉnh Bắc Trung Bộ bao gồm 3 phần:

Thiết bị đầu cuối: Đóng vai trò đo đạc dữ liệu khí tượng sau đó sẽ được tổng hợp và thực hiện giao tiếp với hệ thống máy chủ.

- Thu thập dữ liệu khí tượng: nhiệt độ, độ ẩm, áp suất không khí theo thời gian thực và được tổng hợp nhằm gửi về máy chủ qua đó đóng góp những đặc trưng vào lượng mưa dự báo.
- Giao tiếp với hệ thống máy chủ thông qua tín hiệu Wifi (Wireless Fidelity), giao thức MQTT (Message Queuing Telemetry Transport) nhằm cập nhật dữ liệu thời gian thực và dữ liệu lượng mưa dự báo ở các tỉnh thành nghiên cứu.
- Hiển thị các thông số - thời gian thực và dữ liệu lượng mưa lên màn hình oled được tích hợp trong thiết bị đầu cuối.
- Hỗ trợ tương tác với người dùng thông qua những nút nhấn cũng được tích hợp trên thiết bị nhằm tắt/mở màn hình, chuyển đổi giao diện lượng mưa giữa các tỉnh: Hà Tĩnh, Quảng Bình, Quảng Trị.

Máy chủ điện toán đám mây: Là nơi lưu trữ cơ sở dữ liệu lịch sử những yếu tố khí tượng và lượng mưa, đồng thời liên tục cập nhật dữ liệu đo đạc và lượng mưa dự đoán tại các tỉnh thành đang nghiên cứu thông qua mô hình học sâu với mạng LSTM. Đây cũng là nền tảng cho giao diện trang web.

Thành phần cuối cùng chính là giao diện người dùng - trang web nhằm hỗ trợ quan sát lượng mưa dự báo các điểm nghiên cứu thuộc 3 tỉnh: Hà Tĩnh, Quảng Bình, Quảng Trị. Lượng mưa dự báo được hiển thị dựa trên theo 3 nguồn: mô hình nội suy IFS-ECMWF (Integrated Forecasting System - European Centre for Medium-Range Weather Forecasts), mô hình dự báo lượng mưa dựa trên mạng nơ ron LSTM đối với bộ dữ liệu IFS và lượng mưa theo hệ thống cơ sở dữ liệu Vrain.

1.4 Phương pháp thực hiện

Dưới đây là các phương pháp hiện thực hệ thống bao gồm thiết bị đầu cuối và hệ thống máy chủ cho hệ thống dự báo lượng mưa khu vực các tỉnh Bắc Trung Bộ:

- Nghiên cứu, tìm hiểu các bài báo và công trình nghiên cứu khoa học về các hệ thống dự báo lượng mưa ứng dụng mô hình học máy, học sâu. Hoặc từ những hệ thống sử dụng dữ liệu cảm biến hay dữ liệu vệ tinh, ra đa.
- Phân tích, nghiên cứu các yếu tố khí tượng, thuỷ văn ảnh hưởng đến lượng mưa trong khoảng thời gian 24 giờ tiếp theo.
- Hiện thực chức năng thu thập dữ liệu khí tượng từ môi trường xung quanh từ hệ thống thiết bị nhúng.
- Nghiên cứu và hiện thực các giải thuật dự báo lượng mưa từ những tham số dữ liệu đầu vào (từ cảm biến và từ dữ liệu ra đa).
- Tìm hiểu, hiện thực máy chủ lưu trữ dữ liệu khí tượng thuỷ văn tại khu vực nghiên cứu và liên tục cập nhật dữ liệu theo giờ.
- Lắp đặt phần cứng, đóng gói các thiết bị, mạch điện tử nhằm có thể áp dụng ngoài tự nhiên mà vẫn đảm bảo hiệu năng hoạt động tốt.
- Nghiên cứu, xây dựng quy trình vận hành cho toàn hệ thống, hiện thực giao diện người dùng, trang web và trên màn hình oled được tích hợp trên thiết bị đầu cuối.
- Xây dựng các phương pháp triển khai và đánh giá kết quả của hệ thống, từ đó xác định những điểm còn hạn chế và đề ra hướng phát triển tiếp theo.

Chương 2. TỔNG QUAN ĐỀ TÀI

Các nghiên cứu về mô hình dự đoán lượng mưa trở lại đây thường áp dụng những kiến trúc mô hình học sâu với nhiều yếu tố khí tượng, thuỷ văn từ những cảm biến hay ra-đa, vệ tinh nhằm tối đa độ chính xác của mô hình dự báo lượng mưa. Với lượng dữ liệu cần được xử lý lớn và liên tục theo giờ, đòi hỏi những hệ thống máy chủ cần được phát triển để đáp ứng yêu cầu đặt ra.

2.1 Tình hình nghiên cứu quốc tế

Trong bài báo [11], tác giả đã xây dựng mô hình dự báo dựa trên dữ liệu đầu vào là ảnh vệ tinh và tập trung vào độ phân giải cao của khu vực dự đoán, cụ thể ở đây là 1km. Tuy vậy, thời gian dự đoán của mô hình là một giờ tiếp theo chỉ phù hợp với công việc đưa ra ngay lập tức. Ngoài ra cần phải giải quyết các hiệu ứng đường viền trong quá trình trích xuất dữ liệu từ ảnh vệ tinh.

Tác giả bài báo [12] đã ứng dụng mạng nơ ron CRNN (Convolutional Recurrent Neural Networks) để dự báo lượng mưa từ dữ liệu ra-đa 3D giúp cải thiện khả năng thu thập dữ liệu so với ảnh ra-đa 2D. Tuy nhiên, hiệu suất của mô hình vẫn chưa được đảm bảo và chuỗi thời gian đầu vào của mô hình là 15 bước, có thể khó khăn trong việc dự báo lượng mưa từ những trận bão.

Ở bài báo [13], tác giả đã kết hợp dữ liệu thu thập từ nhiều nguồn khác nhau: ra-đa, vệ tinh và các thông số bề mặt, từ đó giúp cải thiện độ chính xác mô hình bên cạnh ứng dụng 2 mạng nơ ron GWR (Gradient-weighted Class Activation) và ANN (Artificial Neural Network). Tuy nhiên, mô hình cần nghiên cứu và phát triển các phương pháp hồi quy phi tuyến tính khác hoặc cải tiến GWR để tăng cường khả năng xử lý các mối quan hệ phức tạp giữa các yếu tố khí tượng thuỷ văn.

Một bài báo khác về mô hình dự báo đã tổng hợp được dữ liệu từ hệ thống định vị toàn cầu [14] (GNSS - Global Navigation Satellite System) và các cảm biến dự báo, cho ra tỉ lệ phân loại tình trạng mưa hay không mưa đạt từ 63% - 72%. Tuy nhiên, bộ dữ liệu của mô hình vẫn chưa có độ dài đủ lớn (5 năm) để có thể nắm bắt được những trận mưa lớn, đặc biệt là ở mùa mưa khu vực nghiên cứu.

2.2 Tình hình nghiên cứu trong nước

Được đề xuất bởi [15], nhóm tác giả đã giới thiệu mô hình kết hợp giữa hai phương pháp tiền xử lý với hai loại mạng nơ-ron (ANN và SANN - Seasonal Artificial Neural Network) để cải thiện dự đoán lượng mưa hàng tháng. Mô hình đã dựa trên bộ dữ liệu lượng mưa từ trạm thủy văn Cà Mau trong khoảng thời gian 39 năm (1971-2010), đảm bảo tính chính xác và độ tin cậy của mô hình. Tuy vậy, mô hình được áp dụng với lượng mưa dự báo hàng tháng vì thế vẫn chưa thể áp dụng rộng rãi do chưa có số liệu chi tiết theo giờ.

Áp dụng dữ liệu ra-đa thời tiết để cải thiện công tác dự báo mưa lớn từ bão và áp thấp nhiệt đới [16] tại khu vực Bắc Trung Bộ, nhằm cung cấp thông tin thời tiết trong mọi điều kiện, đặc biệt là trong các tình huống thời tiết phức tạp mà các phương pháp quan trắc truyền thống không thể thực hiện được. Bài báo này tập trung vào việc phân tích "độ phản hồi của vật thể" từ những bức ảnh thu thập từ ra-đa thời tiết, kết hợp với các tài liệu lịch sử từ năm 1993 - 2017 như số liệu các thế hệ ra-đa, 21 trạm khí tượng và 3 trạm Hải văn thuộc khu vực Bắc Trung Bộ, bên cạnh các số liệu thông tin về bão, áp thấp nhiệt đới và mưa lớn từ năm 1980 - 2017. Các số liệu sẽ được tổng hợp, tính toán chọn lọc. Thông qua đó xây dựng các công cụ hỗ trợ cảnh báo, dự báo quỹ đạo của bão, áp thấp nhiệt đới và mưa lớn từ trường mây bão gây ngập lụt ở khu vực Bắc Trung Bộ cho các thời đoạn dưới 3 giờ, dưới 6 giờ, dưới 9 giờ và dưới 12 giờ.

Bên cạnh những nguồn dữ liệu được ứng dụng trong mô hình dự báo, thì tại Việt Nam, mô hình nội suy Trung tâm Châu Âu về Dự báo Thời tiết Trung bình (ECMWF - The European Centre for Medium-Range Weather Forecasts) với độ phân giải 9km [17] cho kết quả lượng mưa dự báo tốt hơn các mô hình tái phân tích khác GFS, NOGAPS (Mỹ) hay GSM của Nhật [18]. Mô hình IFS (ECMWF) dự báo tốt ở lượng mưa nhỏ và vừa, nhưng cho hiệu suất kém với những trận mưa to, cụ thể ở bài báo [19] đã đánh giá dự báo mưa ngắn hạn từ mô hình IFS (ECMWF) tại Bắc Trung Bộ, sử dụng số liệu 20 trạm trong 6 năm (2012-2017) kết quả cho thấy chất lượng mô hình giảm ở hạn dự báo 72h, tuy vẫn nắm bắt được một số đợt mưa lớn.

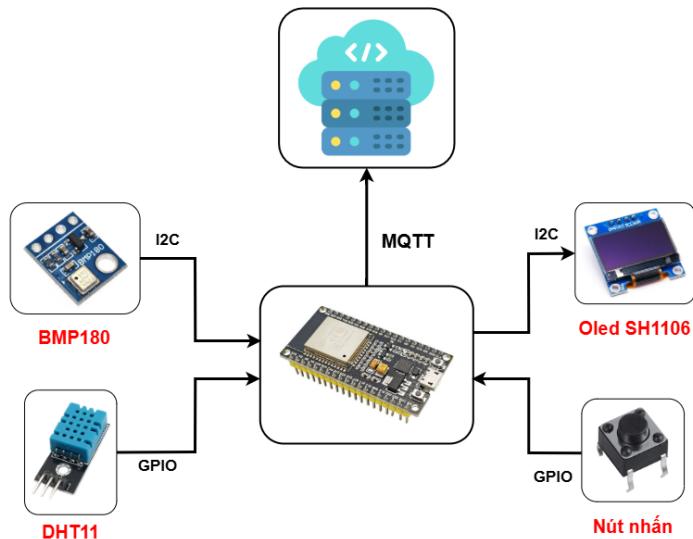
Chương 3. CƠ SỞ LÝ THUYẾT

3.1 Thiết bị đầu cuối

Thiết bị đầu cuối đóng vai trò thu thập dữ liệu khí tượng từ môi trường bên ngoài: nhiệt độ, độ ẩm và áp suất không khí. Dữ liệu sau đó được gửi về hệ thống máy chủ với tần suất 20 phút / lần, qua đó hỗ trợ công việc dự báo lượng mưa các tỉnh nghiên cứu khu vực Bắc Trung Bộ.

Bên cạnh đó người dùng có thể quan sát lượng mưa dự báo 3 tỉnh: Hà Tĩnh, Quảng Bình, Quảng Trị trên màn hình oled được tích hợp sẵn trên thiết bị thông qua các nút nhấn được hỗ trợ cho giao tiếp giữa người dùng nhằm tắt/mở, thay đổi giao diện của màn hình.

Hình 3.1 thể hiện tổng quan các thành phần trong hệ thống thiết bị đầu cuối và giao thức giao tiếp kết nối giữa các thành phần với bo mạch chủ và phương thức truyền dữ liệu giữa thiết bị biên và hệ thống máy chủ.



Hình 3.1: Sơ đồ kết nối các thành phần thiết bị đầu cuối

Quan sát hình 3.1, ta thấy thành phần cấu thành nên hệ thống thiết bị biên nhằm hiện thực chức năng thu thập số liệu khí tượng và giao tiếp với người dùng bao gồm:

- Bo mạch ESP32 (Goouuu-ESP32): Đóng vai trò điều khiển, thu thập và xử lý dữ liệu từ những cảm biến khí tượng và giao tiếp kết nối với máy chủ mô hình.

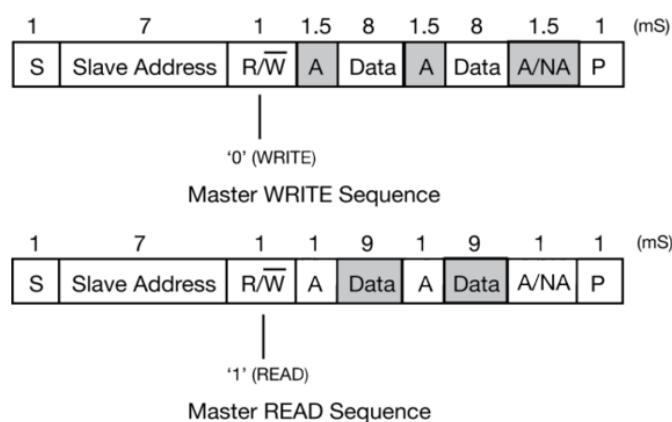
Những yếu tố đầu thu thập từ các cảm biến: nhiệt độ, độ ẩm, áp suất không khí sẽ được đưa về bo Goouuu-ESP32 để được xử lý. Dữ liệu sau khi được thu thập sẽ được chia làm hai hướng: cập nhật số liệu trên màn hình oled và đóng gói lại gửi lên máy chủ để phục vụ công việc dự đoán lượng mưa thông qua kết nối Wifi, giao thức MQTT. Không chỉ vậy, Goouuu-ESP32 còn liên tục cập nhật những số liệu theo yêu cầu từ phía máy chủ khi lượng mưa dự báo có thay đổi.

- Cảm biến đo nhiệt độ, áp suất (BMP180): Giao tiếp thông qua giao thức I2C. Đóng vai trò là thiết bị thu thập nhiệt độ và áp suất không khí môi trường xung quanh. Dữ liệu thu thập được sẽ được gửi về bo mạch Goouuu-ESP32 theo chu kỳ 3 giây một lần.
- Cảm biến đo độ ẩm không khí (DHT11): Là cảm biến đo đặc độ ẩm không khí môi trường xung quanh. Bo mạch Goouuu-ESP32 giao tiếp với cảm biến này thông qua giao thức 1 dây, nghĩa là tín hiệu chỉ cần được truyền thông qua một chân duy nhất. Cảm biến này sẽ gửi dữ liệu về bo mạch chủ mỗi 3 giây một lần.
- Màn hình oled (SH1106): Đóng vai trò như giao diện người dùng ngay tại hệ thống phần cứng, hiển thị thông tin về dữ liệu nhiệt độ, độ ẩm, áp suất không khí từ 2 cảm biến BMP180 và DHT11 cũng như dữ liệu lượng mưa dự đoán. Màn hình oled được điều khiển bởi bo mạch Goouuu-ESP32 thông qua giao thức I2C.
- Nút nhấn (button): Thành phần này đóng vai trò như những đầu vào được điều khiển bởi người dùng. Hệ thống phần cứng có tổng cộng 4 nút: nút nguồn, nút trái, nút giữa và nút phải. Những nút này đóng vai trò như đầu vào bật/tắt màn hình oled và chuyển đổi giữa các màn hình hiển thị.
- Hộp nhựa kỹ thuật: là thành phần đóng vai trò đóng gói toàn bộ hệ thống phần cứng, qua đó cố định các thành phần, không chỉ giúp gia tăng tính thẩm mĩ mà còn có thể bảo vệ các thiết bị, cảm biến bên trong. Từ đó sẵn sàng ứng dụng hệ thống phần cứng ra ngoài môi trường thực tế.

Về giao thức kết nối, như đã thể hiện trên hình 3.1, bo mạch chủ sẽ giao tiếp với cảm biến DHT11 - dùng để đo độ ẩm không khí và nút nhấn 4 chân thông qua việc cấu hình chân GPIO (General Purpose Input/Output). Từ đó có thể xử lý tín hiệu độ ẩm và trạng thái của nút nhấn mỗi khi dữ liệu được truyền từ cảm biến về bo mạch chủ hoặc khi người dùng nhấn nút.

Đối với cảm biến BMP180 - đo nhiệt độ và áp suất không khí và màn hình oled SH1106 - có chức năng hiển thị giao diện người dùng, giao thức được ứng dụng nhằm truyền tải dữ liệu giữa bo mạch Goouuu-ESP32 và 2 thiết bị là giao thức I2C (Integrated Circuit). Đây là giao thức đồng bộ 2 đường truyền tín hiệu (Two Wired Interface – TWI) đó là: SCL (Serial Clock Line) - truyền các tín hiệu xung đồng bộ và SDA (Serial Data Line) - truyền dữ liệu giữa các thiết bị. Các trường dữ liệu trong một khung theo chuẩn I2C ở hình 3.2 bao gồm:

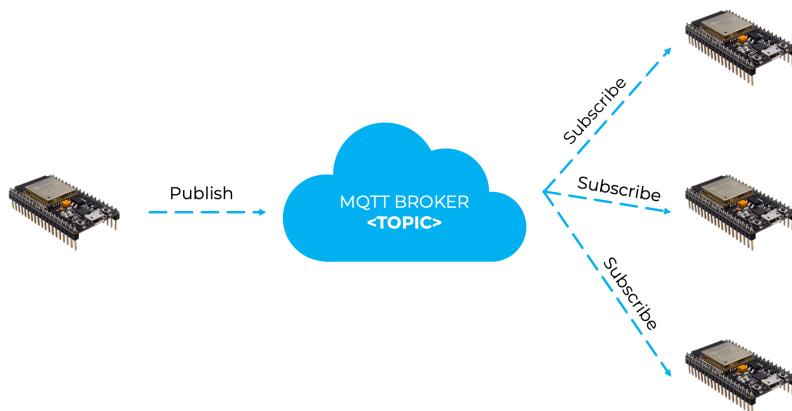
- Khung địa chỉ: Chỉ địa chỉ duy nhất cho thiết bị đích.
- Bit đọc / ghi: Cho biết Master đang gửi dữ liệu đến Slave hay đang đọc từ nó.
- Bit ACK / NACK (Acknowledged / Not Acknowledged). Dùng để so sánh bit địa chỉ vật lý của thiết bị so với địa chỉ được gửi tới.
- Dữ liệu: Đoạn thông tin được truyền đi giữa các thiết bị trong đường truyền.



Hình 3.2: Khung truyền dữ liệu giao thức I2C

Giao thức I2C kết hợp các tính năng tốt nhất của SPI và UART. I2C có thể kết nối nhiều slave với một master duy nhất (như SPI) và có thể có nhiều master điều khiển một hoặc nhiều slave, giúp giảm số lượng dây dẫn cần thiết, tiết kiệm không gian và tăng tính linh hoạt trong việc truyền tải dữ liệu.

Đối với việc truyền tải dữ liệu giữa thiết bị biên và hệ thống máy chủ, giao thức được sử dụng chính là MQTT (Message Queuing Telemetry Transport). Đây là một giao thức gửi tin nhắn nhẹ, được thiết kế đặc biệt cho các thiết bị có tài nguyên hạn chế và mạng không đáng tin cậy, như các thiết bị IoT. Giao thức MQTT hoạt động dựa trên mô hình xuất bản (publish)/đăng ký (subscribe). Các thiết bị IoT có thể đăng ký (subscribe) vào các chủ đề (topics) và gửi (publish) các thông điệp đến các chủ đề đó. Máy chủ MQTT nhận các thông điệp từ các thiết bị đã đăng ký và chuyển tiếp đến các thiết bị khác đang quan tâm đến chủ đề tương ứng. Hình 3.3 minh họa quá trình giao tiếp MQTT giữa các thiết bị trong mạng lưới IoT.



Hình 3.3: Mô hình kết nối MQTT

Để triển khai kiến trúc IoT dựa trên điện toán đám mây, nhóm lựa chọn nền tảng Flespi nhằm quản lý dữ liệu thiết bị và giao tiếp qua giao thức MQTT. Flespi hỗ trợ phiên bản MQTT 5.0 cùng các mức QoS (Quality of Service) 0, 1 và 2, giúp đảm bảo tính linh hoạt và hiệu suất vượt trội. Nền tảng này phù hợp cho các ứng dụng IoT yêu cầu khả năng mở rộng, hiệu năng cao và hoạt động ổn định trong thời gian thực.

Và để hiện thực các chức năng giao tiếp, truyền tải dữ liệu cho bo mạch chủ Goouuu-esp32. Nhóm đã sử dụng ESP-IDF (Espressif IoT Development Framework).

Nền tảng này cung cấp SDK (Software Development Kit) hoàn chỉnh cho việc phát triển ứng dụng trên các nền tảng này, sử dụng các ngôn ngữ lập trình như C và C++. Một số đặc điểm về nền tảng này:

- Hỗ trợ đa giao thức kết nối: ESP-IDF hỗ trợ nhiều giao thức kết nối như Wi-Fi, Bluetooth, Ethernet, giúp phát triển các ứng dụng IoT đa dạng.
- Thư viện phong phú: Cung cấp nhiều thư viện và API để tương tác với các ngoại vi cũng như các giao thức mạng.
- Tích hợp hệ điều hành thời gian thực (RTOS - Real Time Operating System): Tích hợp thư viện FreeRTOS, cho phép quản lý đa nhiệm và tài nguyên hệ thống hiệu quả.
- Nền tảng ESP-IDF đi kèm với các công cụ như trình biên dịch, gỡ lỗi, và hệ thống xây dựng chương trình trên các hệ điều hành khác nhau.
- Nền tảng ESP-IDF đi kèm với tài liệu chi tiết về các thành phần phần mềm và cả các ví dụ có sẵn, giúp các nhà phát triển nắm bắt và áp dụng dễ dàng.

Thông qua nền tảng ESP-IDF, nhóm sẽ hiện thực các chức năng thu thập dữ liệu khí tượng - dựa trên các cảm biến, giao tiếp với hệ thống máy chủ - thông qua giao thức truyền tải không dây và cũng là thiết bị giao tiếp với người dùng thông qua các nút nhấn và màn hình oled.

3.2 Máy học và Học sâu

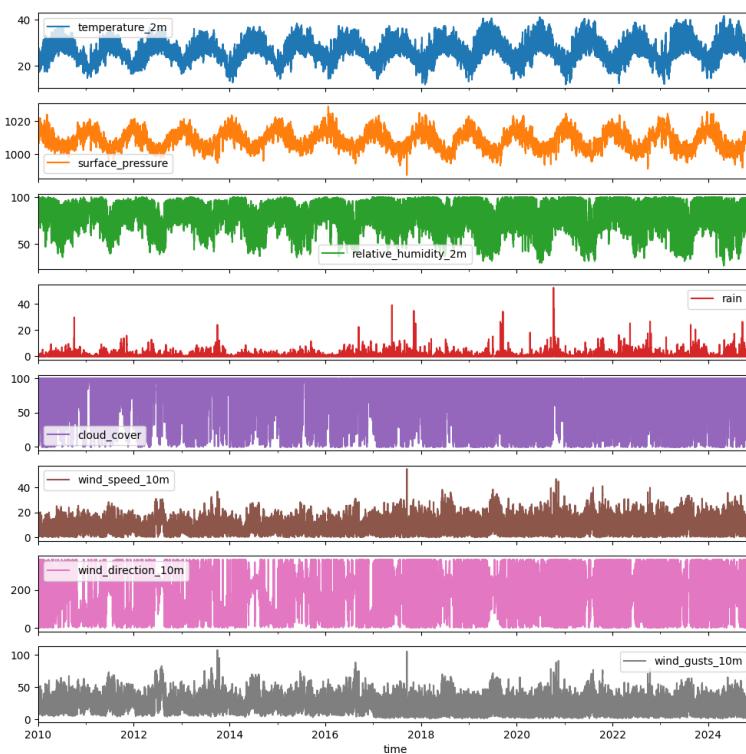
3.2.1 Lý thuyết thống kê

Hệ thống này ứng dụng dữ liệu thời tiết với nhiều yếu tố khí tượng thuỷ văn khác nhau nên việc tính toán, thống kê các đại lượng là điều cần thiết thông qua việc xem các yếu tố này là các đại lượng ngẫu nhiên liên tục với các phân phối tích lũy khác nhau.

Các đặc trưng thống kê thông dụng gồm: Giá trị trung bình, lớn nhất, nhỏ nhất, độ lệch chuẩn,... và các chỉ số khác giúp cung cấp cái nhìn tổng quan về đặc điểm cơ bản của dữ liệu, từ đó hỗ trợ quá trình phân tích và xây dựng mô hình trong thống kê

và học máy. Giá trị trung bình giúp xác định trung tâm của dữ liệu, trong khi độ lệch chuẩn cho thấy mức độ phân tán của dữ liệu xung quanh giá trị trung bình, từ đó hỗ trợ đánh giá sự ổn định. Các giá trị cực đại và cực tiểu cho phép hiểu rõ được phạm vi của dữ liệu, giúp nhận diện các ngoại lệ hoặc điểm bất thường (outliers). Những chỉ số này rất quan trọng trong việc kiểm tra tính đồng nhất và phân bố của dữ liệu, điều này đặc biệt hữu ích trong học máy khi lựa chọn mô hình phù hợp và điều chỉnh tham số. Ngoài ra, chúng còn giúp phát hiện các vấn đề như dữ liệu mất cân bằng (imbalanced data), xác định độ chính xác của mô hình và giúp cải thiện các thuật toán bằng cách hiểu rõ hơn về tính chất của dữ liệu đầu vào.

Xem xét phân phối của các yếu tố thời tiết để đưa ra các quyết định tính toán phù hợp thông qua việc trực quan dữ liệu bằng đồ thị. Việc trực quan hóa giúp ta dễ dàng hiểu và nhận diện các mẫu, xu hướng, và mối quan hệ trong dữ liệu mà có thể không rõ ràng qua các con số hay bảng biểu. Đặc biệt là dữ liệu thời tiết, ta có thể tìm ra được các quy luật theo mùa, tháng, năm thông qua việc trực quan hóa dữ liệu.



Hình 3.4: Trực quan dữ liệu khí tượng thuỷ văn tại trạm Triệu Ái, Quảng Trị

Quan sát hình 3.4, ta thấy rằng dữ liệu thời tiết thường có quy luật tuần hoàn theo mùa hoặc năm. Nhưng cũng có trường hợp như dữ liệu lượng mưa và độ phủ mây, dữ liệu phân phối không đồng đều và cũng không dễ tìm được quy luật. Do là dữ liệu theo giờ nên lượng mưa và độ phủ mây sẽ rất biến động, có thể bằng 0 suốt nhiều giờ liền và sau đó đột ngột tăng lên cao, khiến cho các mô hình thống kê hay học máy cổ điển gặp bất lợi khi tính toán. Do địa điểm khảo sát là các tỉnh thành giáp biển nên hướng gió có xu hướng thay đổi theo ngày đêm. Cụ thể, gió từ biển thổi vào đất liền vào ban ngày bởi ban ngày không khí ở ngoài biển nóng trở nên nhẹ rồi bốc lên cao và bị thổi đẩy vào đất liền, ngược lại, gió từ đất liền thổi ra biển vào ban đêm là bởi ban đêm không khí ở ngoài biển lạnh nên trở nên nặng hơn, đi xuống thấp và gió từ đất liền bị thổi ra biển.

Thông qua những quy luật vật lý, địa lý ta cũng có thể nhận ra các yếu tố thời tiết có ảnh hưởng và tác động qua lại với nhau, vì thế các yếu tố khí tượng thuỷ văn sẽ không thể là các biến độc lập trong thống kê mà phụ thuộc lẫn nhau.

3.2.2 Mô hình Random Forest

Đây là mô hình được phát triển từ mô hình Cây quyết định (Decision Tree), cải tiến những hạn chế của mô hình cũ như dễ gây quá khớp (overfitting) nếu số lượng các đặc trưng lớn, dẫn đến chỉ tạo ra một kịch bản dự báo duy nhất cho mỗi lần quan sát,... vì vậy mô hình Random Forest ra đời nhằm giải quyết những hạn chế đó. Sức mạnh của một cây quyết định đơn lẻ có thể bị hạn chế, nhưng khi nhiều cây quyết định kết hợp với nhau, chúng tạo nên một hệ thống mạnh mẽ hơn - giống như câu nói: "một cây làm chỏng nêu non, ba cây chụm lại nêu hòn núi cao". Ý tưởng của tính tập hợp của nhiều cây (ensemble) đã hình thành nên mô hình Random Forest.

Mô hình Random Forest bao gồm N cây quyết định. Để xây dựng mỗi cây, ta sử dụng một mẫu con ngẫu nhiên của dữ liệu huấn luyện thông qua phương pháp lấy mẫu Bootstrap (random sampling with replacement). Đây là phương pháp lấy mẫu có hoàn lại, có nghĩa là một cá thể có thể xuất hiện nhiều lần trong một lần lấy mẫu.

Với mỗi cây i ($i = 1, 2, \dots, N$), ta lấy một mẫu con B_i ngẫu nhiên từ tập huấn luyện $D = \{(x_j, y_j)\}_{j=1}^n$, trong đó x_j là các đặc trưng và y_j là nhãn của điểm dữ liệu thứ j .

Một cây quyết định được xây dựng trên mẫu con B_i bằng cách tối ưu hóa một hàm mất mát (thường là trung bình sai số bình phương).

Khi thực hiện dự đoán, mỗi cây trong rừng đưa ra một giá trị dự đoán $\hat{y}_i(x)$ cho một điểm đầu vào x . Dự đoán cuối cùng của rừng ngẫu nhiên là giá trị trung bình của các dự đoán từ tất cả các cây:

$$\hat{y}(x) = \frac{1}{N} \sum_{i=1}^N \hat{y}_i(x) \quad (3.1)$$

Trong đó: $\hat{y}_i(x)$ là dự đoán của cây thứ i , N là số lượng cây trong rừng.

Mỗi cây quyết định trong rừng ngẫu nhiên thực hiện hồi quy bằng cách phân chia dữ liệu theo các điểm quyết định, sau đó tính toán giá trị dự đoán cho mỗi lá cây dựa trên giá trị trung bình của nhãn y trong tập hợp các điểm dữ liệu thuộc lá đó.

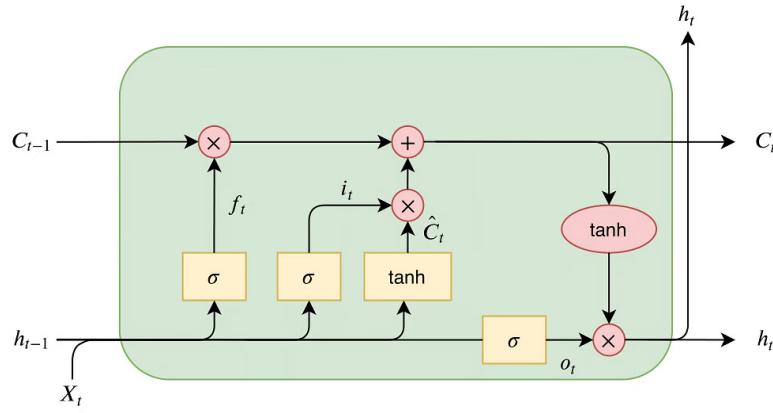
Giả sử, tại một lá cây t , ta có tập hợp các điểm dữ liệu $\{(x_j, y_j)\}_{j \in T}$, với T là chỉ số của các điểm dữ liệu thuộc lá t . Giá trị dự đoán tại lá cây t là trung bình của các giá trị y_j trong T :

$$\hat{y}_t = \frac{1}{|T|} \sum_{j \in T} y_j \quad (3.2)$$

Với $|T|$ là số lượng điểm dữ liệu trong lá t .

3.2.3 Mô hình Long Short Term Memory

Mạng nơ-ron LSTM là một kiến trúc đặc biệt được cải tiến từ mạng nơ-ron hồi quy (Recurrent neural network - RNN) có khả năng học được sự phục thuộc trong dài hạn (long-term dependencies) được giới thiệu bởi Hochreiter & Schmidhuber (1997). Kiến trúc này đã được phổ biến và sử dụng rộng rãi cho tới ngày nay. LSTM đã tỏ ra khắc phục được rất nhiều những hạn chế của RNN trước đây liên quan đến độ dốc. Tuy nhiên cấu trúc của chúng có phần phức tạp hơn mặc dù vẫn dữ được tư tưởng chính của RNN là sự sao chép các kiến trúc theo dạng chuỗi.



Hình 3.5: Kiến trúc mô hình LSTM

Ý tưởng chính của mạng nơ-ron LSTM là thành phần ô trạng thái (cell state) được thể hiện qua đường chạy ngang qua đỉnh đồ thị. Thành phần trạng thái giống như một băng chuyền. Nó chạy thẳng xuống toàn bộ chuỗi, chỉ có một số tương tác tuyến tính nhỏ. Rất dễ dàng để thông tin di chuyển theo nó mà không thay đổi, nên thông tin có thể truyền đi được xa hơn nên nó cũng chính là phần Long Term trong mạng này.

LSTM có khả năng xóa và thêm thông tin vào ô trạng thái và điều chỉnh các luồng thông tin này thông qua các cấu trúc gọi là cổng. Cổng là cơ chế đặc biệt để điều chỉnh luồng thông tin đi qua. Chúng được tổng hợp bởi một tầng ẩn của hàm activation sigmoid và với một toán tử nhân hình dấu X trên sơ đồ.

Hàm sigmoid sẽ cho ra là một giá trị xác suất nằm trong khoảng từ 0 đến 1, thể hiện rằng có bao nhiêu phần thông tin sẽ đi qua cổng. Giá trị bằng 0 ngụ ý rằng không cho phép thông tin nào đi qua, giá trị bằng 1 sẽ cho toàn bộ thông tin đi qua. Một mạng LSTM sẽ có 3 cổng có kiến trúc dạng này để bảo vệ và kiểm soát các ô trạng thái.

Cho một chuỗi đầu vào $\mathbf{x} = (x_1, x_2, \dots, x_T)$, lớp LSTM xuất ra một trạng thái ẩn \mathbf{h}_t và ô trạng thái \mathbf{c}_t . Việc lan truyền tiến (forward propagation) qua một lớp LSTM tại thời điểm t được xác định bởi:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.3)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.4)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.5)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot \tilde{c}_t \quad (3.6)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.7)$$

$$h_t = o_t \odot \tanh(c_t) \quad (3.8)$$

Trong đó σ là hàm sigmoid, \odot biểu thị phép nhân Hadamard là phép nhân được thực hiện theo từng cặp phần tử tương ứng của hai ma trận (element-wise), và W và b trọng số và bias tương ứng.

3.2.4 Các thang đo đánh giá hiệu suất

3.2.4.1 Chỉ số đánh giá cho bài toán phân loại

Nhằm đánh giá khả năng dự đoán của mô hình dựa trên khả năng đánh giá có mưa hay không mưa. Nhóm sử dụng những chỉ số dưới đây nhằm đưa ra độ chính xác trong việc dự đoán đúng hoặc sai trạng thái mưa. Các chỉ số này bao gồm Accuracy, Precision, Recall và F1 Score.

Trước tiên, ta hãy tìm hiểu các giá trị xuất phát trong ma trận nhầm lẫn, nhằm phân loại kết quả dự đoán của mô hình dự báo lượng mưa:

- TP - True Positive: Là giá trị thể hiện số mẫu (số giờ) mà mô hình dự báo chính xác là có mưa.
- TN - True Negative: Tổng số giờ mà mô hình dự báo chính xác là không mưa.
- FP - False Positive: Giá trị thể hiện số lượng mẫu mà mô hình dự đoán là có mưa nhưng thực tế là không mưa.
- FN - False Negative: Giá trị thể hiện số giờ mà mô hình dự đoán là không mưa nhưng thực tế là có mưa.

Như vậy, từ những giá trị trong ma trận nhầm lẫn, ta có thể tính toán được các chỉ số Accuracy, Precision, Recall và F1 Score và đánh giá được khả năng dự đoán trạng thái mưa của mô hình:

- Accuracy: Tỷ lệ số giờ dự đoán đúng (bao gồm cả có mưa và không mưa) trên tổng số mẫu . Được tính theo công thức:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.9)$$

- Precision: Là tỉ lệ giữa số giờ được dự đoán là có mưa đúng với thực tế so với tổng số giờ được dự báo là có mưa. Trong bài toán phân lớp nói chung, chỉ số này quan trọng khi các sai sót dự đoán dương cho ra kết quả không mong muốn.

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.10)$$

- Recall: Tỷ lệ giữa số giờ được mô hình dự đoán chính xác là có mưa so với tổng số giờ có mưa trên thực tế. Chỉ số này quan trọng trong các bài toán mà việc bỏ sót dữ liệu dương là điều không tốt.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.11)$$

- F1 Score: Giá trị trung bình dung hoà giữa Precision và Recall. Giá trị này quan trọng khi chúng ta phát triển mô hình theo hướng đạt trọng số Precision hay Recall cao hơn bởi thường làm giảm chỉ số còn lại. Giá trị F1 score càng tăng thể hiện mô hình càng có hiệu năng tốt.

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.12)$$

Đây là các chỉ số quan trọng trong việc đánh giá mô hình dự báo trạng thái mưa (có mưa hoặc không) - một bài toán đơn giản hơn so với dự báo lượng mưa theo chuỗi thời gian (bài toán hồi quy).

3.2.4.2 Chỉ số đánh giá cho bài toán hồi quy

Trong quá trình huấn luyện mô hình, nhóm sử dụng các thông số sau đây nhằm đánh giá độ lệch giữa lượng mưa dự đoán và dữ liệu lượng mưa thực tế. Những thông số này nhằm đánh giá độ chính xác từ số liệu dự báo của mô hình, từ đó có thể khắc phục những sai sót, hạn chế đồng thời cải thiện hiệu suất dự báo.

- Trung bình sai số toàn phương (MSE - Mean squared error) là giá trị trung bình của bình phương giữa giá trị dự đoán và giá trị thực tế. Giá trị MSE càng nhỏ, mô hình càng chính xác.

$$MSE = \frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2 \quad (3.13)$$

- Lấy căn bậc hai sai số bình phương trung bình (căn bậc hai của MSE) là chỉ số RMSE (Root Mean Squared Error). Giá trị RMSE càng thấp thấp biểu thị chất lượng dự báo tốt và mức độ sai số nhỏ.

$$RMSE = \sqrt{MSE} \quad (3.14)$$

- Trung bình sai số tuyệt đối (Mean Absolute Error - MAE) là trung bình của giá trị tuyệt đối các sai số. Giá trị MAE càng nhỏ cho thấy số liệu dự báo càng chính xác. Chỉ số MAE đặc biệt hữu ích trong việc xác định sai số tổng thể của dự báo mà không cần biết sai số có hướng nào.

$$MAE = \frac{1}{N} \sum_{i=1}^N |Y_i - \hat{Y}_i| \quad (3.15)$$

Trong đó, Y_i là lượng mưa thực tế, \hat{Y}_i giá trị dự đoán, và N là tổng số mẫu.

Đây là các chỉ số đánh giá quan trọng nhằm hỗ trợ phân tích khả năng dự đoán lượng mưa của mô hình. Đặc biệt với hạn mức dự báo của mô hình là 24 giờ - cũng là quãng thời gian tối mà bộ dữ liệu IFS-ECMWF được khuyến nghị sử dụng trong các mô hình lớn hơn.

Việc tối ưu hóa được thực hiện bằng trình tối ưu hóa Adam (Adaptive Moment Estimation) , một thuật toán nhằm tối ưu tốc độ học thích ứng, cập nhật các trọng số bằng cách sử dụng:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) \nabla L(\theta_t) \quad (3.16)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) (\nabla L(\theta_t))^2 \quad (3.17)$$

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \quad (3.18)$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \quad (3.19)$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t \quad (3.20)$$

Trong đó, m_t là v_t lần lượt là đường trung bình của độ dốc và độ dốc bình phương; η là tốc độ học; ϵ là một hằng số nhỏ để tránh việc chia cho 0; β_1 và β_2 là tốc độ phân rã theo cấp số nhân.

3.3 Nguồn dữ liệu

3.3.1 Mã nguồn mở OPEN-METEO

OPEN-METEO là một nền tảng cung cấp dữ liệu thời tiết miễn phí thông qua các API (Application Programming Interface). Đây là nền tảng cung cấp dữ liệu thời tiết, khí tượng thuỷ văn có thể truy cập và sử dụng một cách dễ dàng.

OPEN-METEO là một hệ thống mã nguồn mở đáng tin cậy, với nguồn dữ liệu đến từ các cơ quan khí tượng uy tín như ECMWF, NOAA (National Oceanic and Atmospheric Administration) của Hoa Kỳ. Qua đó cung cấp dữ liệu thời tiết chính xác cho bất kỳ địa điểm nào trên thế giới. Những đặc điểm của hệ thống mã nguồn mở này bao gồm:

- Đây là mã nguồn miễn phí và không yêu cầu API Key: Hệ thống này cho phép truy cập dữ liệu thời tiết miễn phí chỉ cần sử dụng các API được quy định.
- Người dùng có thể truy cập hệ thống này nhằm quan sát và sử dụng những dữ liệu từ quá khứ (từ năm 1940) đến nay hoặc dữ liệu dự đoán trong tương lai (thời hạn dự báo là 14 ngày) nhờ sự kết hợp của nhiều mô hình nội suy trên thế giới.

- Đa dạng những dữ liệu khí tượng thuỷ văn: nhiệt độ, độ ẩm, áp suất không khí, tốc độ và hướng gió, tia UV, ...

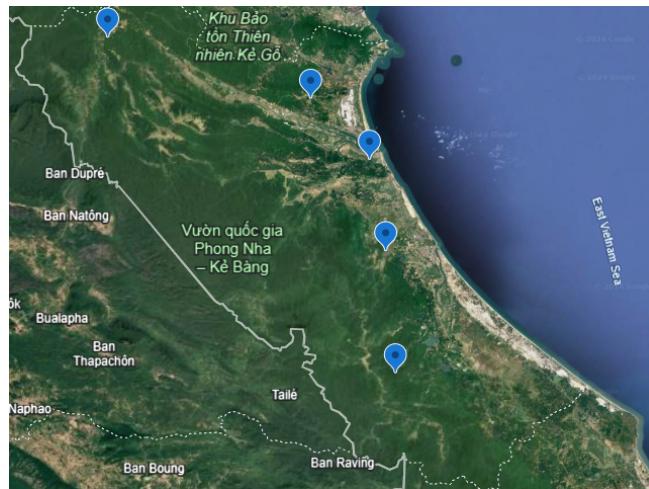
Có thể thấy, đây là một hệ thống mã nguồn mở hỗ trợ rất tốt trong việc thu thập dữ liệu lịch sử khu vực nghiên cứu. Trong khuôn khổ sản phẩm khoá luận tốt nghiệp, nhóm sử dụng dịch vụ Open Meteo nhằm thu thập dữ liệu khí tượng thuỷ văn (bao gồm lượng mưa) từ mô hình IFS-ECMWF trong quá khứ, cụ thể là từ năm 1990 đến nay. Dữ liệu chuỗi thời gian được thu thập theo giờ, góp phần quan trọng trong việc huấn luyện và xây dựng mô hình dự báo lượng mưa ở các tỉnh Bắc Trung Bộ.

Với tính chất của mô hình nội suy IFS-ECMWF với độ phân giải là 9km - nghĩa là các điểm dữ liệu trên bản đồ sẽ cách nhau 9km. Đây là một khoảng cách gần và hỗ trợ việc thu thập dữ liệu tại một địa điểm bất kỳ trở nên chính xác và thu thập được nhiều pha thời tiết do thời gian thu thập dữ liệu quá khứ có thể lên đến hàng chục năm.

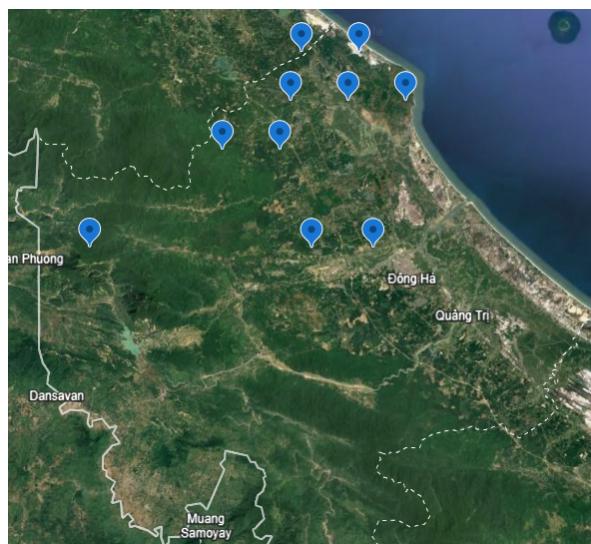
Nhóm sẽ so sánh, thu thập mỗi tỉnh 3 điểm dữ liệu từ nhiều điểm để có thể ứng dụng những yếu tố khí tượng thuỷ văn theo chuỗi thời gian không chỉ cho việc dự báo lượng mưa theo dữ liệu lượng mưa IFS mà còn cho lượng mưa Vrain. Hình 3.6, thể hiện một vài điểm được thu thập theo mô hình nội suy IFS tại địa bàn tỉnh Hà Tĩnh. Tương tự là hình 3.7 và 3.8 thể hiện các điểm nút lưới tương ứng tại 2 tỉnh thành Quảng Bình và Quảng Trị.



Hình 3.6: Điểm dữ liệu theo mô hình IFS ở tỉnh Hà Tĩnh



Hình 3.7: Điểm dữ liệu theo mô hình IFS ở tỉnh Quảng Bình



Hình 3.8: Điểm dữ liệu theo mô hình IFS ở tỉnh Quảng Trị

3.3.2 Hệ thống các trạm đo mưa VRAIN

Vrain là một hệ thống theo dõi lượng mưa trên toàn quốc của Việt Nam thuộc công ty Cổ phần Tư vấn và Phát triển kỹ thuật Tài nguyên nước (WATEC) là doanh nghiệp Khoa học & Công nghệ của thành phố Đà Nẵng, được thành lập vào năm 2003.

Hệ thống Vrain bao gồm các trạm đo, hệ thống máy chủ tiếp nhận và xử lý dữ liệu, trang web xem số liệu và quản trị. Khi trời mưa, cảm biến đo mưa sẽ liên tục cập nhật lượng mưa và truyền thông tin về hệ thống. Tất cả thông tin về thời tiết, mưa và lượng mưa sẽ được cập nhật từng phút trên trang web dành cho máy tính và trên ứng

dụng dành cho điện thoại thông minh.

Trạm đo mưa và bộ thu thập, truyền dữ liệu sử dụng điện lưới quốc gia vào những ngày mưa và điện mặt trời những ngày nắng. Ngoài ra, hệ thống lắp đặt pin dự phòng để duy trì hệ thống trong 40 ngày nếu mất điện tại khu vực. (theo [báo Nhân Dân](#)).

Theo thông tin về thiết bị cảm biến đo đặc lượng mưa của Vrain (theo trang [Vrain](#)) bao gồm các thông số kỹ thuật sau:

- Diện tích mặt hứng mưa của phễu: 314cm².
- Độ phân giải: 0,2mm.
- Cường độ mưa: từ 0-240mm/giờ.
- Sai số: ± 2
- Chất liệu: Inox 304, hợp kim kẽm sơn Epoxy, Composite pha hạt kim loại.
- Vỏ ngoài có thể tháo ra được, có lưỡi chắn rác

Quan sát hình 3.9, ta thấy hình dáng bên ngoài của phễu đựng mưa của hệ thống Vrain, đây là dụng cụ được đặt ngoài trời nhằm đo đặc lượng mưa hàng giờ một cách tự động trước khi được tổng hợp và gửi về hệ thống máy chủ.



Hình 3.9: Thiết bị cảm biến đo mưa của Vrain

Về bộ thu thập, xử lý và truyền dữ liệu thì có đặc điểm:

- Vi xử lý ARM Cortex - M3, CPU 32bit có tần số hoạt động 32MHz.
- Nguồn điện: Sử dụng năng lượng mặt trời hoặc điện lưới. Và nguồn điện dự phòng 3.7VDC/6.8Ah Lithium-ion hoạt động từ 45-60 ngày (khi mất điện).
- Tích hợp Modem truyền số liệu qua mạng GSM/GPRS/3G/4G. Và có phương thức truyền dữ liệu: 2G/3G/4G/SMS. Ngoài ra hệ thống cho phép cập nhật firmware từ xa thông qua GPRS/3G/4G.
- Tần suất ghi và truyền dữ liệu: 60 phút/lần và hỗ trợ thẻ nhớ lưu trữ dữ liệu: 32GB.
- Cho phép thực hiện cấu hình thông qua các phương thức: SMS/GPRS/3G/4G để thay đổi tần suất thu thập dữ liệu.

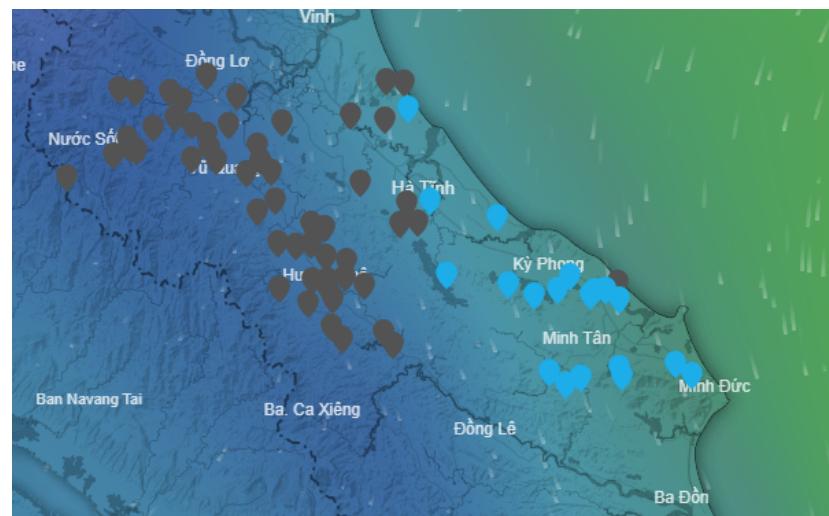
Hình 3.10 cho thấy bộ xử lý bên trong cảm biến đo mưa của hệ thống đo mưa Vrain có khả năng và chức năng như những điểm đã được liệt kê ở trên.



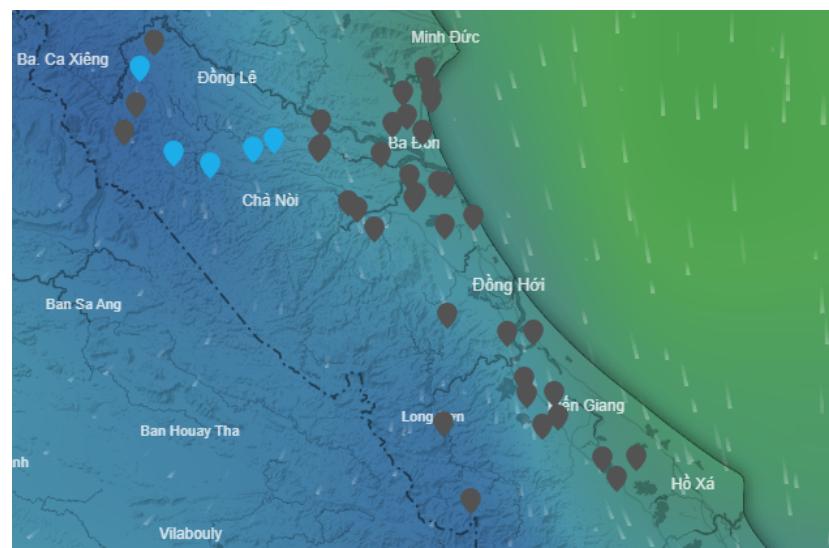
Hình 3.10: Bộ xử lý bên trong cảm biến đo mưa của Vrain

Một điểm đáng chú ý là hệ thống Vrain cung cấp giao diện người dùng trang web và ứng dụng để quan sát lượng mưa các tỉnh thành có đặt cảm biến. Hình 3.11, 3.12 và 3.13 thể hiện vị trí các trạm đo mưa khu vực tỉnh Hà Tĩnh vào ngày 21/12/2024. Quan sát các hình ảnh dưới đây, những vị trí hiển thị màu đen thể hiện những khu

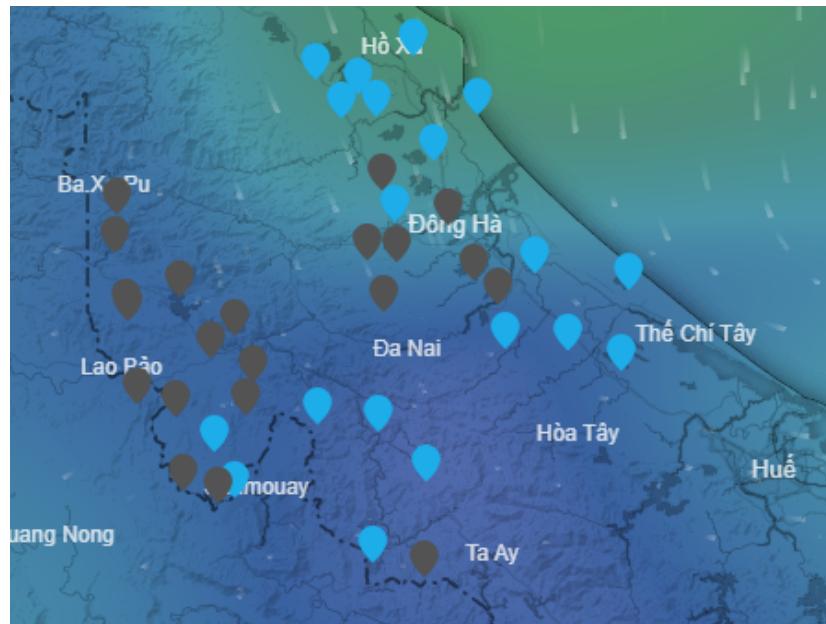
vực hiện tại không có mưa, và nếu có mưa sẽ hiển màu xanh dương - mưa nhỏ (lượng mưa trong 24 giờ nhỏ hơn 16 mm), hoặc màu xanh lá đối với mưa vừa (lượng mưa từ 16mm đến dưới 51mm), màu cam ứng với mưa to (lượng mưa trong 1 ngày từ 51mm đến 100mm) hay màu đỏ - mưa rất to (lượng mưa lớn hơn 100 mm).



Hình 3.11: Vị trí các trạm đo mưa Vrain tỉnh Hà Tĩnh



Hình 3.12: Vị trí các trạm đo mưa Vrain tại tỉnh Quảng Bình



Hình 3.13: Vị trí các trạm đo mưa Vrain tại tỉnh Hà Tĩnh

Trong khuôn khổ của khoá luận tốt nghiệp, nhóm ứng dụng dữ liệu lượng mưa lịch sử hệ thống đo mưa Vrain như nguồn tham khảo giúp người dùng có thể quan sát và so sánh lượng mưa dự đoán theo hệ thống Vrain và dữ liệu từ mô hình nội suy IFS-ECMWF.

3.3.3 Bộ dữ liệu ứng dụng cho mô hình

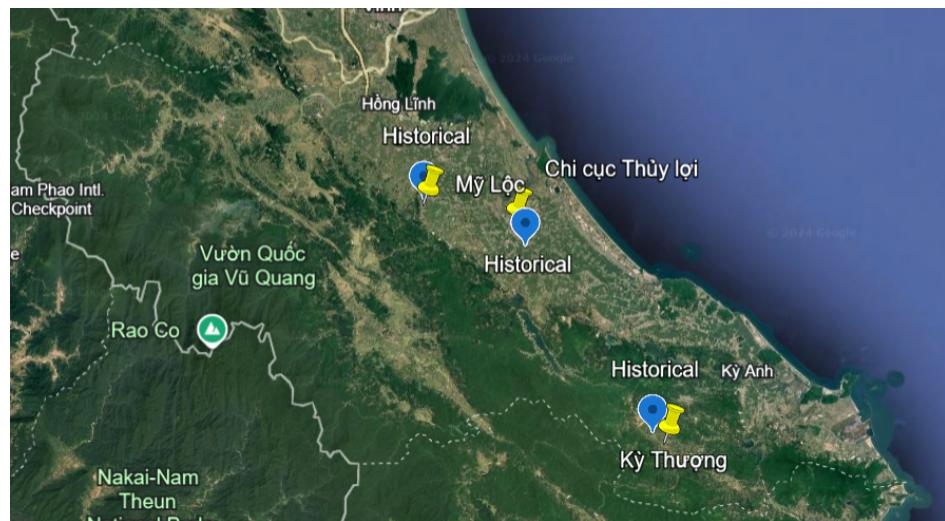
Thông qua hai phần trên, ta đã biết về dữ liệu từ mô hình nội suy IFS-ECMWF và dữ liệu lượng mưa theo Vrain. Nhóm sẽ tận dụng cả 2 nguồn dữ liệu này để đưa ra bộ dữ liệu ứng dụng cho việc dự đoán lượng mưa.

Nhắc lại về mô hình IFS-ECMWF, đây là một mô hình nội suy có độ phân giải 9km và thời gian lịch sử dữ liệu rất dài (sớm nhất từ năm 1980). Còn đối với Vrain, đây là hệ thống các cảm biến đo đặc lượng mưa được phân bố rộng khắp cả nước, tuy nhiên dữ liệu lịch sử thì khá ngắn (sớm nhất từ năm 2021).

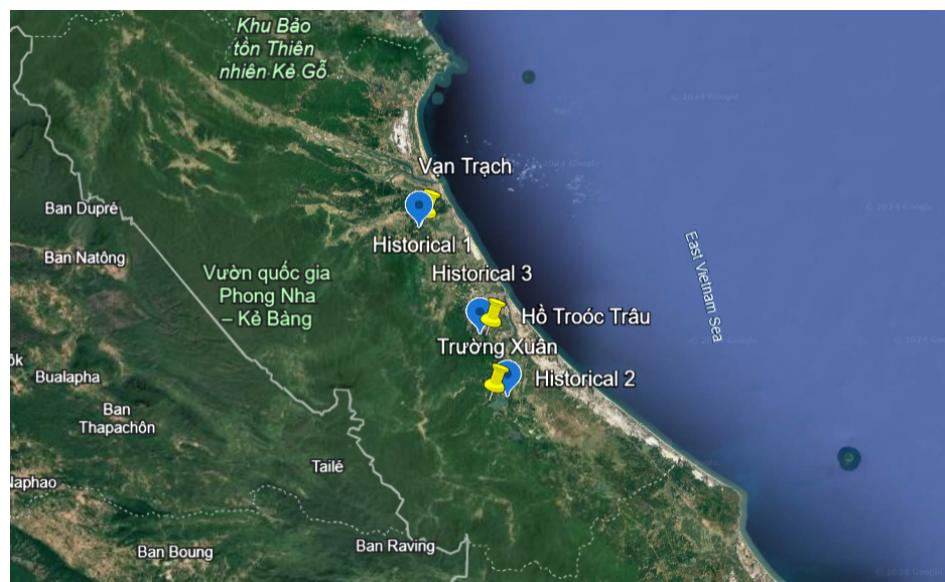
Vì vậy, nhóm sẽ xây dựng bộ dữ liệu với các yếu tố khí tượng thuỷ văn ảnh hưởng đến lượng mưa như nhiệt độ, độ ẩm, áp suất không khí, tốc độ gió, ... từ dữ liệu lịch mô hình tái phân tích IFS. Và giá trị đích sẽ là lượng mưa theo Vrain. Như vậy bộ dữ liệu thu thập được sẽ có độ dài 4 năm và được chia theo giờ.

Qua quá trình phân tích và sàng lọc, ta lựa chọn các cặp điểm sau đây để xây dựng

bộ dữ liệu. Có 3 tỉnh thành: Hà Tĩnh, Quảng Bình, Quảng Trị và mỗi tỉnh thành có 3 địa điểm dùng để xây dựng bộ dữ liệu. Như vậy mỗi tỉnh sẽ có 3 cặp điểm so sánh. Hình 3.14, hình 3.15 và hình 3.16 thể hiện 3 cặp điểm ở mỗi tỉnh nghiên cứu.



Hình 3.14: Cặp điểm dữ liệu tỉnh Hà Tĩnh



Hình 3.15: Cặp điểm dữ liệu tỉnh Quảng Bình



Hình 3.16: Cặp điểm dữ liệu tỉnh Quảng Trị

Các cặp dữ liệu được lựa chọn ở 3 hình 3.14, 3.15 và 3.16 đều được lựa chọn dựa trên các tiêu chí: cả 2 điểm phải gần nhau, điểm dữ liệu của hệ thống đo mưa Vrain phải ưu tiên đo đạc từ năm 2021 và liên tục đến thời điểm hiện tại. Như vậy ta sẽ có một bộ dữ liệu lịch sử của 4 năm liên tục (năm 2021 đến hết năm 2024). Cụ thể về tọa độ và khoảng cách giữa các cặp điểm sẽ được thể hiện ở bảng 3.1.

Tỉnh	Tên trạm	Khoảng cách cặp điểm (km)
Hà Tĩnh	Kỳ Thượng	2.4
	Chi cục Thủy Lợi	4.4
	Mỹ Lộc	0.4
Quảng Bình	Vạn Trạch	0.8
	Trường Xuân	4.1
	Hồ Trooc Trâu	1.2
Quảng Trị	Cửa Tùng	2.1
	Hướng Linh	2.6
	Triệu Ái	1.7

Bảng 3.1: Bảng khoảng cách giữa các cặp điểm thu thập

Như đã nói ở trên, nhóm sẽ xây dựng 2 bộ dữ liệu ứng với 2 nguồn lượng mưa lịch. Hay nói cụ thể hơn, cả hai bộ dữ liệu đều có yếu tố đầu vào là các chỉ số khí tượng thuỷ văn trong quá khứ từ mô hình nội suy IFS-ECMWF, nhưng giá trị đích sẽ có 2 dạng: dữ liệu lượng mưa lịch sử từ mô hình IFS và dữ liệu lượng mưa trong quá khứ hệ thống đo mưa Vrain.

Bộ dữ liệu dựa trên mô hình nội suy IFS-ECMWF có khoảng thời gian từ năm 2010 đến năm 2024, và bộ dữ liệu với giá trị đích là lượng mưa Vrain có độ dài 4 năm: từ năm 2021 đến hết năm 2024.

Chương 4. PHƯƠNG PHÁP NGHIÊN CỨU VÀ THỰC HIỆN

4.1 Tổng quan hệ thống

Toàn bộ hệ thống mô hình dự báo lượng mưa khu vực các tỉnh Bắc Trung Bộ được chia làm 3 phần: Thiết bị đầu cuối đóng vai trò thu thập dữ liệu khí tượng: nhiệt độ, độ ẩm, áp suất không khí và tương tác với hệ thống máy chủ. Máy chủ của hệ thống có vai trò nhận dữ liệu từ phần cứng, kết hợp với các yếu tố khác ảnh hưởng đến lượng mưa, đưa ra dự đoán về lượng mưa trong 24 giờ tiếp theo mà một tỉnh phải hứng chịu. Và cuối cùng là giao diện người dùng, giao diện web nhằm thể hiện cụ thể những lượng mưa dự đoán của khu vực nghiên cứu và màn hình ở thiết bị phần cứng thể hiện lượng mưa trung bình trong 1 giờ, 6 giờ và 24 giờ tiếp theo.

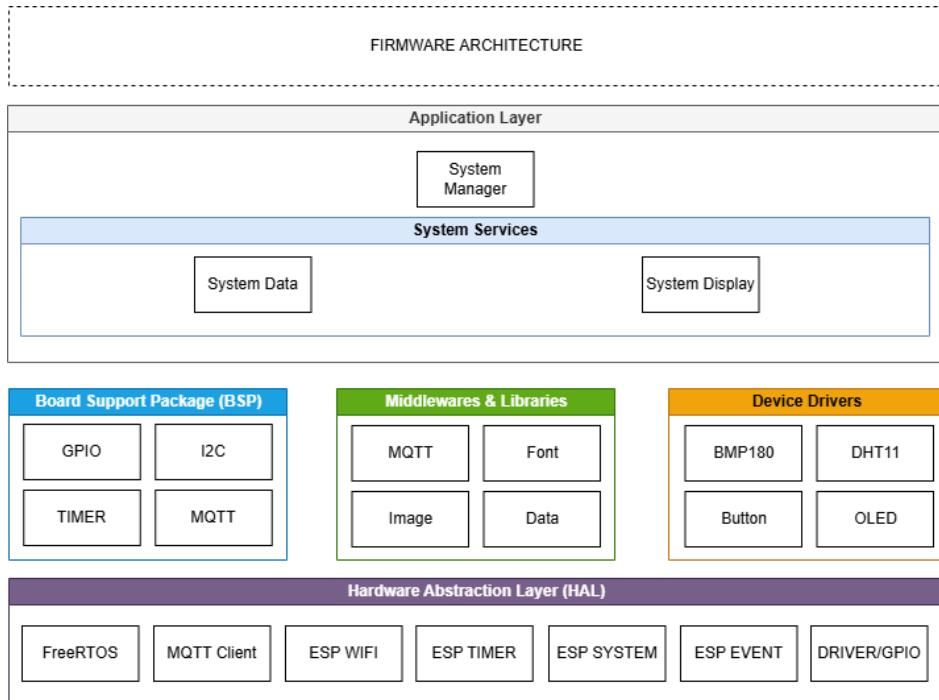
Như vậy, đầu vào dữ liệu của mô hình dự báo từ 2 nguồn dữ liệu: cảm biến đo đặc (phần cứng) và dữ liệu từ mô hình tái phân tích IFS-ECMWF. Đầu ra sẽ là lượng mưa dự báo từ 3 nguồn: mô hình nội suy IFS-ECMWF, mô hình dự đoán có kết hợp các yếu tố phần cứng (cảm biến) và lượng mưa theo cơ sở dữ liệu Vrain.

4.2 Quy trình hoạt động thiết bị đầu cuối

Phần này sẽ mô tả về kiến trúc phân lớp của thiết bị đầu cuối, giúp người đọc hiểu thêm về cách phát triển và hiện thực chức năng hoạt động của thiết bị và kịch bản hoạt động được thiết kế theo biểu đồ tuần tự nhằm giúp người dùng nắm rõ các bước xử lý theo trình tự thời gian của hệ thống.

4.2.1 Kiến trúc phân lớp thiết bị đầu cuối

Với những đặc điểm kỹ thuật về các thành phần trong hệ thống phần cứng đã được liệt kê ở chương 3, để có thể hiện thực được các chức năng của hệ thống một cách trực quan nhằm hỗ trợ tối đa công việc phát triển và gỡ lỗi. Nhóm sẽ chia cấu trúc hệ thống nhúng thành các lớp khác nhau với mỗi lớp đóng một vai trò riêng biệt, các lớp này được sắp xếp chồng lên nhau nhằm hiện thực chức năng phân lớp. Phương pháp thiết lập kiến trúc này thường được áp dụng trong các hệ thống nhúng, với vai trò giúp nhà phát triển có thể nâng cấp, bảo trì và phát triển sản phẩm một cách khoa học hơn.



Hình 4.1: Kiến trúc hệ thống

Quan sát hình 4.1, ta thấy kiến trúc hệ thống phần cứng đã được chia thành các lớp HAL (Hardware Abstraction Layer), lớp BSP (Board Support Package), lớp Driver, lớp Middleware, lớp System và lớp Application. Mỗi lớp đều có những thành phần riêng bên trong và đều đóng một vai trò quan trọng trong việc cụ thể hóa chức năng của toàn bộ hệ thống. Cụ thể:

- Lớp HAL: Cung cấp một giao diện chuẩn giúp trừu tượng hóa phần cứng, cho phép phần mềm ứng dụng giao tiếp với phần cứng bằng cách cung cấp các API để truy cập tài nguyên phần cứng như GPIO, UART, SPI, I2C, ADC, DAC,... Cụ thể trong nền tảng ESP-IDF, lớp HAL này bao gồm các thư viện về FreeRTOS, MQTT client, ESP WIFI,.. nhằm hiện thực các kết nối trên bo mạch ESP32.
- Lớp BSP: Cung cấp các trình điều khiển và cấu hình cần thiết để hỗ trợ một bo mạch cụ thể. BSP bao gồm các thiết lập cho các thiết bị ngoại vi, bộ khởi tạo hệ thống và cấu hình tài nguyên phần cứng. Trong hệ thống phần cứng này, lớp BSP có trách nhiệm thực hiện các chức năng liên quan đến chân GPIO, I2C, MQTT và định thời áp dụng cho bo mạch Goouuu-ESP32.

- Lớp Driver: Cung cấp các chức năng để điều khiển và giao tiếp với các thiết bị phần cứng như cảm biến, cụ thể ở đây là các thiết bị: nút nhấn, cảm biến đo độ ẩm (DHT11), cảm biến đo nhiệt độ và áp suất không khí (BMP180) và màn hình OLED (SH1106). Lớp Driver này đảm bảo rằng các thiết bị phần cứng hoạt động một cách hiệu quả và tối ưu, từ đó nâng cao hiệu suất toàn bộ hệ thống.
- Lớp Middleware: Đóng vai trò như một lớp trung gian: Cung cấp các dịch vụ và chức năng chung cho các ứng dụng, cũng như các dịch vụ quản lý dữ liệu như truyền thông qua mạng. Trong mô hình này, lớp Middleware bao gồm các xử lý trung gian liên quan đến giao tiếp giao thức MQTT giữa bo Goouuu-ESP32 và máy chủ, xử lý dữ liệu để hiển thị là truyền đi nơi khác, bên cạnh đó là 2 thư viện phông chữ và hình ảnh để hiển thị trên màn hình OLED.
- Lớp System: Có vai trò đảm bảo rằng tất cả các thành phần trong hệ thống nhúng hoạt động cùng nhau một cách hài hòa và hiệu quả. Đây là lớp điều khiển, hiện thực các chức năng của hệ thống thông qua truyền những câu lệnh ở các tầng thấp hơn (như Middleware, Driver, và BSP). Ở trong hệ thống phần cứng, lớp System có vai trò điều khiển và hiện thực các chức năng lưu trữ, xử lý và hiển thị dữ liệu.
- Lớp Application: là tầng trên cùng và là nơi triển khai logic ứng dụng cụ thể mà hệ thống sẽ thực hiện. Đây là nơi mà các yêu cầu từ người dùng hoặc hệ thống sẽ được xử lý bởi các dịch vụ từ các tầng thấp hơn. Cụ thể trong hệ thống phần cứng mà nhóm đang sử dụng, lớp này có chức năng khởi tạo, điều khiển các trạng thái hoạt động của hệ thống giúp hiện thực chức năng mong muốn.

4.2.2 Kịch bản hoạt động

Sau khi đã hoàn thiện các kiến trúc hệ thống phần cứng, tiếp theo ta sẽ đến phần hiện thực các chức năng thông qua các kịch bản, trạng thái hoạt động một cách tuần tự nhằm đảm bảo hệ thống hoạt động ổn định và không có sai sót.

Với kích thước nhỏ (1.3 inch) của màn hình oled SH1106, thì hệ thống sẽ tập trung hiển thị 2 loại màn hình: một màn hình thể hiện số liệu dữ liệu khí tượng thuỷ

văn đo đạc được mỗi chu kỳ: nhiệt độ, độ ẩm và áp suất không khí. Một màn hình còn lại thể hiện lượng mưa trung bình của 1 giờ, 6 giờ và 24 giờ tiếp theo từ các nguồn dự báo (IFS-ECMWF, IFS modified, VRAIN) ứng với mỗi tỉnh thành nghiên cứu trong khu vực các tỉnh Bắc Trung Bộ.

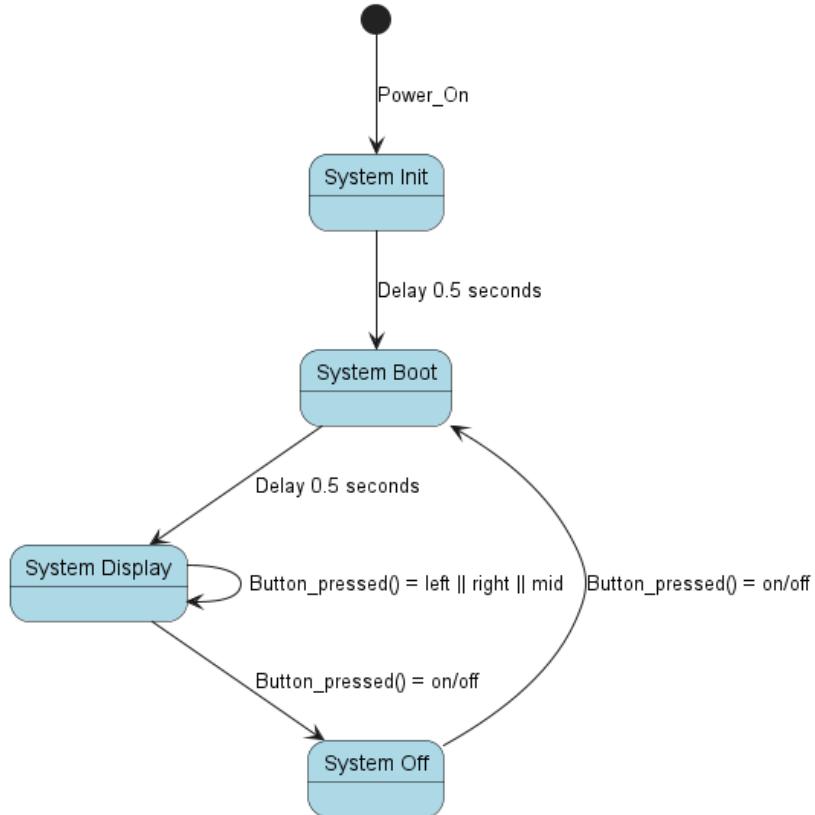
Như vậy, sau khi đã chia dữ liệu hiển thị thành 2 trạng thái khác nhau, ta có thể phân chia toàn bộ trình tự hoạt động của hệ thống phần cứng thành 5 trạng thái khác nhau:

- Trạng thái khởi tạo: Đây là trạng thái mở đầu khi người dùng cung cấp dòng điện cho hệ thống phần cứng. Trạng thái này sẽ hiển thị các giao thức và thiết bị cảm biến cần hoạt động.
- Trạng thái mở đầu: Đây là trạng thái hệ thống bắt đầu công việc của mình bằng cách xử lý những dữ liệu đầu tiên cần được hiển thị trên màn hình oled. Màn hình OLED lúc này sẽ hiển thị logo trường đại học UIT để báo hiệu hệ thống đang khởi động.
- Trạng thái hiển thị dữ liệu màn hình: Đây là trạng thái mà người dùng sẽ quan sát số liệu mà các cảm biến đo đạc được ở môi trường xung quanh hoặc màn hình OLED hiển thị những dữ liệu mưa dự báo với cường độ trung bình ứng với mỗi tỉnh.
- Trạng thái tắt màn hình: Đây là trạng thái mà màn hình OLED sẽ được tắt đi, tuy nhiên hệ thống vẫn sẽ hoạt động ngầm bởi dòng điện chưa bị ngắt. Nhằm khi người dùng mở lại màn hình OLED thì dữ liệu sẽ được cập nhật nhanh chóng.

Các trạng thái được chuyển tiếp với nhau thông qua 4 nhấn nút và được thể hiện thông qua sơ đồ trạng thái - 4.2 dưới đây:

- Nút tắt mở: Dùng để tắt mở màn hình OLED.
- Nút trái/phải: Dùng để thay đổi tình thành quan sát dữ liệu (Hà Tĩnh, Quảng Bình, Quảng Trị).

- Nút giữa: Dùng để thay đổi màn hình quan sát dữ liệu (dữ liệu cảm biến hay dữ liệu lượng mưa).

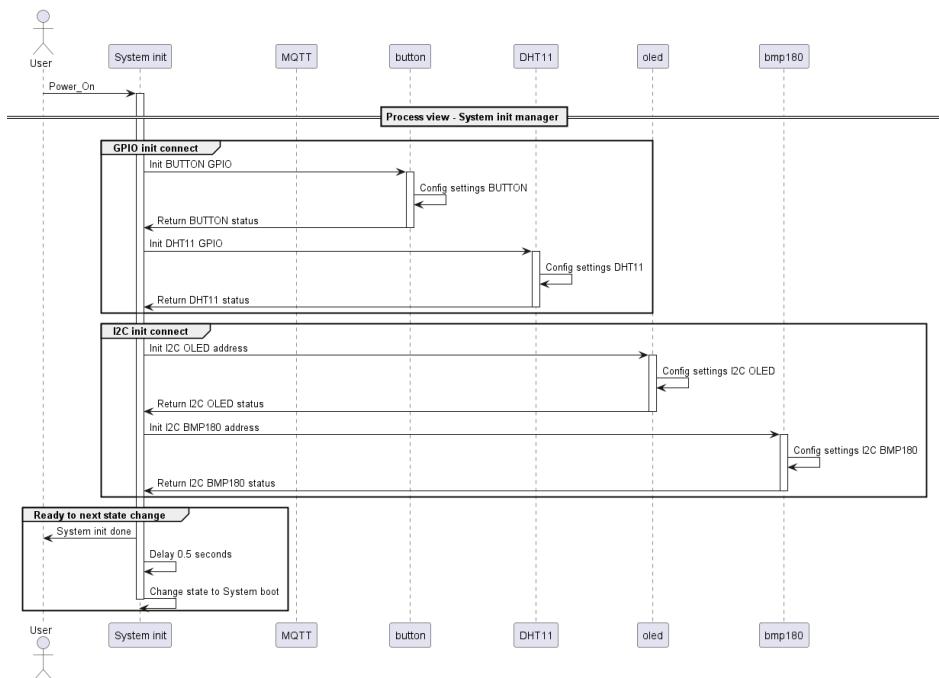


Hình 4.2: Sơ đồ chuyển trạng thái hoạt động

4.2.2.1 Trạng thái khởi tạo

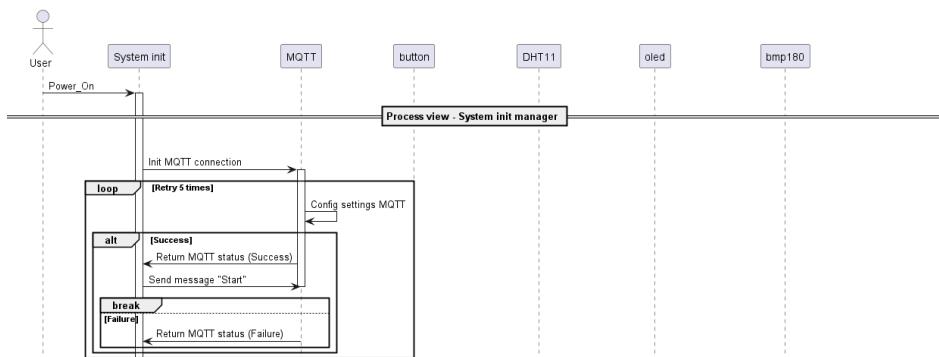
Trạng thái khởi tạo là trạng thái đầu tiên trong hệ thống khi được cấp nguồn điện. Ở trạng thái này, hệ thống sẽ khởi động các giao thức, thiết bị, cảm biến và kết nối với máy chủ.

Cụ thể, hệ thống sẽ khởi tạo các chân GPIO kết nối với 2 thiết bị là nút nhấn và cảm biến đo độ ẩm DHT11, sau khi đã chắc chắn cả 2 thiết bị đều hoạt động tốt. Tiếp theo hệ thống tiếp tục khởi tạo 2 thiết bị là màn hình OLED SH1106 và cảm biến đo nhiệt độ và áp suất không khí BMP180. Quan sát hình 4.3 để thấy rõ từng bước khởi tạo hệ thống với các thiết bị ngoại vi và giao thức kết nối với chúng.



Hình 4.3: Nửa trên biểu đồ tuần tự trạng thái khởi tạo

Sau khi đã khởi tạo các kết nối với thiết bị phần cứng, bước tiếp theo đó là hệ thống sẽ khởi tạo kết nối với máy chủ ở tầng mạng. Giao thức được sử dụng là giao thức MQTT, vì vậy hệ thống sẽ khởi tạo giao thức này và gửi tín hiệu lên máy chủ nhằm thông báo sẵn sàng kết nối. Số lần gửi yêu cầu tối đa là 5 lần. Sau khi đã kết nối thành công, hệ thống sẽ gửi tin nhắn "Start" lên máy chủ để thông báo khởi động đã kết nối, đồng thời sẵn sàng nhận dữ liệu phản hồi: thời gian hiện tại và lượng mưa dự báo cho 3 tỉnh thành. Hình 4.4 dưới đây đã thể hiện tuần tự các bước khởi tạo kết nối với máy chủ.



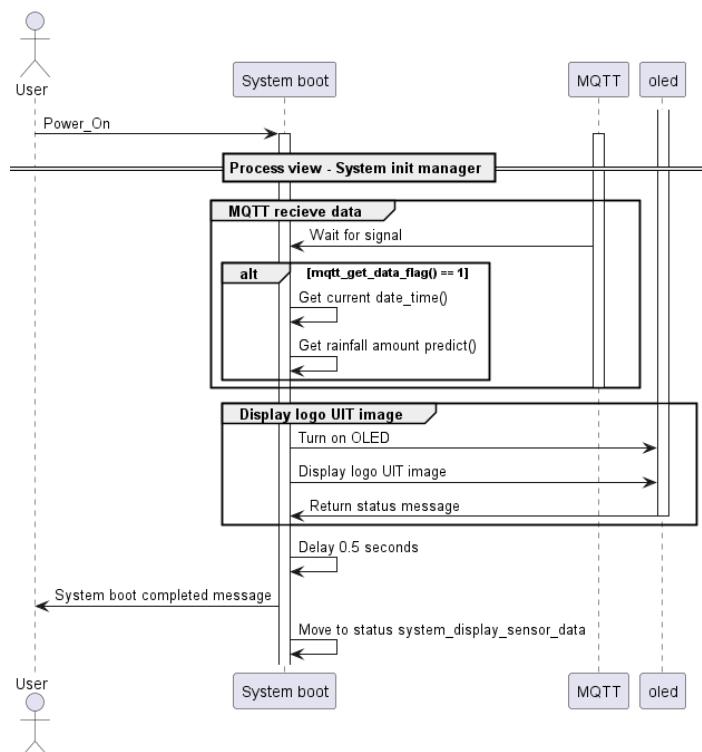
Hình 4.4: Nửa dưới biểu đồ tuần tự trạng thái khởi tạo

Sau khi đã khởi tạo hoàn tất, hệ thống sẽ chuyển đến trạng thái tiếp theo, trạng thái mở đầu.

4.2.2.2 Trạng thái mở đầu

Ở trạng thái mở đầu, hệ thống sẽ chờ đợi dữ liệu được cập nhật từ máy chủ. Khi có cờ tín hiệu được gửi đến, hệ thống lập tức xử lý tín hiệu và lưu vào bộ nhớ, sẵn sàng cho trạng thái tiếp theo - hiển thị màn hình. Cụ thể, dữ liệu được nói đến ở đây là thời gian và lượng mưa dự báo ở các tỉnh thành Bắc Trung Bộ đến từ máy chủ mô hình. Trước khi qua trạng thái kế tiếp, hệ thống sẽ hiển thị hình ảnh đại diện trường Đại học Công Nghệ thông tin trên màn hình OLED nhằm báo hiệu cho người dùng hệ thống đã sẵn sàng hiển thị dữ liệu.

Hình 4.5 dưới đây đã thể hiện các bước tuần tự mà hệ thống hoạt động khi bước vào trạng thái mở đầu.

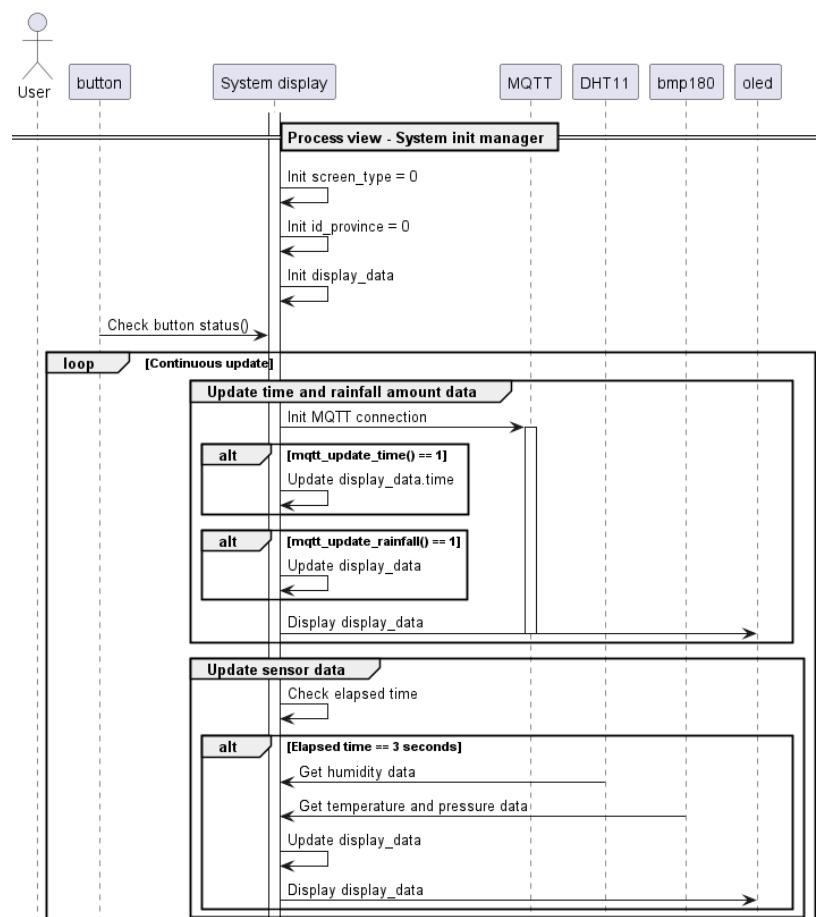


Hình 4.5: Biểu đồ tuần tự trạng thái mở đầu

4.2.2.3 Trạng thái hiển thị màn hình

Trạng thái hiển thị màn hình là trạng thái hoạt động chính của hệ thống. Đóng vai trò quan trọng như một giao diện hỗ trợ người dùng quan sát dữ liệu đo đạc và thu thập được.

Dữ liệu từ cảm biến sẽ được cập nhật mỗi 3 giây, còn dữ liệu từ phái máy chủ: thời gian hiện tại, lượng mưa dự báo sẽ được cập nhật liên tục theo giờ. Nhằm luôn có thể hiển thị lượng mưa dự báo trong 24 giờ tiếp theo. Quan sát hình 4.6 dưới đây để hiểu rõ quy trình cập nhật dữ liệu hệ thống.



Hình 4.6: Phần trên biểu đồ tuần tự trạng thái hiển thị dữ liệu

Như đã nói ở các phần trước, có 2 loại giao diện màn hình hiển thị: dữ liệu cảm biến và lượng mưa dự báo. Không chỉ vậy, còn có 3 tỉnh thành khác nhau với các dữ liệu khác nhau. Vì vậy để quản lý các giao diện này, hệ thống đã sử dụng 2 biến số quản lý: *screen_type* và *id_province* để hệ thống hoạt động hiệu quả.

Theo quy ước của hệ thống: *screen_type* có 2 giá trị:

- 0: Hiển thị dữ liệu cảm biến.
- 1: Hiển thị dữ liệu lượng mưa dự báo.

Còn đối với biến *id_province* thì có 3 giá trị:

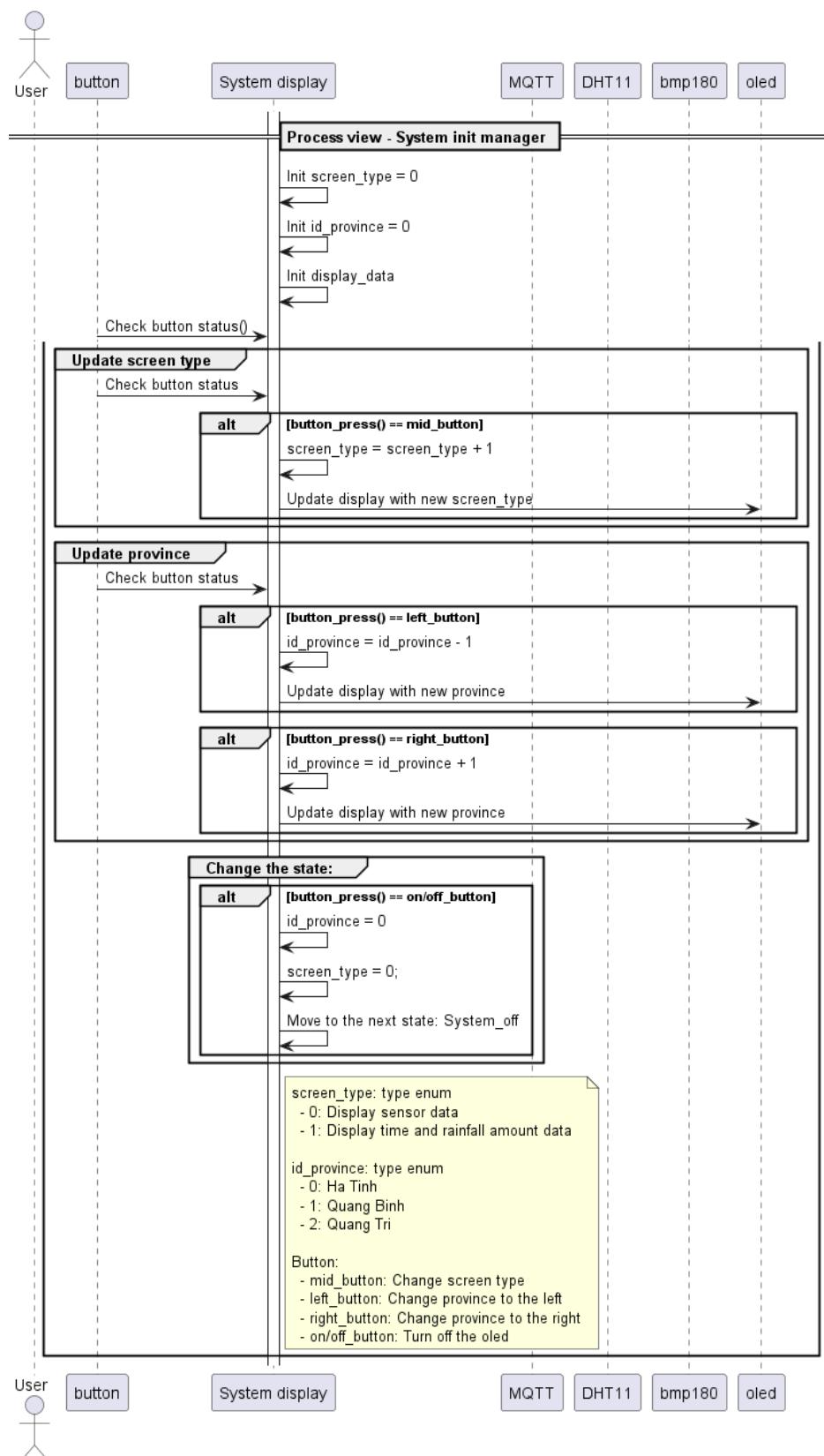
- 0: Tỉnh Hà Tĩnh.
- 1: Tỉnh Quảng Bình.
- 2: Tỉnh Quảng Trị.

Khi bước vào trạng thái, giá trị mặc định của 2 biến quản lý trên đều là 0. Nghĩa là giao diện đầu tiên hiển thị dữ liệu cảm biến của tỉnh Hà Tĩnh. Sau đó giao diện có thể thay đổi tùy thuộc vào người dùng khi nhấn nút.

Khi người dùng nhấn nút trái hoặc phải. Màn hình sẽ thay đổi tĩnh thành - đồng nghĩa thay đổi biến *id_province* nhưng vẫn giữ tiếp tục loại dữ liệu quan sát - giữ nguyên giá trị biến *screen_type*. Cụ thể, khi người dùng nhấn nút bên trái, giá trị *id_province* sẽ được dịch sang trái, điều này là bởi các tĩnh thành được xếp xoay vòng với nhau, vì vậy có thể chuyển đến tĩnh thành khác chỉ bằng một lần nhấn nút trái/phải. Ngược lại với nút trái, thì khi người dùng nhấn nút phải, màn hình giao diện hiển thị tĩnh thành nằm bên phải tĩnh trước đó (*id_province* tăng lên 1).

Khi người dùng nhấn nút giữa, giao diện màn hình sẽ thay đổi loại dữ liệu hiển thị (*screen_type* thay đổi) nhưng vẫn tĩnh thành đang quan sát (giá trị *id_province* không đổi). Hai giá trị của biến *id_province* luôn được thay đổi qua lại với nhau, vì vậy người dùng có thể thay đổi giao diện màn hình thông qua một lần nhấn nút giữa.

Khi người dùng muốn tắt màn hình oled, thì cần thao tác nhấn nút tắt/mở. Như vậy, hệ thống sẽ chuyển qua trạng thái tắt màn hình. Hình 4.7 thể hiện thứ tự thực hiện các bước hoạt động của hệ thống phần cứng.

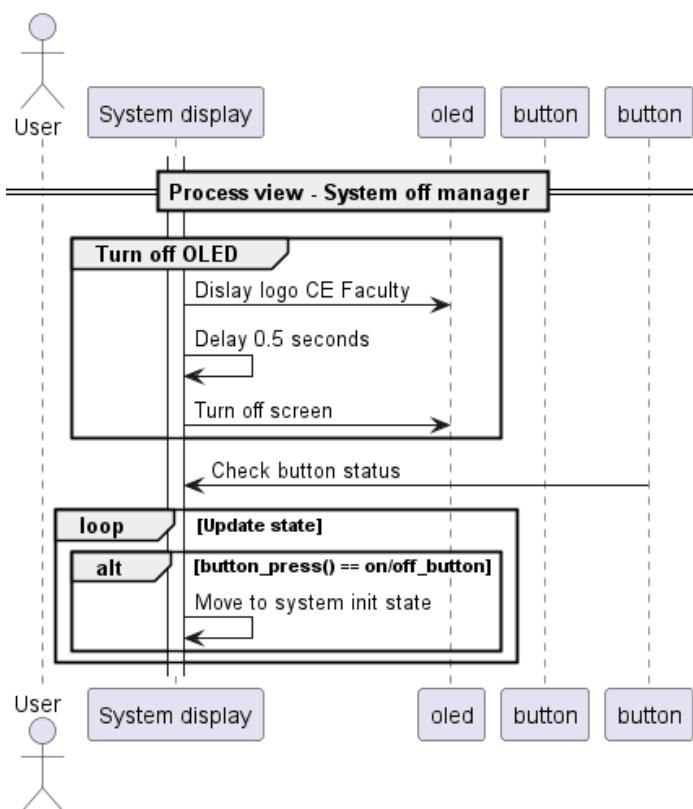


Hình 4.7: Nửa dưới biểu đồ tuần tự trạng thái hiển thị dữ liệu

4.2.2.4 Trạng thái tắt màn hình

Ở trạng thái này, hệ thống sẽ tắt màn hình OLED. Đầu tiên, màn hình OLED sẽ hiển thị hình nền đại diện khoa Kỹ thuật máy tính - Trường Đại học Công nghệ thông tin trong vòng 0.5 giây. Sau đó hệ thống sẽ tắt màn hình OLED.

Khi người dùng nhấn nút tắt/mở một lần nữa, trạng thái tiếp theo của hệ thống đó là trạng thái mở đầu, và hiện thực lại các công việc thuộc về trạng thái mở đầu đã được mô tả ở phía trên.



Hình 4.8: Biểu đồ tuần tự trạng thái tắt màn hình

Khi hệ thống bước vào trạng thái tắt màn hình, tuy màn hình oled sẽ không hiển thị nhưng các thành phần khác của hệ thống vẫn hoạt động bình thường, điều này giúp tiết kiệm thời gian khi hệ thống quay lại trạng thái khởi tạo với có sẵn kết nối MQTT (nếu có) và các giao thức giao tiếp với cảm biến đã được sử dụng từ trước đó.

4.3 Hệ thống máy chủ

Hệ thống máy chủ dự báo lượng mưa được xây dựng dựa trên nền tảng kiến trúc pipeline (ống dẫn), bao gồm bốn thành phần chính:

- Mô đun thu thập dữ liệu thời tiết: Dữ liệu thời tiết dạng thô được thu thập từ hệ thống mã nguồn mở Open-Meteo dựa trên các API cho sẵn, từ đó cung cấp các thông tin mà mô hình cần sử dụng như nhiệt độ, độ ẩm, tốc độ gió và các yếu tố khí tượng khác.
- Mô đun tiền xử lý dữ liệu: Dữ liệu thô sau khi thu thập sẽ được chuẩn hóa và loại bỏ nhiễu. Qua đó trích xuất các đặc trưng quan trọng nhằm phục vụ quá trình học máy, đảm bảo dữ liệu đầu vào có chất lượng tốt và phù hợp với mô hình dự báo.
- Mô hình dự báo lượng mưa: Dựa trên mạng nơ-ron LSTM để phân tích và dự đoán lượng mưa nhờ khả năng xử lý chuỗi thời gian một cách hiệu quả, cụ thể trong khuôn khổ của khoá luận là 24 giờ. Điều này cho phép mô hình không chỉ nhận diện các xu hướng, các pha mưa mà còn hiểu rõ mối quan hệ phức tạp giữa các yếu tố khí tượng trong dữ liệu thời tiết, từ đó đưa ra các dự đoán có độ chính xác cao.
- Hệ thống quản trị cơ sở dữ liệu: Dữ liệu khí tượng thuỷ văn trong quá khứ và lượng mưa dự đoán trong tương lai được lưu trữ trong cơ sở dữ liệu, đảm bảo khả năng truy xuất và quản lý dữ liệu hiệu quả. Vì vậy, hệ thống cung cấp một API trực quan nhằm để hiển thị dữ liệu và kết quả dự báo, cung cấp giao diện thân thiện với người dùng, hỗ trợ các bên liên quan trong việc khai thác thông tin.

4.3.1 Mô đun thu thập dữ liệu và trích chọn đặc trưng

Mô-đun thu thập dữ liệu dự báo lượng mưa được xây dựng nhằm đảm bảo khả năng truy xuất và tích hợp dữ liệu thời tiết từ nguồn mã nguồn mở Open-Meteo, cung cấp thông tin trong lịch sử 20 năm với độ chi tiết cao. Dữ liệu bao gồm 26 đặc trưng được thu thập tại 9 địa điểm, phân bố đều cho 3 tỉnh thành nghiên cứu bao gồm:

- Tỉnh Hà Tĩnh: Kỳ Thượng, Chi cục Thủy Lợi, Mỹ Lộc.
- Tỉnh Quảng Bình: Vạn Trạch, Trường Xuân, Hồ Troóc Trâu.
- Tỉnh Quảng Trị: Cửa Tùng, Hướng Linh, Triệu Ái.

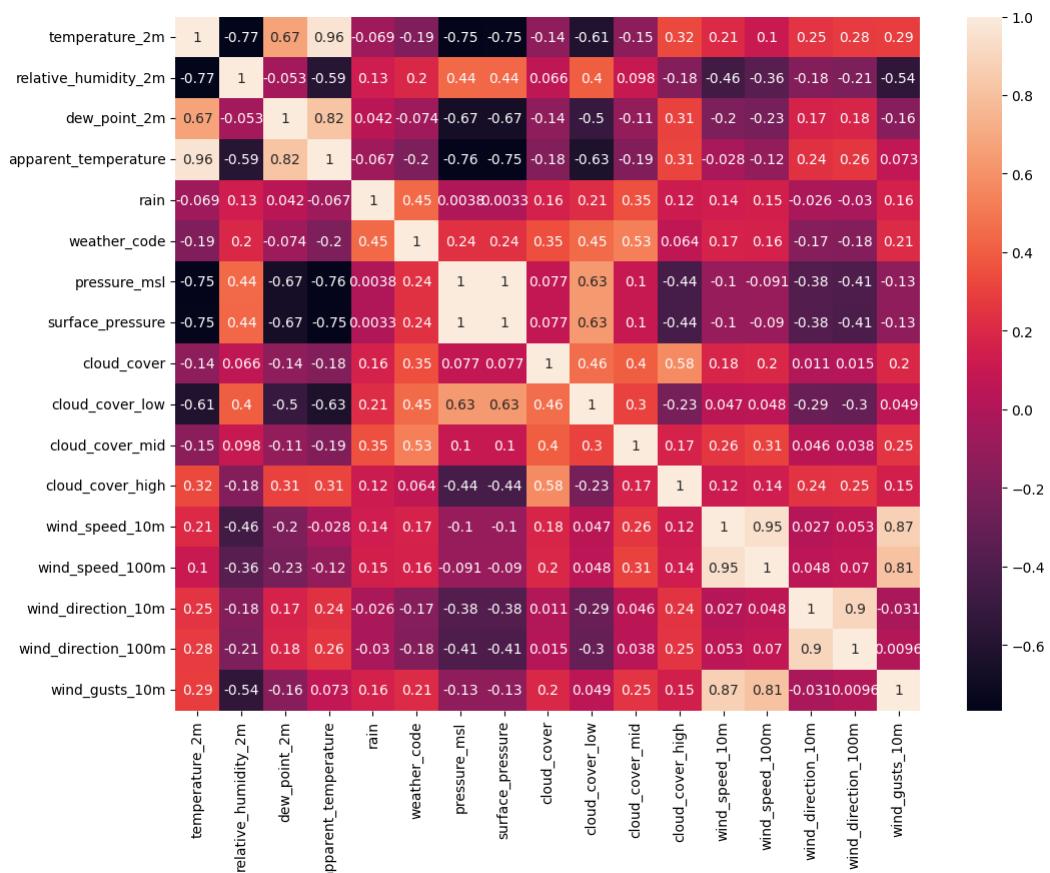
Hệ thống được thiết kế để tích hợp API của Open-Meteo, áp dụng một loạt các cơ chế tiên tiến nhằm đảm bảo tính hiệu quả, ổn định và chính xác trong việc thu thập dữ liệu thời tiết, đồng thời tối ưu hóa các quy trình xử lý và giảm thiểu rủi ro khi thực hiện các yêu cầu đến máy chủ như:

- Quản lý phiên lưu trữ (Cache session management): Khi hệ thống gửi một yêu cầu API và nhận được dữ liệu phản hồi, dữ liệu này sẽ được lưu trữ trong bộ nhớ đệm (cache). Các yêu cầu truy vấn sau đó cho cùng một dữ liệu sẽ được trả về trực tiếp từ bộ nhớ đệm thay vì gửi yêu cầu mới đến API. Qua đó tiết kiệm thời gian xử lý và băng thông mạng và giảm thiểu áp lực lên máy chủ API, đảm bảo hiệu suất hoạt động ổn định hơn.
- Kỹ thuật tự động gọi lại các yêu cầu (request) sau một khoảng thời gian chờ tăng dần theo cấp số mũ: Nếu một yêu cầu gửi đến API bị lỗi (do kết nối mạng, quá tải máy chủ, hoặc các lý do khác), hệ thống sẽ chờ một khoảng thời gian trước khi thử gửi lại yêu cầu. Khoảng thời gian chờ này sẽ tăng dần theo cấp số mũ (ví dụ: 1 giây, 2 giây, 4 giây, 8 giây...). Lợi ích của cơ chế này không chỉ làm giảm nguy cơ quá tải máy chủ API do gửi yêu cầu liên tục mà còn tăng khả năng thành công của yêu cầu trong những điều kiện mạng không ổn định.
- Thực hiện giới hạn số lần gọi API trong một phút, vì API này không cho phép gọi liên tục. Mục đích là ngăn ngừa lỗi "Rate Limit Exceeded" từ máy chủ khi tuân thủ quy định về giới hạn tần suất gọi các API của hệ thống mã nguồn mở Open-Meteo và giúp đảm bảo hệ thống hoạt động ổn định và không bị gián đoạn trong quá trình thu thập dữ liệu.
- Xử lý các biến thời gian và chuỗi dữ liệu thời gian theo múi giờ UTC+7: Dữ liệu được thu thập dựa trên múi giờ UTC+7 (múi giờ của Việt Nam) và đặt kiểu

dữ liệu thời gian về dạng ISO8601 - định dạng thích hợp cho máy tính xử lý và phù hợp với các mô hình học máy. Những thiết lập này nhằm đảm bảo tính đồng nhất và đồng bộ cho các quy trình thu thập và xử lý dữ liệu trong quá trình dự báo lượng mưa.

Dữ liệu thu thập được lưu trữ trong các tệp CSV, tuy nhiên, những tệp này ban đầu không chứa thông tin vị trí địa điểm thu thập. Do đó, mỗi bảng dữ liệu ứng với mỗi trạm được bổ sung thêm cột thông tin về vị trí địa lý nhằm đảm bảo tính đồng nhất và hữu ích cho các bước xử lý tiếp theo.

Sau khi đã thu thập dữ liệu các yếu tố khí tượng ảnh hưởng đến lượng mưa. Để đánh giá mối quan hệ giữa các biến (các yếu tố trong một tập dữ liệu), nhóm đã lập ma trận tương quan như (Hình 4.9) để đo lường mức độ và hướng của mối quan hệ giữa hai hoặc nhiều biến.



Hình 4.9: Ma trận hệ số tương quan của dữ liệu tại Triệu Ái

Quan sát hình 4.9, ta thấy những yếu tố mà nhóm sử dụng bao gồm 16 yếu tố: Số liệu nhiệt độ (nhiệt độ 2m, nhiệt độ biểu kiến, điểm sương), độ ẩm và áp suất không khí, độ che phủ mây ở các độ cao khác nhau và các đặc tính gió (tốc độ và hướng ở các độ cao khác nhau). Mỗi ô trong ma trận hệ số thể hiện hệ số tương quan giữa hai biến - tương ứng với 2 yếu tố khí tượng. Hệ số tương quan này có giá trị từ -1 đến 1, với ý nghĩa:

- Giá trị gần 1: Hai biến có mối quan hệ dương rất mạnh, nghĩa là khi giá trị của biến này tăng thì giá trị của biến kia cũng có xu hướng tăng. Và mối quan hệ dương (tương quan thuận) được thể hiện bằng màu đỏ, với việc màu đỏ càng đậm, mối tương quan thuận càng lớn.
- Giá trị gần -1: Hai biến có mối quan hệ âm rất mạnh, nghĩa là khi giá trị của biến này tăng thì giá trị của biến kia có xu hướng giảm. Và mối quan hệ này được thể hiện bằng màu tím, với mối tương quan nghịch càng mạnh thì màu tím càng đậm.
- Giá trị gần 0: Thể hiện rằng hai biến hầu như không có mối quan hệ tuyến tính với nhau. Và được thể hiện bằng màu đen.

4.3.2 Tiền xử lý dữ liệu

4.3.2.1 Biến đổi lượng mưa

Dữ liệu lượng mưa được biến đổi logarit để xử lý sự phân bố lệch phải của nó. Lý do cho việc này chính là hầu hết các giá trị lượng mưa đều nhỏ (thấp, gần bằng hoặc bằng 0), nhưng đôi khi lại có các giá trị lớn bất thường - khi diễn ra một trận mưa lớn

Vì thế dẫn đến việc đuôi của đồ thị biểu thị lượng mưa thường dài về phía bên phải (lệch phải), từ đó gây khó khăn cho các mô hình học máy và học sâu vì các mô hình này hoạt động tốt hơn trên dữ liệu có phân phối gần chuẩn. Biểu thức biến đổi theo logarit được thực hiện thông qua biểu thức sau:

$$y = \ln(x + 1) \quad (4.1)$$

Với x là giá trị ban đầu, và sau khi mạng nơ ron LSTM đã tính toán ra được lượng mưa dự báo, để chuyển đổi lượng này về giá trị có thể quan sát được, ta thực hiện phép nghịch đảo:

$$x = e^y - 1 \quad (4.2)$$

Phép biến đổi này giúp bình thường hóa việc phân phối các giá trị lượng mưa và ổn định phương sai, khiến giá trị lượng mưa phù hợp hơn cho việc huấn luyện mạng nơ ron thông qua việc giúp các mạng nơ-ron xử lý hiệu quả hơn, vì các trọng số và độ dốc không bị ảnh hưởng mạnh bởi các giá trị lượng mưa lớn bất thường như ban đầu.

4.3.2.2 Chuẩn hóa đặc trưng

Mỗi đặc trưng thời tiết được chuẩn hóa bằng cách sử dụng chuẩn hóa theo tỷ lệ Min-Max. Đây là một kỹ thuật phổ biến trong quá trình tiền xử lý dữ liệu, phương pháp này đảm bảo rằng các đặc trưng thời tiết được đưa về một phạm vi cố định, giúp cải thiện hiệu quả huấn luyện mô hình. Dưới đây là công thức chuẩn hóa Min-Max:

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (4.3)$$

Trong đó: x là giá trị gốc của đặc trưng (ví dụ nhiệt độ, độ ẩm, áp suất không khí, ...), x_{min} là giá trị nhỏ nhất của đặc trưng trong tập dữ liệu, x_{max} là giá trị lớn nhất của đặc trưng trong tập dữ liệu và x_{scaled} là giá trị đặc trưng sau khi chuẩn hóa, nằm trong khoảng $[0, 1]$.

Điều này đảm bảo tất cả các đặc trưng đều nằm trong phạm vi $[0,1]$, nhằm ngăn chặn bất kỳ đặc trưng đơn lẻ nào chiếm ưu thế trong quá trình đào tạo mô hình (tránh việc tồn tại biến độc lập). Do trong bài toán dự báo thời tiết, các đặc trưng đầu vào có thể có các đơn vị đo lường và phạm vi giá trị rất khác nhau nên nếu không thực hiện phương pháp chuẩn hóa, có thể nảy sinh vấn đề rằng có các đặc trưng có phạm vi giá trị lớn sẽ có ảnh hưởng lớn hơn trong quá trình huấn luyện mô hình, ngược lại các đặc trưng khác có thể bị xem nhẹ. Bên cạnh đó là tăng độ chính xác của mô hình và giúp quá trình tối ưu hóa trở nên ổn định và hội tụ nhanh hơn.

4.3.2.3 Tạo các chuỗi dữ liệu - sequence

Việc triển khai sử dụng phương pháp cửa sổ trượt (sliding window) để tạo chuỗi đầu vào-đầu ra, với kích thước cửa sổ đầu vào w_{in} và w_{out} . Đây là một kỹ thuật phổ biến trong các bài báo dự báo chuỗi thời gian (time series forecasting). Kỹ thuật này được sử dụng để tạo ra các cặp dữ liệu đầu vào và đầu ra từ dữ liệu gốc, cho phép mô hình học được mối quan hệ giữa các quan sát trong quá khứ và dự đoán cho tương lai. Dưới đây là công thức chi tiết giải thích phương pháp này:

$$X[t] = [X[t-w_{in}+1], \dots, X[t]] \quad (4.4)$$

$$Y[t] = [Y[t+1], \dots, Y[t+w_{out}]] \quad (4.5)$$

Trong đó $X[t]$ đại diện cho vectơ đặc trưng của các đại lượng đầu vào và $Y[t]$ đại diện cho giá trị lượng mưa - giá trị biến mục tiêu. Và w_{out} là kích thước cửa sổ đầu ra (số bước thời gian dự đoán trong tương lai).

4.3.3 Mô-đun dự báo lượng mưa

4.3.3.1 Kiến trúc mô hình LSTM

Mô hình có cấu trúc với các lớp sau:

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 24, 128)	74,752
BatchNormalization	(None, 24, 128)	512
dropout (Dropout)	(None, 24, 128)	0
lstm_1 (LSTM)	(None, 64)	49,408
BatchNormalization 1	(None, 64)	256
dropout_1 (Dropout)	(None, 64)	0
dense (Dense)	(None, 64)	4,160
BatchNormalization 2	(None, 64)	256
dense_1 (Dense)	(None, 32)	2,080
BatchNormalization 3	(None, 32)	128
dense_2 (Dense)	(None, 24)	792
Total params:		132,344 (516.97 KB)
Trainable params:		131,768 (514.72 KB)
Non-trainable params:		576 (2.25 KB)

Bảng 4.1: Tóm tắt mô hình

- **Lớp đầu vào:** Shape: (24 bước thời gian, 17 đặc trưng)
- **Lớp LSTM thứ nhất:** Chứa 128 đơn vị ẩn với tỷ lệ bỏ học là 20%. Lớp này được cấu hình với *return_sequences = True*, nghĩa là nó xuất ra trạng thái ẩn ở mỗi bước thời gian, cho phép lớp LSTM thứ hai nhận toàn bộ chuỗi đầu ra.
- **Lớp LSTM thứ hai:** Chứa 64 đơn vị ẩn và được định cấu hình bằng giá trị *return_sequences = False*. Lớp này chỉ xuất trạng thái ẩn ở bước thời gian cuối cùng.
- **Các lớp kết nối đầy đủ:** Lớp thứ nhất chứa 64 đơn vị ẩn và lớp thứ hai chứa 32 đơn vị đều áp dụng hàm kích hoạt ReLU.
- **Lớp đầu ra:** Cũng là một lớp kết nối đầy đủ với 24 đơn vị ẩn (kích thước của đầu ra).

Áp dụng chuẩn hóa theo Batch tại mỗi lớp để giảm thiểu hiện tượng phân phối dữ liệu thay đổi, qua đó cải thiện hiệu suất và tốc độ huấn luyện của mô hình. Hàm kích hoạt ReLU được áp dụng ở lớp cuối để đảm bảo giá trị đầu ra không âm, giúp cải thiện độ chính xác dự đoán lượng mưa.

Cấu hình cho quá trình huấn luyện:

- **Hàm mất mát:** Trung bình sai số toàn phương (MSE). Như đã nói ở Chương 3. Chỉ số MSE nhằm đánh giá mức độ sai lệch giữa giá trị dự đoán và thực tế. Qua đó làm nổi bật các dự đoán có sai số lớn (do bình phương sai số).
- **Trình tối ưu hóa:** Adam cùng với thay đổi tốc độ học thích ứng. Các giá trị khởi tạo gồm: $\alpha = 0.001$, $\beta_1 = 0.9$ và $\beta_2 = 0.999$. Trình tối ưu hóa này giúp tự động điều chỉnh tốc độ học, giúp quá trình hội tụ nhanh và ổn định hơn.

Triển khai callback: Đây là công việc giúp kiểm soát quá trình huấn luyện để đảm bảo hiệu suất và tránh lãng phí tài nguyên.

- **Dừng lại sớm khi độ kiên nhẫn = 10.** Điều này đồng nghĩa với việc dừng quá trình huấn luyện sớm nếu mô hình không cải thiện sau một số epoch nhất định

(cụ thể là bằng 10). Qua đó tránh hiện tượng overfitting (mô hình học quá kỹ vào dữ liệu huấn luyện, mất khả năng tổng quát hóa). Cách hoạt động thông qua việc mô hình sẽ theo dõi giá trị hàm mất mát trên tập kiểm tra, nếu giá trị này không giảm trong 10 epoch liên tiếp, quá trình huấn luyện sẽ dừng.

- Giảm tốc độ học khi hàm mất mát đang cao. Cách hoạt động chính là nếu hàm mất mát không cải thiện trong một số epoch, mô hình sẽ giảm tốc độ học bằng một hệ số (ví dụ: giảm 50%). Mục đích của việc này giúp chính là giúp mô hình tiến gần hơn đến điểm hội tụ khi đang trong giai đoạn tối ưu cuối cùng.
- Lưu lại checkpoint trong quá trình huấn luyện. Đây là công việc lưu lại các tham số trọng số và cấu hình của mô hình có hàm mất mát nhỏ nhất. Nhằm mục đích lưu lại trạng thái của mô hình sau mỗi epoch, qua đó cho phép khôi phục mô hình từ điểm checkpoint tốt nhất nếu có sự cố hoặc mô hình bị dừng sớm.

Các kỹ thuật được áp dụng trong quá trình huấn luyện mô hình được mô tả ở trên không chỉ giúp cải thiện hiệu suất và độ chính xác trong quá trình huấn luyện mô hình mà còn giúp tối ưu tài nguyên tính toán. Các phương pháp, kỹ thuật trên không chỉ đảm bảo mô hình học tốt hơn mà còn hỗ trợ kiểm soát quá trình huấn luyện, giúp ngăn chặn overfitting và tiết kiệm thời gian.

4.3.3.2 Hệ thống dự báo theo thời gian thực

Hệ thống quản lý thời gian thực đóng vai trò: Thực thi các chức năng theo lịch trình, ứng dụng các API đầu cuối và xử lý lỗi và nhật ký hệ thống. Điều này đặc biệt quan trọng đối với những hệ thống yêu cầu tính chính xác và hiệu quả trong việc xử lý dữ liệu, dự đoán hoặc quản lý như mô hình dự báo lượng mưa nhóm đang nghiên cứu.

Đối với việc thực thi chức năng theo lịch trình, hệ thống sử dụng APScheduler cho các hoạt động tự động. Mục đích của việc này là lập lịch và quản lý các tác vụ tự động trong hệ thống Python. APScheduler cho phép định nghĩa các tác vụ cần thực hiện và thiết lập thời gian thực hiện các tác vụ đó. Các tác vụ được lên kế hoạch thực hiện ở dưới đây bao gồm:

- Cập nhật dữ liệu theo giờ: Hệ thống tự động cập nhật dữ liệu từ hệ thống mã nguồn mở Open-Meteo và dữ liệu từ hệ thống cảm biến (thiết bị biến) nhằm thu thập dữ liệu thời tiết đến thời điểm hiện tại, qua đó đảm bảo mô hình hoạt động với dữ liệu mới nhất.
- Tạo kết quả dự báo: Ứng dụng những dữ liệu đã cập nhật từ 2 nguồn trên để thực hiện công việc dự báo lượng mưa và lưu trữ lại kết quả dự báo. Sẵn sàng cho những công việc kế tiếp.
- Đồng bộ hóa cơ sở dữ liệu: Hệ thống tiến hành cập nhật và đồng bộ dữ liệu giữa các thành phần trong hệ thống để tránh xung đột hoặc lỗi dữ liệu. Dữ liệu này sau đó sẽ được gửi đến giao diện trang web và thiết bị đầu cuối để hỗ trợ người dùng quan sát lượng mưa dự báo.

Triển khai REST API để truy cập dữ liệu. Mục đích là để cung cấp một cách thức tiêu chuẩn cho phép các thành phần hệ thống truy xuất và thao tác dữ liệu:

- GET /api/dat đầu cuối: Đây là phương thức dùng để truy xuất giá trị lượng mưa dự đoán và giá trị thực tế.
- Hỗ trợ tham số truy vấn để người dùng truy xuất dữ liệu theo vị trí cụ thể, bởi hệ thống dự báo lượng mưa này hoạt động với 9 trạm nghiên cứu khác nhau, trải dài trên 3 tỉnh.

Việc thực thi chương trình bao gồm xử lý lỗi toàn diện. Công việc này giúp giảm thiểu rủi ro hệ thống gặp lỗi và không bị phát hiện, bên cạnh việc tăng độ tin cậy và khả năng bảo trì hệ thống.

- Bắt và ghi nhật ký ngoại lệ cho các yêu cầu API. Qua đó phát hiện và xử lý các lỗi phát sinh trong quá trình xử lý yêu cầu. Ví dụ như lỗi không thể tìm thấy dữ liệu cho một địa điểm hay định dạng của yêu cầu truy xuất không hợp lệ, ...
- Kiểm tra định dạng và cấu trúc dữ liệu. Giúp đảm bảo dữ liệu đầu vào và đầu ra của API đúng định dạng cho trước, qua đó ngăn chặn lỗi khi xử lý dữ liệu không hợp lệ.

- Giám sát và ghi lại nhật ký quy trình hệ thống. Mục đích là để theo dõi hoạt động của hệ thống từ đó phát hiện các vấn đề về hiệu suất hoạt động hoặc lỗi tiềm ẩn bên trong. Qua đó giúp các nhà phát triển hiểu rõ quy trình hoạt động của hệ thống, từ đó có thể cải tiến hệ thống trong tương lai.

4.3.4 Hệ thống quản trị cơ sở dữ liệu

Các hoạt động dưới đây của hệ thống quản trị cơ sở dữ liệu nhằm đảm bảo dữ liệu được tổ chức và xử lý một cách chính xác, cụ thể:

- Xử lý trùng lặp với độ phân giải dựa trên dấu thời gian (timestamp) - trường thông tin thường đi cùng với mỗi bản ghi dữ liệu được tạo ra hoặc cập nhật: Cách xử lý là giữ lại bản ghi mới nhất, thông qua việc so sánh các dấu thời gian giữa những dữ liệu trùng lặp, bản dữ liệu nào có dấu thời gian gần hiện tại hơn - mới hơn được giữ lại, các bản ghi cũ hơn bị loại bỏ. Công việc xử lý này đảm bảo dữ liệu được cập nhật là duy nhất, không trùng lặp, từ đó tránh dẫn đến việc làm tăng kích thước cơ sở dữ liệu, bảo đảm độ chính xác của mô hình khi áp dụng thực tế.
- Căn chỉnh (alignment) dữ liệu chuỗi thời gian: đây là quá trình tổ chức lại dữ liệu sao cho các điểm dữ liệu được căn chỉnh với nhau đúng theo yêu cầu: theo hằng giờ. Điều này rất quan trọng khi bộ dữ liệu của hệ thống dự báo lượng mưa bao gồm nhiều yếu tố khí tượng thuỷ văn khác nhau, và đến từ 2 nguồn: hệ thống cảm biến và mô hình nội suy IFS-ECMWF. Đảm bảo sự đồng nhất trong bộ dữ liệu và giảm thiểu sai sót trong quá trình hoạt động.
- Xử lý múi giờ hệ thống: Vì múi giờ của hệ thống mã nguồn mở Open-Meteo mặc định là bằng 0, khác múi giờ Việt Nam (UTC+7). Vì vậy hệ thống quản trị cơ sở dữ liệu phải được điều chỉnh về múi giờ của Việt Nam, ứng với vị trí địa lý và sự đồng nhất về chuỗi thời gian được thiết lập.

Chương 5. ĐÁNH GIÁ KẾT QUẢ THỰC NGHIỆM

5.1 Vận hành thiết bị đầu cuối

5.1.1 Đóng gói thiết bị

Thông qua những linh kiện điện tử, bo mạch đã được liệt kê ở trên, bây giờ nhóm sẽ thực hiện việc đóng gói sản phẩm để tăng tính thẩm mĩ và dễ dàng sử dụng. Để làm được công việc trên, nhóm đã lựa chọn vỏ bên ngoài là hộp nhựa kỹ thuật, từ hộp nhựa đó, nhóm sẽ cắt và khoan những lỗ xung quanh hộp nhằm đủ diện tích cho công tắc, nút nhấn và màn hình kết hợp với những dụng cụ như máy mài, máy khoan, keo dán,... Bên trong hộp chính là các cảm biến, nguồn điện,.. giúp cho thiết bị vận hành ổn định.

Hình 5.1 thể hiện hình dáng bên ngoài thiết bị phần cứng trong hệ thống dự báo lượng mưa. Có thể thấy, tại bề mặt trên của hộp đã được khoan và xử lý để xuất hiện vị trí cho màn hình OLED và nút nhấn thao tác. Chức năng các nút từ trái qua phải: Nút tắt/mở, nút giữa, nút trái và nút phải.



Hình 5.1: Bên ngoài thiết bị đầu cuối

Bên trong hộp chính là các sản phẩm còn lại của phần cứng thiết bị, bao gồm bo mạch điều khiển chính (Goouuu ESP32) và các cảm biến đo khí tượng và nguồn điện,... Quan sát hình 5.2 để quan sát các thành phần thiết bị được cố định và đóng gói bên trong hộp nhựa.



Hình 5.2: Bên trong thiết bị đầu cuối

5.1.2 Hiển thị dữ liệu

5.1.2.1 Người dùng nhấn nút tắt/mở

Khi người dùng nhấn nút tắt/mở, nếu thiết bị đang tắt sẽ được bật và sẵn sàng quan trắc dữ liệu khí tượng xung quanh, màn hình sẽ hiển thị hình ảnh logo Trường Đại học Công Nghệ thông tin - Đại học Quốc gia Thành phố Hồ Chí Minh. Trong quá trình này, thiết bị cũng sẽ khởi tạo kết nối Wifi và giao thức MQTT rồi sau đó gửi tín hiệu cập nhật dữ liệu lên hệ thống máy chủ.

Sau đó màn hình sẽ hiển thị giao diện thông số cảm biến khí tượng tại thành phố Hồ Chí Minh, như hình 5.3. Một điểm cần lưu ý là thời gian khởi tạo, mặc định của thiết bị là ngày "01/01/2025 lúc 00:00" nên nếu màn hình oled hiển thị thông tin thời gian trên có nghĩa rằng thiết bị chưa được kết nối Wifi và MQTT.



Hình 5.3: Màn hình hiển thị dữ liệu cảm biến

5.1.2.2 Người dùng nhấn nút giữa

Khi người dùng nhấn nút giữa, màn hình vẫn sẽ chuyển sang địa phận tỉnh Hà Tĩnh và giao diện hiển thị sẽ là dự báo lượng mưa, giống như hình 5.4.



Hình 5.4: Màn hình hiển thị dữ liệu lượng mưa dự báo tỉnh Hà Tĩnh

Vì kích thước không đủ của màn hình oled, nên mỗi giao diện của một tỉnh chỉ hiển thị dữ liệu lượng mưa trung bình cả 3 trạm nghiên cứu và lượng mưa hiển thị trên màn hình là lượng mưa trung bình của toàn tỉnh trong 1 giờ, 6 giờ và 24 giờ tiếp theo.

5.1.2.3 Người dùng nhấn nút trái/phải

Trường hợp người dùng nhấn nút trái phải, thì màn hình sẽ thay đổi sang giao diện dự báo lượng mưa tỉnh khác, quan sát hình 5.5 và 5.6 để thấy giao diện màn hình dữ liệu cảm biến và lượng mưa tỉnh Quảng Bình và Quảng Trị.



Hình 5.5: Giao diện dữ liệu lượng mưa dự báo tỉnh Quảng Bình



Hình 5.6: Giao diện dữ liệu lượng mưa dự báo tỉnh Quảng Trị

5.2 Giao tiếp với hệ thống máy chủ

Dữ liệu cảm biến sau khi được thu thập từ các cảm biến, bên cạnh việc cập nhật trên màn hình oled cũng sẽ được gửi về máy chủ thông qua giao thức MQTT và nền tảng trung gian để hiện thực công việc đó là flespi.

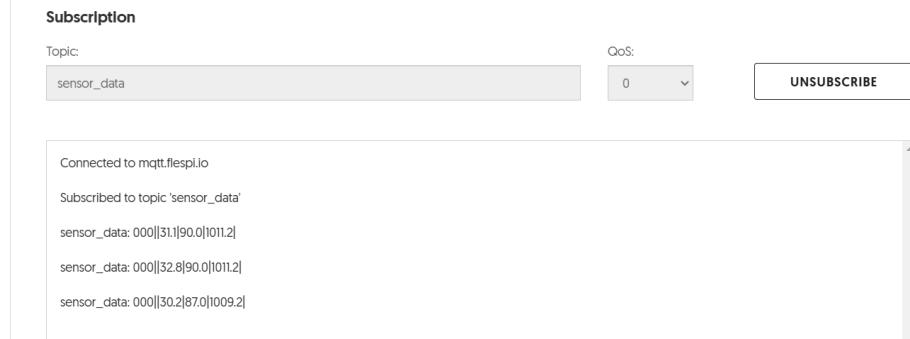
Hình 5.7 thể hiện những trường cần khai báo để hỗ trợ việc quan sát dữ liệu được luân chuyển.

Hình 5.7: Token và chủ đề theo dõi trên nền tảng flespi

Tin nhắn gửi dữ liệu lên nền tảng flespi là 3 giá trị, nhiệt độ, độ ẩm, áp suất không khí. Để báo hiệu cho phía máy chủ biết rằng đây là tin nhắn chứa giá trị cảm biến, phía đầu tin nhắn được thêm bằng chuỗi "000||".

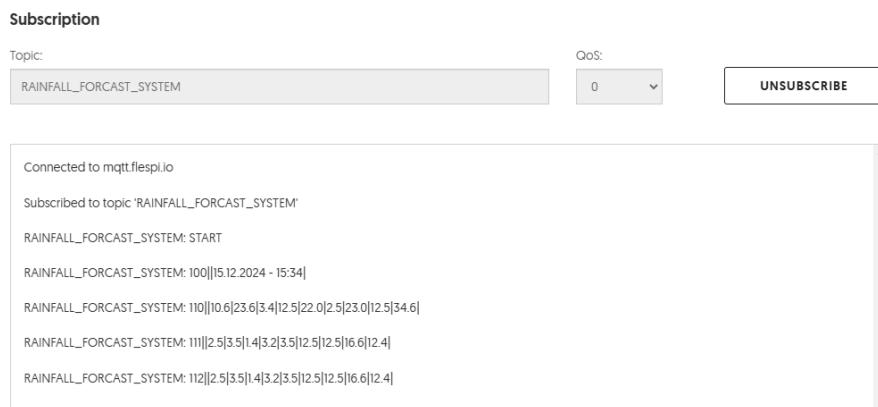
Như vậy, cứ mỗi 3 giây hệ thống máy chủ sẽ nhận được những giá trị cảm biến thông qua nền tảng flespi. Những giá trị này sẽ là 3 trong những đầu vào mô hình dự báo lượng mưa được đặt ở máy chủ.

Ví dụ với giá trị nhiệt độ, độ ẩm, áp suất không khí lần lượt là 31.1 độ C, 90.0% và 1011.2 hPa thì tin nhắn được gửi đi sẽ là: "000||31.1|90.0|1011.2|"



Hình 5.8: Dữ liệu cảm biến được gửi đến nền tảng flespi

Sau mỗi một phút, bên phía thiết bị đầu cuối sẽ gửi thông tin yêu cầu cập nhật lên hệ thống máy chủ thông qua nền tảng flespi này, hệ thống máy chủ sau khi nhận được yêu cầu sẽ lần lượt gửi 4 tin nhắn phản hồi bao gồm: Thời gian hiện tại - bắt đầu bằng "100||...", sau đó là lượng mưa dự báo ứng với tỉnh Hà Tĩnh, Quảng Bình, và Quảng Trị - lần lượt là "110||...", "111||..." và "112||....". Thiết bị đầu cuối sau khi nhận được thông tin cập nhật từ phía máy chủ sẽ thay đổi giao diện màn hình nhằm hiển thị số liệu lượng mưa mới nhất. Quan sát hình 5.9 để thấy quy trình cập nhật dữ liệu thông qua giao thức MQTT trên nền tảng Flespi giữa thiết bị đầu cuối và hệ thống máy chủ.

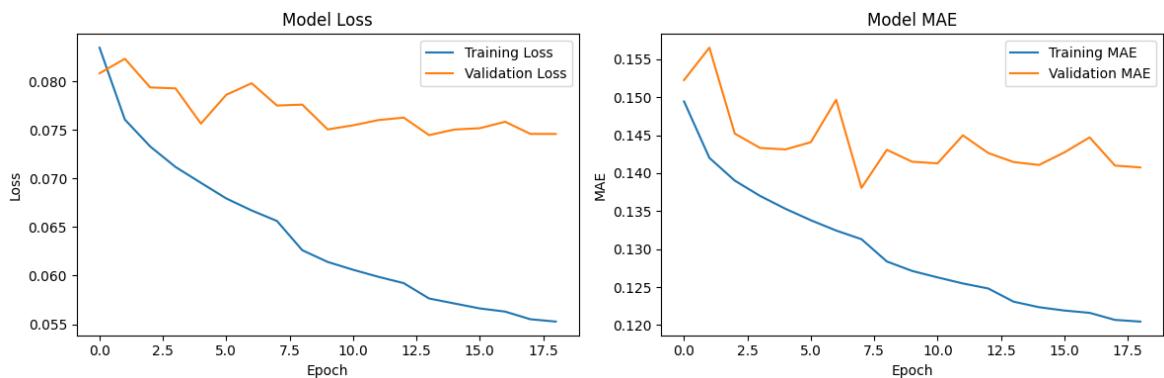


Hình 5.9: Dữ liệu cập nhật giữa hệ thống máy chủ và thiết bị đầu cuối

5.3 Đánh giá kết quả huấn luyện

5.3.1 Mô hình dự báo theo dữ liệu IFS

Kết quả trên Hình 5.10 cho thấy quá trình huấn luyện mô hình diễn ra rất thuận lợi. Khi trải qua càng nhiều epoch (số lần toàn bộ dữ liệu được đưa qua mô hình) thì cả hai đường cong đồ thị Model Loss và Validation Loss đều có xu hướng giảm, chứng minh mô hình đang dần học được và sai số đang giảm dần.



Hình 5.10: Đánh giá quá trình huấn luyện trên tập Validation

Dựa trên hình ảnh đồ thị nằm bên trái, ta thấy rằng:

- Model Loss: Biểu thị mức độ sai số giữa dự đoán của mô hình và giá trị thực tế trên tập huấn luyện. Đường cong này giảm dần theo thời gian, cho thấy mô hình đang dần học được và cải thiện về chính xác.
- Validation Loss: Biểu thị mức độ sai số giữa dự đoán của mô hình và giá trị thực tế trên tập kiểm định (validation). Đường cong này dao động nhưng có xu hướng giảm dần theo thời gian, cũng chỉ rằng mô hình có khả năng tổng quát hoá tốt hơn.

Đối với hình ảnh đồ thị bên phải - biểu đồ MAE:

- Training MAE: Ta thấy chỉ số MAE trên tập Test giảm dần trong suốt thời gian thể hiện kết quả tốt thông qua việc mô hình càng ngày giảm thiểu độ sai lệch khi dự đoán.

- Validation MAE: Chỉ số này dao động ở mức cao cho thấy việc mô hình có thể gặp khó khăn trong quá trình tổng quát hóa dữ liệu.

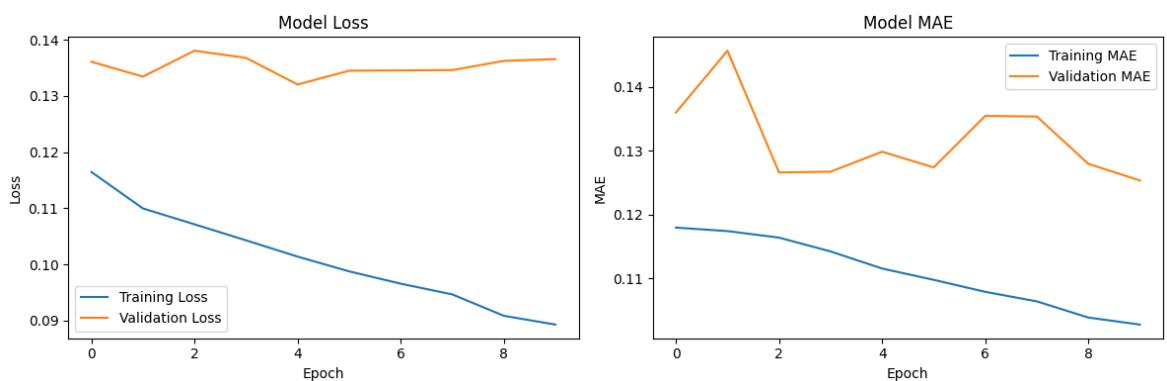
Các chỉ số tập TEST:

- MSE: 0.5959. Giá trị này thể hiện độ lệch bình phương giữa giá trị dự đoán và giá trị thực tế, giá trị này tương đối thấp chứng tỏ mô hình dự đoán khá tốt dẫu cho vẫn còn những sai sót.
- MAE: 0.3013. Chỉ số này đo lường sai số trung bình tuyệt đối giữa dự đoán và giá trị thực tế. Sai số này khá nhỏ, chứng tỏ rằng mô hình cho ra kết quả dự đoán sai lệch về lượng mưa ít.

Các kết quả cho thấy mô hình sẽ thiếu chính xác ở những khoảng thời gian có mưa nhiều, đặc biệt là vào mùa mưa khi xuất hiện những trận mưa có cường độ lớn bất thường, và kết quả dự đoán không thể chạm được ngưỡng cực trị so với thực tế.

5.3.2 Mô hình dự báo theo dữ liệu IFS kết hợp Vrain

Kết quả trên hình 5.11 cho thấy qua việc huấn luyện mô hình gặp bất lợi. Có thể thấy hàm mất mát của Validation rất biến động, lên xuống thất thường. Điều này là do dữ liệu dùng cho huấn luyện không được nhiều (khoảng 4 năm), dẫn đến mô hình chưa thể học được các quy luật của lượng mưa dài hạn.

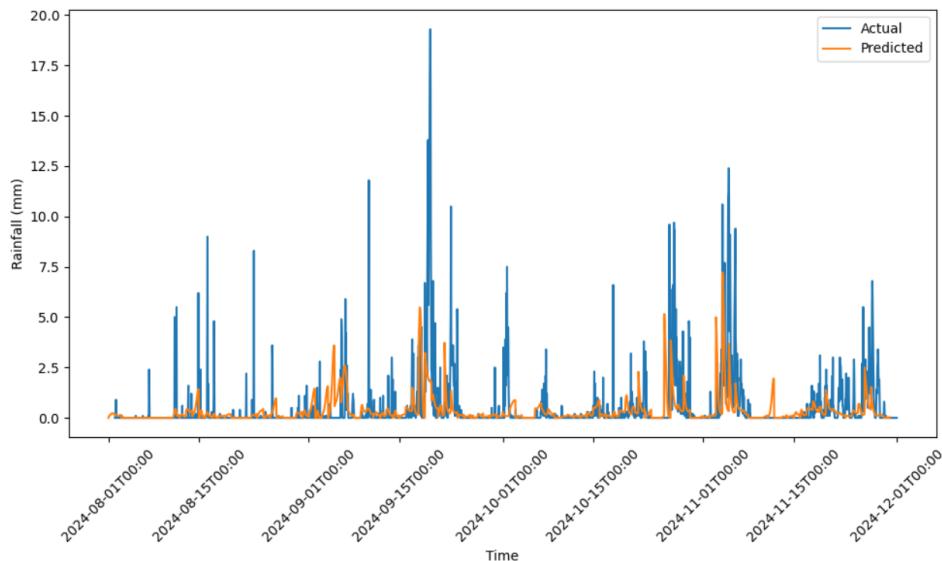


Hình 5.11: Đánh giá quá trình huấn luyện trên tập Validation

Tham số MAE trên tập TEST: 0.6639 cho thấy lượng mưa dự đoán có độ lệch 0.6639 (mm) - không đáng kể, vì lượng mưa 0.6639 mm là rất nhỏ.

5.3.3 Đánh giá mô hình trên tập Test bộ dữ liệu IFS

Hình 5.12 hiển thị biểu đồ so sánh lượng mưa dự đoán và lượng mưa thực tế theo bộ dữ liệu nội suy IFS-ECMWF trong khoảng thời gian từ ngày 01/08/2024 đến ngày 01/12/2024 tại trạm Chi Cục Thuỷ lợi - tỉnh Hà Tĩnh.



Hình 5.12: lượng mưa dự đoán và thực tế theo bộ dữ liệu IFS

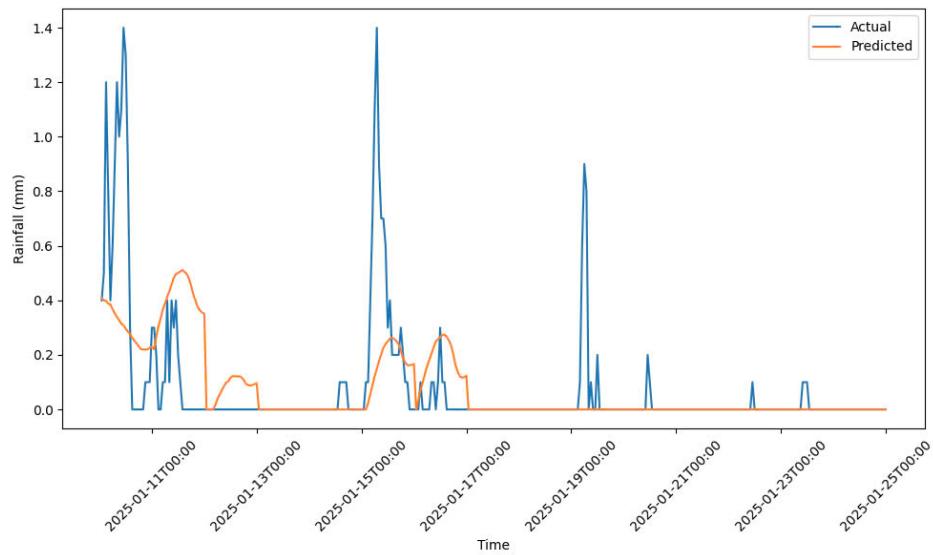
Biểu đồ cho thấy kết quả dự báo trong 4 tháng cho thấy lượng mưa trong thực tế có cường độ bất thường và liên tục có lượng mưa cao hơn mức bình thường do ảnh hưởng của những cơn bão hoặc áp thấp nhiệt đới,... vì vậy việc dự đoán lượng mưa trong khoảng thời gian này sẽ khó khăn hơn.

Quan sát biểu đồ cho thấy lượng mưa dự báo tuy chưa chạm được đỉnh của dữ liệu thực tế, nhưng đã có thể nắm bắt được xu hướng mưa của những trận mưa lớn bất thường. Dựa vào đó, ta có thể biết được trời có mưa hay không hoặc lượng mưa nằm ở mức nhỏ, vừa, to hoặc rất to.

Các thông số đánh giá bao gồm:

- RMSE: 1.445. Chỉ số này cho thấy lượng mưa dự đoán với sai số trung bình khoảng 1.445 mm lượng mưa.
- MAE: 0.561. Sai số này nằm trong mức chấp nhận được, nếu xét về yêu cầu dự đoán xu hướng hoặc giá trị gần đúng thay vì độ chính xác tuyệt đối.

Ngoài ra, nhóm cũng đưa ra biểu đồ so sánh lượng mưa dự đoán và thực tế trong khoảng thời gian 14 ngày, từ ngày 10/01/2025 - 24/01/2025 cũng tại trạm Chi Cục Thuỷ Lợi - tỉnh Hà Tĩnh. Có thể thấy, đây là khoảng thời gian mùa khô, nên đa số thời gian ghi nhận là không mưa, cá biệt chỉ ghi nhận những cơn mưa xuân, với cường độ không lớn - tối đa là 1.4 mm/h. Quan sát hình 5.13 để phân tích khả năng dự đoán của hệ thống.



Hình 5.13: lượng mưa dự đoán và thực tế trong tháng 1 năm 2025

Với các chỉ số đánh giá hiệu suất: RMSE: 0.21 và MAE: 0.089 - thấp hơn rất nhiều so với quãng thời gian nửa cuối năm 2024 theo hình 5.12: 1.445 và 0.561.

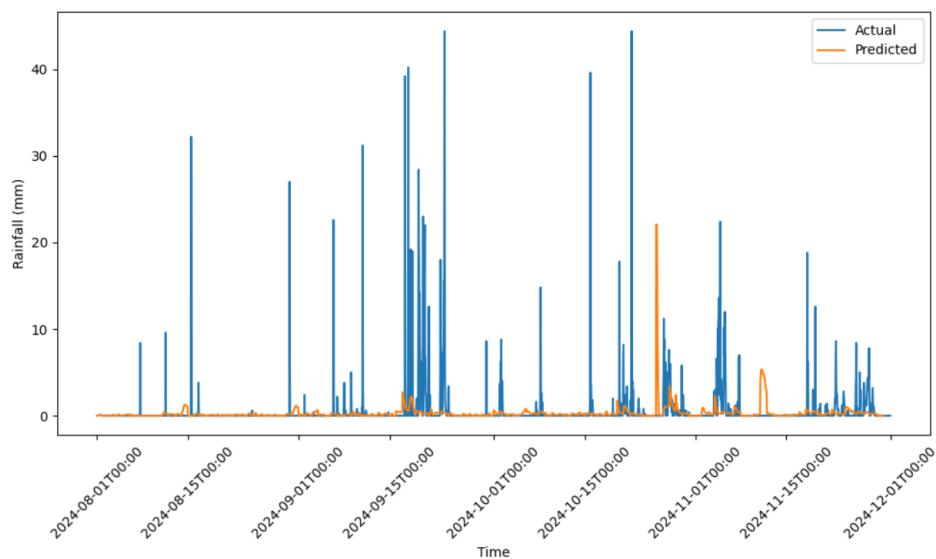
Về thống kê khả năng dự đoán có mưa hay không: Vì bộ dữ liệu IFS-ECMWF có độ chia nhỏ nhất - lượng mưa nhỏ nhất là 0.1 mm/h. Vì vậy với nếu lượng mưa dự đoán nằm dưới khoảng này (bé hơn 0.1 mm) thì được tính là không mưa. Với tổng cộng 360 mẫu ($24 * 15 = 360$ giờ), thì:

- Số giờ dự báo có mưa chính xác - TP: 53.
- Số giờ dự báo không mưa chính xác - TN: 245.
- Số giờ dự báo có mưa nhưng lại ghi nhận không mưa - FP: 42.
- Số giờ dự báo không mưa nhưng thực tế có mưa - FN: 20.

- Chỉ số Accuracy: 0.83.
- Chỉ số Precision: 0.56.
- Chỉ số Recall: 0.73.
- Chỉ số F1 Score: 0.63.

5.3.4 Đánh giá mô hình trên tập Test bộ dữ liệu kết hợp Vrain

Hình 5.14 thể hiện biểu đồ so sánh giữa lượng mưa dự báo và giá trị thực tế cũng tại trạm Chi Cục thuỷ lợi - Hà Tĩnh trong khoảng thời gian 4 tháng (01/08/2024 đến 01/12/2024).



Hình 5.14: Lượng mưa dự đoán và thực tế theo bộ dữ liệu IFS và Vrain

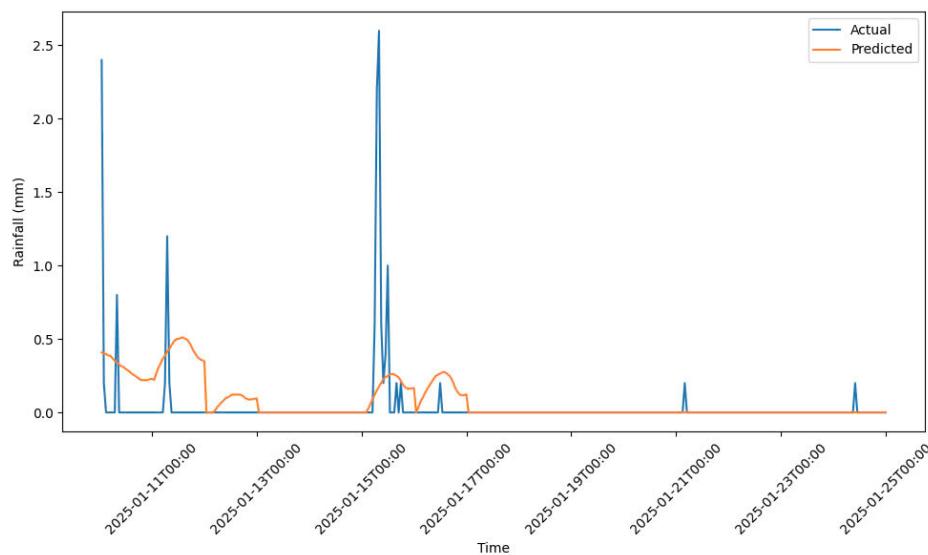
Quan sát hình 5.14, ta thấy rằng thông số lượng mưa theo hệ thống quan trắc tự động Vrain trong những tháng mưa tại trạm Chi Cục Thuỷ Lợi - Hà Tĩnh nói riêng và các tỉnh Bắc Trung Bộ nói chung có sự biến động mạnh mẽ hơn so với bộ dữ liệu IFS thông thường.

Lượng mưa lớn nhất theo giờ ghi nhận được vượt mức 40mm. Đây là một con số không nhỏ bởi nếu tính rộng ra, lượng mưa trong 24 giờ tại một khoảng thời gian bất kỳ trong mùa đều có thể vượt mức mưa vừa (lớn hơn 51mm) và mưa to (lớn hơn 100mm), vì vậy việc nắm bắt chính xác lượng mưa là điều không hề dễ dàng.

Quan sát hình 5.14 với thông số RMSE: 3.394, ta thấy lượng mưa dự đoán không thể bắt được xu hướng và lượng mưa lớn bất thường của thực tế, với độ lệch bình phương trung bình là 3.394 mm, vì vậy hệ thống dự báo nếu dựa vào bộ dữ liệu IFS kết hợp với Vrain này, hiện tại thích hợp cho việc dự báo trời mưa hay không thay vì dự đoán được lượng mưa sẽ xảy đến.

Điều này là bởi bộ dữ liệu lượng mưa Vrain chỉ có độ dài khoảng 4 năm (từ đầu năm 2021 đến nay), và lượng mưa lịch sử - đo dựa trên thiết bị phễu đo mưa tự động đều cho ra lượng mưa lịch sử với khoảng rất lớn - thấp nhất: 0mm/giờ và cao nhất: trên 40mm/giờ.

Tương tự như phần 5.3.3 - hình 5.13, nhóm sẽ đưa ra hình ảnh so sánh giữa lượng mưa dự báo và thực tế mô hình dự đoán bộ dữ liệu kết hợp lượng mưa Vrain trong khoảng thời gian từ ngày 10/01/2025 - 24/01/2025. Quan sát hình 5.15 để phân tích hiệu suất dự đoán tại trạm Chi Cục Thuỷ Lợi - tỉnh Hà Tĩnh.



Hình 5.15: Lượng mưa dự đoán và thực tế trong tháng 1 năm 2025

Như đã phân tích ở phần 5.3.4, lượng mưa thực tế cao nhất ghi nhận được là rất thấp - 2.5 mm/giờ do đây là quãng thời gian mùa khô nên chưa xuất hiện xu hướng mưa đột biến. Với các chỉ số đánh giá thấp hơn so với giai đoạn những tháng cuối năm 2024 - RMSE: 0.225 và MAE: 0.073.

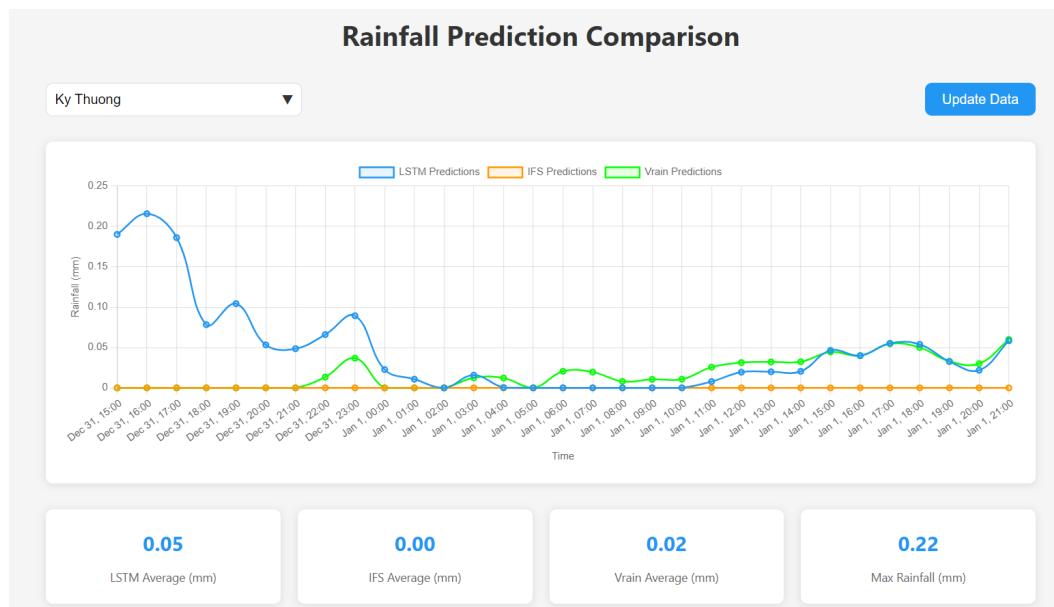
Về thông số dự báo chính xác có mưa hay không, do bộ dữ liệu kết hợp lượng mưa Vrain có độ chia nhỏ nhất là 0.2 mm (đối với IFS: 0.1 mm) nên phương pháp đánh giá sẽ tương tự nhưng thay đổi về số ngưỡng. Nghĩa là nếu lượng mưa dự đoán thấp hơn 0.2 mm thì sẽ được gán nhãn là không mưa. Như vậy, các con số thống kê và tỉ lệ đánh giá có mưa hoặc không trong 360 giờ như sau:

- Số giờ dự báo có mưa chính xác - TP: 13.
- Số giờ dự báo không mưa chính xác - TN: 263.
- Số giờ dự báo có mưa nhưng lại ghi nhận không mưa - FP: 58.
- Số giờ dự báo không mưa nhưng thực tế có mưa - FN: 2.
- Chỉ số Accuracy: 0.82.
- Chỉ số Precision: 0.18.
- Chỉ số Recall: 0.87.
- Chỉ số F1 Score: 0.3.

5.3.5 Giao diện người dùng

Nhằm hỗ trợ người dùng quan sát lượng mưa dự báo các khu vực dự báo tại các tỉnh thành nghiên cứu, nhóm sử dụng giao diện web để hiển thị những thông tin trên.

Giao diện web được ứng dụng ngrok - một công cụ giúp tạo đường hầm (tunnel) từ máy cục bộ đến một đường dẫn công khai trên internet. Công cụ này giúp chia sẻ máy chủ nội bộ (local server) với bên ngoài mà không cần phải cấu hình mạng như mở cổng hoặc thiết lập DNS.



Hình 5.16: Giao diện web trạm Kỳ Thuợng - Hà Tĩnh

Quan sát hình 5.16, ta thấy giao diện web ứng với mỗi trạm nghiên cứu sẽ có 3 đường thẳng ứng với lượng mưa 3 thông số lượng mưa dự báo: Mô hình LSTM của nhóm nghiên cứu ứng với màu xanh nước biển, mô hình nội suy IFS-ECMWF ứng với màu cam và lượng mưa dự báo Vrain ứng với màu xanh lục.

Chương 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN

6.1 Kết luận

Sau 6 tháng tìm hiểu đề tài và nghiên cứu phát triển, nhóm đã thành công hoàn thiện một hệ thống dự báo lượng mưa áp dụng trên khu vực các tỉnh Bắc Trung Bộ, với 3 tỉnh thành: Hà Tĩnh, Quảng Bình, Quảng Trị. Với việc xây dựng được bộ dữ liệu huấn luyện mô hình từ 2 nguồn dữ liệu: mô hình nội suy IFS-ECMWF, hệ thống quan trắc Vrain và dữ liệu cảm biến trong quá trình đo đạc thực tế, nhóm đã đưa ra lượng mưa dự báo cho các tỉnh thành khu vực Bắc Trung Bộ trong 24 giờ tiếp theo, giúp hỗ trợ người dân phòng trừ những thiệt hại nặng nề do những cơn mưa có cường độ lớn bất thường gây ra.

Nhóm đã xây dựng một hệ thống thiết bị đo đạc dữ liệu cảm biến bao gồm bo mạch chủ Goouuu-esp32, các cảm biến đo nhiệt độ, độ ẩm, áp suất không khí - DHT11 và BMP180. Bên cạnh đó là màn hình Oled và các nút nhấn giúp hỗ trợ người dùng thao tác với thiết bị. Các thành phần phần cứng của thiết bị đều là những linh kiện giá cả phải chăng, giúp đáp ứng tiêu chí tiết kiệm chi phí nhưng vẫn đảm bảo được về mặt hiệu năng thông qua việc hiện thực những chức năng và yêu cầu đề ra: giao tiếp với cảm biến và hệ thống máy chủ, hiển thị dữ liệu lên màn hình oled.

Về hệ thống máy chủ, mô hình dựa trên kiến trúc LSTM dự báo hoạt động khá tốt, các thông số đánh giá cho thấy mô hình rất đáng để tin cậy. Cơ sở dữ liệu dựa trên định dạng CSV rất dễ sử dụng và không tồn nhiều tài nguyên.

6.2 Khó khăn gặp phải

Trong quá trình hiện thực và phát triển hệ thống, nhóm cũng gặp phải không ít khó khăn, thách thức liên quan đến cơ sở lý thuyết, hoàn thiện hệ thống phần cứng thiết bị đầu cuối và hệ thống máy chủ để dự báo lượng mưa.

Về mặt cơ sở lý thuyết, mô hình dự báo lượng mưa nhóm đề xuất dựa trên 16 yếu tố khí tượng thuỷ văn làm đầu vào, số lượng yếu tố này tuy không ít nhưng cũng không phải là nhiều, xét riêng đến bản thân mô hình dự báo lượng mưa IFS-ECMWF thì dữ liệu đầu vào của mô hình đã lên đến hàng chục thậm chí hàng trăm yếu tố khác nhau,

bởi mỗi trong các yếu tố trên đều có ảnh hưởng qua lại, tác động lên nhau. Không chỉ vậy, việc dữ liệu lượng mưa Vrain lâu nhất là 4 năm tính đến thời điểm làm khoá luận (từ năm 2021 đến hết năm 2024) cũng gây khó khăn trong việc xây dựng mô hình dự báo bởi bộ dữ liệu này có thể chưa bao quát hết được những pha khí tượng có thể xảy ra trong một năm.

Về việc thu thập cập nhật dữ liệu theo thời gian, nếu ở hệ thống mã nguồn Open-Meteo cung cấp các API miễn phí nhằm có thể lấy dữ liệu một cách dễ dàng theo tọa độ và dữ liệu thu thập, thì ở bên hệ thống Vrain lại không hỗ trợ khả năng này, vì vậy quá trình phát thu thập dữ liệu lượng mưa lịch sử ở hệ thống Vrain đều phải làm thủ công, làm tốn không ít thời gian bởi dữ liệu theo giờ chỉ được lấy nếu thời hạn dữ liệu là 1 tháng, tuy nhiên mô hình nghiên cứu đến 9 điểm tọa độ chia đều có 3 tỉnh và mỗi địa điểm có thời gian đo đặc là 4 năm.

Ngoài ra, địa điểm nghiên cứu là các tỉnh thành Bắc Trung Bộ, không phải là thành phố Hồ Chí Minh cũng là một yếu tố khó khăn trong việc áp dụng hệ thống phần cứng để kiểm tra độ chính xác khi có sự xuất hiện các yếu tố cảm biến (bên cạnh các giá trị từ mô hình nội suy).

Về hệ thống thiết kế phần cứng, một điều khó khăn đối với nhóm chính là lúc đóng gói các thành phần lại với nhau để tiếng hành áp dụng trong thực tế. Số lượng linh kiện không nhỏ khiến việc đóng gói không dễ dàng. Ngoài ra những cảm biến đo đặc cũng có sai số không nhỏ, với ± 2 độ C với nhiệt độ, 5% với độ ẩm không khí và 0.02 hPa đối với áp suất không khí. Những sai số này cũng ảnh hưởng đến độ chính xác của lượng mưa dự báo so với thực tế.

Về mô hình dự báo, sử dụng công nghệ Học sâu dựa trên các mạng nơ-ron thần kinh nhân tạo khiến cho mô hình trở nên phức tạp và đòi hỏi máy tính có cấu hình tốt cũng như phải có GPU (Graphics Processing Unit - Đơn vị xử lý đồ họa) để hoạt động trơn tru. Việc huấn luyện tốn rất nhiều thời gian vì kích cỡ của dữ liệu là rất lớn (lên đến hơn 1 triệu dòng) và kiến trúc LSTM có quá nhiều hàm số tính toán.

Dữ liệu thời tiết có sự phân bố không đồng đều và đặc tính của mỗi loại yếu tố thời tiết là khác nhau. Có một số yếu tố mang tính tuần hoàn theo mùa trong khi một

số khác lại tuần hoàn theo ngày,... Dữ liệu bị phân tán rất cao và rất khó tìm được yếu tố quan trọng cho mô hình. Đặc biệt là dữ liệu về lượng mưa và độ phủ mây không biểu hiện một đường cong tron thường thấy nên không thể tính đạo hàm tại tất cả các điểm (không khả vi và liên tục) và độ dốc cũng biến động rất lớn.

Việc xây dựng hệ thống tự động tìm nạp dữ liệu và tự động dự báo theo giờ gấp phải vấn đề về giới hạn lượt gọi API từ Open-Meteo. Để có thể kiểm thử hệ thống cần phải đợi một khoảng thời gian sau mỗi lần thực thi.

6.3 Hướng phát triển

Về mặt cơ sở lý thuyết, có thể phát triển về bộ dữ liệu huấn luyện cho mô hình. Cụ thể, có thể ứng dụng những mô hình nội suy là mô hình cải tiến của IFS đó là WRF3kmIFS hay WRF3kmIFS-DA được đề cập ở bài báo [20]. Đây là những mô hình có độ phân giải thấp hơn so với IFS (3km so với 9km), từ đó cung cấp những điểm nút lưới sát gần hơn với địa điểm nghiên cứu, góp phần làm gia tăng độ chính xác lượng mưa dự báo. Ngoài ra những mô hình này đã được chứng minh rằng có khả năng dự báo lượng mưa tốt hơn trong thời hạn dài hơn (nhiều hơn 24 giờ) hoặc khi lượng mưa lớn hơn 50mm/h.

Hoặc có thể nghiên cứu về hệ số tương quan giữa lượng mưa từ hai nguồn trên nhằm đưa ra liên hệ giữa 2 loại giá trị, như trong bài báo [21] đã đưa ra một phương trình hồi quy $Y = 0.9717 * X$ với (X : lượng mưa dự báo, Y là lượng mưa theo trạm mặt đất lưu vực hồ Dầu Tiếng). Tuy nhiên, do hạn chế về nguồn dữ liệu dùng để phân tích (chỉ có dữ liệu của một cơn Bão năm 2018), nên kết quả phân tích có thể chưa phản ánh đúng bản chất về mối liên hệ giữa mưa dự báo toàn cầu và mưa thực đo.

Bên cạnh đó, nhóm đề xuất một phương pháp đó là tính độ lệch giữa lượng mưa bên phía IFS và bên phía Vrain trong 4 năm (2021 - 2024), sau đó có thể áp dụng độ lệch đó với toàn bộ dữ liệu lượng mưa trong 20 năm bên phía IFS. Phương pháp này đòi hỏi khả năng thống kê bởi độ lệch về lượng mưa giữa 2 mùa trong năm: mùa nắng và mùa mưa là khác nhau hoặc chia ra theo tháng nhằm giúp việc tính độ lệch càng mịn, đòi hỏi người thực hiện phải nắm vững cả về kiến thức về khí tượng thủy văn ở khu vực Bắc Trung Bộ, Việt Nam và kiến thức lập trình, toán học.

Về thiết bị đo đạc, nhóm đề xuất sử dụng nhiều cảm biến hơn để đo đạc dữ liệu như tốc độ gió, hướng gió, nhiệt độ, áp suất tại một độ cao cụ thể (2m hoặc 20m so với mực nước biển), đây cũng là những yếu tố mà mô hình IFS đang áp dụng để dự báo lượng mưa hàng giờ. Ngoài ra, hệ thống cũng có thể nâng cao chất lượng dữ liệu thu thập được với độ sai lệch số thấp hơn bằng cách ứng dụng các cảm biến chất lượng hơn nhằm nâng cao khả năng dự báo.Thêm vào đó, có thể ứng dụng thêm những tính năng như tìm địa điểm hiện tại mà hệ thống đang được lắp đặt, ứng dụng các giao thức khác như BLE, TCP để thiết lập, cài đặt cấu hình cho thiết bị.

Dữ liệu cần được phân tích và khai thác sâu hơn để tìm được những đặc tính cũng như quy luật chưa tìm thấy. Cần thực hiện thêm các phương pháp xử lý dữ liệu như tăng cường dữ liệu, phát hiện outlier, kỹ thuật trích chọn đặc trưng (data engineering và data extraction),... để dữ liệu trở nên cân bằng hơn và phù hợp với mô hình hơn; hay sử dụng các mô hình hồi quy đánh giá các đặc trưng theo đa biến và kỹ thuật PCA (Principal Components Analysis) để tìm ra những tổ hợp đặc trưng phù hợp với mô hình.

Mô hình chưa được xây dựng quá tốt, đặc biệt là các siêu tham số chưa được tối ưu tốt dẫn đến việc huấn luyện mất nhiều thời gian cũng như độ chính xác giảm đi. Có thể phát triển các mô hình phức tạp hơn như (ConvLSTM, BiLSTM, Transformer,...). Cần so sánh kết quả với các mô hình thống kê truyền thống trước đây như các mô hình hồi quy, ARIMA,... Đưa ra thêm các độ đo hiệu suất mô hình sử dụng trọng số.

Cơ sở dữ liệu còn quá đơn giản, cần nâng cấp lên các hệ quản trị cơ sở dữ liệu tiên tiến hơn như MySQL, Firebase,...

Giao diện thiết bị đo đạc, và giao diện web hiện nay còn khá thô sơ, chưa được đẹp mắt dẫu đã có thể hiện thực được chức năng cơ bản là hiển thị lượng mưa dự báo ở các tỉnh thành từ mô hình. Tuy nhiên, nhằm để tăng tính hướng người dùng của hệ thống, nhóm muốn đầu tư nhiều thời gian hơn để phát triển các tính năng, cải thiện thẩm mỹ giao diện và hiệu năng của các thành phần này góp phần tăng tính ứng dụng của hệ thống.

TÀI LIỆU THAM KHẢO

- [1] N. K. Vân, “Vai trò của hình thái địa hình đối với mưa lớn ở vùng bắc trung bộ và sự phân hóa giữa bắc và nam đèo ngang,” *Tạp chí CÁC KHOA HỌC VỀ TRÁI ĐẤT*, vol. 34, p. 1, 2012.
- [2] P. T. Hà, “Tác động của biến đổi khí hậu đến lũ quét, sạt lở đất,” *Báo Tài nguyên và Môi trường*, 2014.
- [3] T. T. T. Huyền, N. T. Thảo, N. V. Vinh, M. V. Thông, and N. T. Hằng, “Dự báo tác động của biến đổi khí hậu đến vùng bắc trung bộ, giai đoạn 2016-2030,” *Tạp chí Tài nguyên và Môi trường*, 2021, phân hiệu Trường Đại học TN&MT Hà Nội tại tỉnh Thanh Hóa.
- [4] L. Espeholt, S. Agrawal, C. Sønderby, M. Kumar, J. Heek, C. Bromberg, C. Gazen, R. Carver, M. Andrychowicz, J. Hickey *et al.*, “Deep learning for twelve hour precipitation forecasts,” *Nature communications*, vol. 13, no. 1, pp. 1–10, 2022.
- [5] F. Alqahtani, “Ai-driven improvement of monthly average rainfall forecasting in mecca using grid search optimization for lstm networks,” *Journal of Water and Climate Change*, vol. 15, no. 4, pp. 1439–1458, 2024.
- [6] A. Zanella, S. Zubelzu, M. Bennis, M. Capuzzo, and P. Tarolli, “Internet of things for hydrology: Potential and challenges,” in *2023 18th Wireless On-Demand Network Systems and Services Conference (WONS)*. IEEE, 2023, pp. 114–121.
- [7] M. S. Binetti, C. Campanale, C. Massarelli, and V. F. Uricchio, “The use of weather radar data: Possibilities, challenges and advanced applications,” *Earth*, vol. 3, no. 1, pp. 157–171, 2022.

- [8] U. Chrobak, “We can no longer rely on historical data to predict extreme weather,” *Popular Science*, 2020.
- [9] V. Academy, “La nina: Việt nam chịu ảnh hưởng như thế nào?” *VSG Academy*, 2024, viện Vũ trụ và Địa khôn gian Vega.
- [10] M. Thành, D. Hưng, V. Tý, V. Dung, and H. Ngà, “Bắc trung bộ thích ứng biến đổi khí hậu - bài 1: Những tín hiệu phức tạp đáng lo ngại,” *Báo Tin Tức*, 2024.
- [11] S. Agrawal, L. Barrington, C. Bromberg, J. Burge, C. Gazen, and J. Hickey, “Machine learning for precipitation nowcasting from radar images,” *arXiv preprint arXiv:1912.12132*, 2019.
- [12] Q.-K. Tran and S.-k. Song, “Multi-channel weather radar echo extrapolation with convolutional recurrent neural networks,” *Remote Sensing*, vol. 11, no. 19, p. 2303, 2019.
- [13] Y. Wehbe, M. Temimi, and R. F. Adler, “Enhancing precipitation estimates through the fusion of weather radar, satellite retrievals, and surface parameters,” *Remote Sensing*, vol. 12, no. 8, p. 1342, 2020.
- [14] P. Benevides, J. Catalao, and G. Nico, “Neural network approach to forecast hourly intense rainfall using gnss precipitable water vapor and meteorological sensors,” *Remote Sensing*, vol. 11, no. 8, p. 966, 2019.
- [15] D. Tran Anh, T. Duc Dang, and S. Pham Van, “Improved rainfall prediction using combined pre-processing methods and feed-forward neural networks,” *J*, vol. 2, no. 1, pp. 65–83, 2019.
- [16] L. Đức Cương and Đặng Ngọc Sơn, “Nghiên cứu sử dụng số liệu radar phục vụ cảnh báo, dự báo bão kèm mưa lớn và xây dựng các cấp độ rủi ro thiên tai ở khu vực bắc trung bộ,” *Tạp chí Khí tượng Thủy văn*, no. 691, pp. 11–20, 2018.

- [17] V. V. Hòa, N. Đăng Quang, and Đăng Anh Tuấn, “Số liệu dự báo của ecmwf và khả năng ứng dụng trong dự báo hạn mưa ở việt nam,” *Tạp chí Khí tượng Thủy văn*, no. 669, pp. 1–6, 2016.
- [18] V. V. Hòa, “Nghiên cứu so sánh kỹ năng dự báo mưa lớn khu vực miền trung và tây nguyên của một số mô hình toàn cầu,” *Tạp chí Khí tượng Thủy văn*, no. 667, pp. 1–8, 2016.
- [19] T. Đức Bá, V. V. Hòa, and Đoàn Quang Trí, “Đánh giá chất lượng dự báo mưa hạn ngắn của mô hình ifs trên khu vực bắc trung bộ,” *Tạp chí Khí tượng Thủy văn*, no. 697, pp. 33–43, 2019.
- [20] N. Nga, C. Thanh, M. Hưng, and D. Tiên, “Đánh giá kĩ năng dự báo mưa định lượng từ mô hình quy mô toàn cầu và khu vực phân giải cao cho khu vực bắc bộ,” *Tạp chí Khí tượng Thủy văn*, no. 730, pp. 79–92, 2021.
- [21] Đinh Công Sản, L. N. Thanh, and N. V. Lanh, “Dự báo mưa trên lưu vực hồ dồn tiếng từ tài liệu dự báo thời tiết toàn cầu phục vụ dự báo dòng chảy lũ đến và điều tiết hồ trong mùa lũ,” *Tạp chí Khoa học và Công nghệ Thủy lợi*, no. 56, pp. 1–10, 2019.

PHỤ LỤC