HỆ THỐNG GIÁM SÁT VÀ DỰ BÁO LƯỢNG MƯA CÁC TỈNH BẮC TRUNG BỘ

Nguyễn Tiến Toàn, 21521548

Khoa Kỹ thuật máy tính, Trường Đại học Công Nghệ Thông tin

TS. Đoàn Duy

Khoa Kỹ thuật máy tính, Trường Đại học Công Nghệ Thông tin

*Tóm tắt*—Những thiên tai liên quan đến khí tượng thuỷ văn thường bắt đầu từ những trận mưa có cường độ lớn bất thường như lũ lụt, lũ quét, sạt lở gây ra thiệt hại lớn về tài sản người và của. Cùng với sự ảnh hưởng ngày càng lớn của biến đổi khí hậu, những hệ thống, mô hình dự báo lượng mưa ngày càng được phát triển nhằm hỗ trợ công tác phòng chống thiên tai cũng như dựa vào khả năng dự đoán lượng mưa có thể trút xuống khu vực trong tương lai gần. Trong khuôn khổ của khoá luận, nhóm đề xuất một mô hình dự báo lượng mưa ứng dụng kĩ thuật học sâu với dữ liệu dựa trên mô hình nội suy IFS-ECMWF cho địa bàn 3 tỉnh khu vực Bắc Trung Bộ (Hà Tĩnh, Quảng Bình, Quảng Trị). Hệ thống sẽ thu thập những dữ liệu khí tượng thuỷ văn từ mô hình IFS rồi từ đó đi qua thuật toán LSTM với những mạng nơ-ron nhằm đưa ra lượng mưa dự đoán trong 24 giờ tiếp theo. Hệ thống bao gồm một máy chủ lưu trữ lượng mưa lịch sử và lượng mưa dự đoán và cung cấp một nền tảng giao diện người dùng nhằm cung cấp một góc nhìn toàn cảnh dựa chuỗi thời gian lượng mưa hiển thị.

Từ khóa—Khí tượng thuỷ văn, lượng mưa, mô hình dự báo, nội suy, IFS-ECMWF, Bắc Trung Bộ, học sâu, LSTM.

# GIỚI THIỆU

Khu vực Bắc Trung Bộ thường xuyên chịu ảnh hưởng của mưa lớn do vị trí địa lý [1], gây ra nhiều thiệt hại nghiêm trọng [2]. Do đó, dự báo lượng mưa có vai trò quan trọng trong phòng chống thiên tai và nghiên cứu xu hướng khí tượng thủy văn, trong bối cảnh biến đổi khí hậu diễn ra ngày càng phức tạp [3].

Các hệ thống dự báo lượng mưa hiện đại đều ứng dụng trí tuệ nhân tạo với các mô hình học máy và học sâu để dự đoán lượng mưa trong cả ngắn hạn [4] và dài hạn [5]. Và chúng dựa trên nhiều yếu tố khí tượng thủy văn được thu thập từ các trạm đo, vệ tinh hoặc ra-đa. Tuy nhiên, các hệ thống dựa vào cảm biến có thể cho sai số phần cứng [6]. Trong khi dữ liệu vệ tinh và ra-đa bị ảnh hưởng bởi các yếu tố như mây và sương mù [7]. Thêm vào đó, hiện tượng thời tiết cực đoan ngày càng phức tạp cũng ảnh hưởng đến hiệu quả dự báo của các mô hình dựa trên dữ liệu quá khứ [8]. Như hiện tượng La Nina và biến đổi khí hậu gây ra những trận mưa lớn đột biến tại khu vực Bắc Trung Bộ, với trung bình 9-10 cơn bão và 3-4 cơn áp thấp nhiệt đới hằng năm [9] [10].

Các nghiên cứu gần đây về dự báo lượng mưa có những cách tiếp cận khác nhau: Sử dụng ảnh vệ tinh với độ phân giải 1km [11], ứng dụng mạng CRNN với dữ liệu ra-đa 3D [12], kết hợp nhiều nguồn dữ liệu với mạng GWR và ANN [13], và tích hợp dữ liệu GNSS [14]. Mỗi phương pháp đều có ưu điểm riêng nhưng vẫn tồn tại những hạn chế như thời gian dự báo ngắn, hiệu suất chưa ổn định, và thiếu dữ liệu dài hạn để nắm bắt các hiện tượng mưa cực đoan.

Tại Việt Nam, cũng đã có nhiều nghiên cứu về như kết hợp mô hình ANN và SANN với dữ liệu 39 năm tại Cà Mau [15], hay phân tích dữ liệu ra-đa thời tiết kết hợp với số liệu lịch sử từ 1980-2017 tại khu vực Bắc Trung Bộ [16]. Đặc biệt, mô hình ECMWF với độ phân giải 9km [17] cho kết quả tốt hơn các mô hình GFS, NOGAPS và GSM [18], dù còn hạn chế trong dự báo mưa lớn và hạn dự báo 72h tại khu vực Bắc Trung Bộ [19].

Trong khuôn khổ khoá luận, nhóm đề xuất một mô hình dự báo lượng mưa dựa trên dữ liệu đầu vào là những yếu tố khí tượng thuỷ văn từ mô hình nội suy IFS-ECMWF sau đó được xử lý qua mạng nơ-ron LSTM nhằm đưa ra dự báo lượng mưa mà một khu vực có thể hứng chịu. Khu nghiên cứu gồm 3 tỉnh thành Bắc Trung Bộ: Hà Tĩnh, Quảng Bình và Quảng Trị với mỗi tỉnh sẽ 3 điểm dự báo. Thời gian dự báo là 24 giờ.

# GIẢI PHÁP

## Xây dựng bộ dữ liệu

Mô hình sẽ truy cập, thu thập dữ liệu lịch sử khí tượng thuỷ văn mô hình nội suy IFS-ECMWF thông qua nền tảng mã nguồn mở Open Meteo. Bộ dữ liệu có khoảng thời gian từ năm 2010 đến nay và được chia thành 2 phần: 80% cho huấn luyện và 20% đánh giá.

Để đánh giá tác động của các yếu tố (nhiệt độ, độ ẩm, độ bao phủ mây …) ảnh hưởng đến lượng mưa, thuật toán Random Forest Regression được sử dụng để chọn lọc những đặc trưng có tác động lớn nhất.

Trong đó: Iilà tầm quan trọng của đặc trưng i, T là tổng số cây quyết định, ΔFij là sự thay đổi độ chính xác cây quyết định j khi i bị loại bỏ và Δi là tổng sự thay đổi chính xác của tất cả các cây quyết định khi đặc trưng i bị loại bỏ.

Đầu vào mô hình dự báo là chuỗi dữ liệu gồm 24 bước (tương ứng với 24 giờ trước đó) và đầu ra lượng mưa dự báo của 24 giờ tiếp theo dựa theo kiến trúc 2 lớp mạng nơ-ron LSTM.

## Thuật toán dự báo lượng mưa

Chuỗi dữ liệu sau khi được thu thập sẽ được tiền xử lý trước khi được đưa vào LSTM. Đầu tiên, lượng mưa (mm) sẽ được chuyển đổi bằng phép toán *logarit* nhằm giảm độ lệch, qua đó giúp mô hình nắm bắt tốt hơn các biến động về cường độ mưa.

Tiếp theo, dữ liệu được chuẩn hoá theo tỷ lệ Min-Max, qua đó cải thiện độ chính xác mô hình với x là giá trị gốc của đặc trưng, xscaled là giá trị đã chuẩn hoá, nằm trong khoảng [0,1].

Chuỗi dữ liệu sau khi đã chuẩn hoá sẽ được đưa vào lần lượt 2 lớp mạng nơ ron LSTM với lớp LSTM thứ nhất bao gồm *128 đơn vị* (*units*), với giá trị đầu ra là vectơ trạng thái ẩn ht128 chiều. Để tính toán trạng thái ẩn mạng nơ ron, mô hình sử dụng hàm kích hoạt *tanh*.

A diagram of a flowchart

Description automatically generated

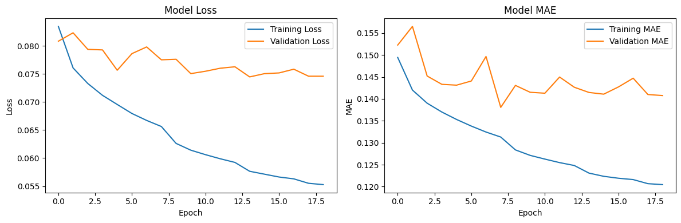
Chuỗi trạng thái ẩn tính toán được chính là đầu vào cho lớp LSTM thứ hai bao gồm *64 đơn vị*, với các phép toán tương tự với lớp ẩn thứ nhất. Đầu ra của lớp ẩn thứ 2 là một vectơ 64 chiều trạng thái ẩn ht.

Lớp đầu ra LSTM chính là một lớp Dense sẽ xử lý vectơ 24 chiều. Tuy nhiên, nhằm đảm bảo lượng mưa không âm, mô hình ứng dụng hàm kích hoạt *ReLU*:

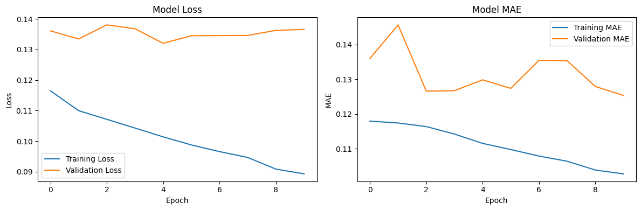
Giá trị ypred sau đó sẽ được chuyển đổi thành lượng mưa dự báo sau khi thực hiện phép lấy mũ (exponentiation).

# KẾT QUẢ

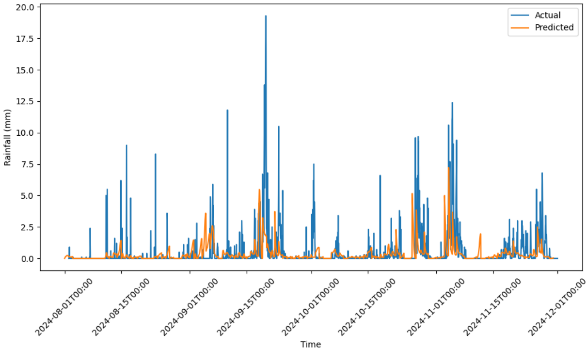
Quá trình huấn luyện mô hình dự đoán theo dữ liệu IFS diễn ra thuận lợi, và sai số đang giảm dần. Chỉ số đánh giá tập Test RMSE: 0.5959 và MAE: 0.3013.



Tuy nhiên việc huấn luyện mô hình bộ dữ liệu IFS kết hợp với bộ dữ liệu IFS kết hợp lượng mưa Vrain cho thấy bất lợi, thông qua việc chỉ số hàm mất mát Validation rất biến động, dẫn đến mô hình chưa học được các quy luật mưa dài hạn. Các chỉ số đánh giá bao gồm: RMSE: 3.23 và MAE: 0.6639.

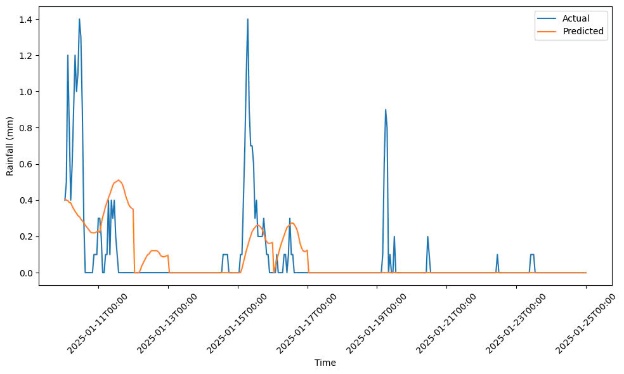


Đánh giá trên tập Test dài hạn, mô hình dự báo trên bộ dữ liệu IFS-ECMWF tại trạm Chi Cục Thuỷ lợi – tỉnh Hà Tĩnh từ ngày 01/08/2024 đến 01/12/2024 cho thấy lượng mưa dự báo có thể nắm bắt được xu hướng của của những trận mưa lớn. Các chỉ số đánh giá: RMSE: 1.445 và MAE: 0.561.



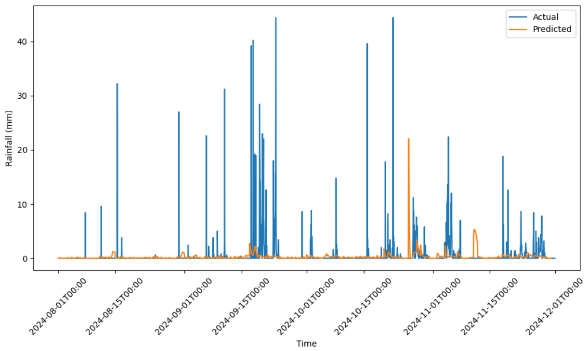
Đối với quãng thời gian mùa khô, ví dụ như trong 14 ngày tháng một năm 2025 (10/01/2025 – 24/01/2025) cũng tại trạm Chi Cục Thuỷ Lợi, thì mô hình cho hiệu suất dự báo tốt hơn: RMSE: 0.21 và MAE: 0.089.

Với khả năng dự báo trạng thái mưa, hiệu suất được thể hiện thông qua các chỉ số: Accuracy: 0.83, Precision: 0.56, Recall: 0.73 và F1 Score: 0.63.



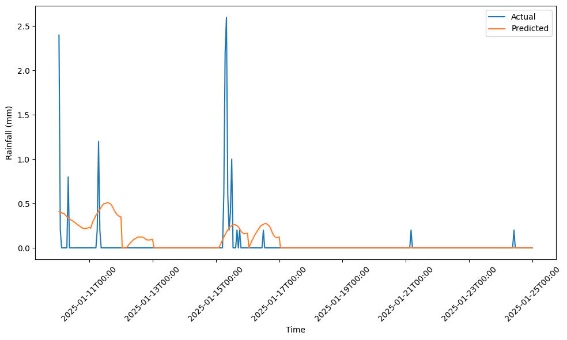
Đối với bộ dữ liệu IFS kết hợp Vrain cũng tại trạm Chi Cục Thuỷ lợi, cùng quãng thời gian thì mô hình lại không thể nắm bắt được xu thế lượng mưa. Điều này là do dữ liệu lượng mưa lịch sử của Vrain có biến động nhiều hơn so với lượng mưa lịch sử IFS và bộ dữ liệu từ Vrain chỉ có thời gian 4 năm. Với các chí số đánh giá là RMSE: 3.39 và MAE = 0.82.

Điều này chứng tỏ mô hình ứng với bộ dữ liệu Vrain hiện tại thích hợp hơn cho việc dự báo trạng thái mưa thay vì dự báo lượng mưa theo chuỗi thời gian 24 giờ tiếp theo.



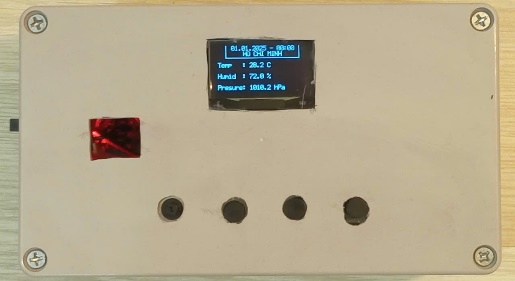
Tương tự như vậy, nhóm cũng đưa ra biểu đồ so sánh giữa lượng mưa dự đoán và thực tế trong tháng 1 năm 2025 (từ ngày 10/01 đến ngày 24/01) nhằm thể hiện hiệu suất dự đoán của mô hình trong thời gian mùa khô.

Quan sát hình dưới đây, có thể thấy lượng mưa cao nhất trong quãng thời gian này là 2.5 mm/h cho thấy không xuất hiện các xu hướng mưa lớn đột biến, với đa số thời gian ghi nhận thực tế là không mưa. Các chỉ số đánh giá độ lệch: RMSE: 0.225 và MAE: 0.073.



Ngoài ra, đối với khả năng dự báo trạng thái mưa, mô hình cho ra các chỉ số: Accuracy: 0.82, Precision: 0.18, Recall 0.87 và F1 Score: 0.3.

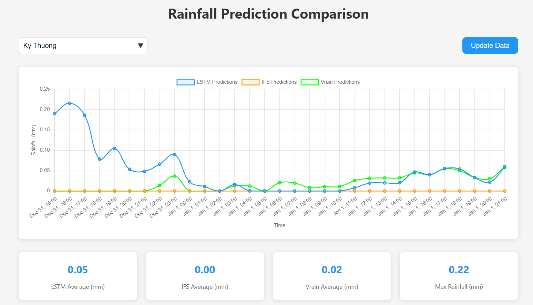
Ngoài ra nhóm đã hoàn thiện thiết bị phần cứng có khả năng thu thập dữ liệu khí tượng và hiển thị lượng mưa dự báo lên màn hình oled và giao tiếp với hệ thống máy chủ thông qua giao thức MQTT.





Những nút nhấn được tích hợp trên thiết bị đầu cuối đóng vai trò để tắt/mở giao diện và thay đổi giao diện màn hình, giữa giao diện thông tin khí tượng và giao diện màn hình lượng mưa dự báo cũng như giao diện dự đoán lượng mưa trung bình trong 1 giờ, 6 giờ và 24 giờ tiếp theo giữa 3 tỉnh: Hà Tĩnh, Quảng Bình, Quảng Trị.

Nhóm cũng hiện thực giao diện trang web để hỗ trợ người dùng quan sát lượng mưa dự đoán trong 24 giờ tiếp theo tại các trạm nghiên cứu trên mỗi tỉnh.



# KẾT LUẬN

Hiện tại, nhóm đã hiện thực mô hình đã có thể dự báo lượng mưa trong 24 giờ tiếp theo. Thiết bị biên có khả năng thu thập dữ liệu cảm biến và giao tiếp với máy chủ thông qua MQTT. Hệ thống máy chủ có khả năng thu thập dữ liệu và dự đoán lượng mưa trong 24 giờ kế tiếp và giao diện trang web nhằm hiển thị lượng mưa dự báo cho người dùng.

Về hướng phát triển tiếp theo, các mô hình WRF3kmIFS và WRF3kmIFS-DA [20] với độ phân giải 3km cho độ chính xác cao hơn mô hình IFS (9km), đặc biệt với dự báo trên 24 giờ và lượng mưa trên 50mm/h. Một hướng tiếp cận khác là nghiên cứu hệ số tương quan giữa lượng mưa dự báo và thực tế, như phương trình hồi quy Y = 0.9717X được đề xuất trong [21], dù còn hạn chế về nguồn dữ liệu nghiên cứu.

Ngoài ra, mô hình cần cải thiện thuật toán và các tham số nhằm có lượng mưa dự báo chính xác hơn. Hoặc nâng cấp các thiết bị cảm biến dùng để đo đạc hoặc cải tiến giao diện người dung – trang web nhằm tăng tính thẩm mỹ khi người dùng sử dụng.

# 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 OnDemand 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 không 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 b.o 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 Đặng Anh Tuấn, “Số liệu dự báo của ecmwf và khả năng ứng dụng trong dự báo hạn mùa 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. Tiến, “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 dầu 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*