



APPLYING ARTIFICIAL NEURAL NETWORK (ANN) MODELS TO CALCULATE THE ROLLER LENGTH OF HYDRAULIC JUMPS IN HORIZONTAL RECTANGULAR CHANNELS

Hung Viet Ho*

Faculty of Water Resources Engineering, Thuyloi University, 175 Tay Son, Hanoi, Viet Nam

ARTICLE INFO

TYPE: Research Article

Received: 28/01/2026

Revised: 08/04/2026

Accepted: 13/04/2026

Published online: 15/04/2026

<https://doi.org/10.47869/tcsj.77.3.3>

* *Corresponding author*

Email: hohung.thuyluc@tlu.edu.vn

Abstract. The hydraulic jump roller length plays a decisive role in determining the size and efficiency of energy dissipation structures downstream of spillways and culverts. Therefore, accurate estimation of this parameter is a critical requirement in the design of hydraulic engineering works. This study developed and evaluated eight Artificial Neural Network (ANN) models to predict the hydraulic jump roller length within horizontal prismatic rectangular channels. Four input variables and one model output were selected based on the Buckingham π theorem to ensure representativeness and dimensional homogeneity, an approach that has not been thoroughly investigated previously. In addition, machine learning models were employed to evaluate how gravity, drag, and channel width affect roller length, improving prediction accuracy. Test results show that all ANN models achieved lower errors compared to traditional empirical equations, with Nash–Sutcliffe Efficiency (NSE) coefficients exceeding 0.975. Notably, the two-hidden-layer ANN with 16 neurons per layer attained the highest NSE value (0.992) and the lowest relative error (4.32%), confirming this configuration as the optimal predictive model and a viable alternative to existing empirical methods.

Keywords: roller length, hydraulic jumps, ML model, hydraulics, ANN.

@ 2026 University of Transport and Communications



ÁP DỤNG CÁC MÔ HÌNH MẠNG NƠ-RON NHÂN TẠO (ANN) TÍNH TOÁN CHIỀU DÀI KHU XOÁY CỦA NƯỚC NHẢY TRONG KÊNH CHỮ NHẬT NĂM NGANG

Hồ Việt Hùng*

Khoa Kỹ thuật tài nguyên nước, Trường Đại học Thủy lợi, 175 Tây Sơn, Hà Nội, Việt Nam

THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học

Ngày nhận bài: 28/01/2026

Ngày nhận bài sửa: 08/04/2026

Ngày chấp nhận đăng: 13/04/2026

Ngày xuất bản Online: 15/04/2026

<https://doi.org/10.47869/tcsj.77.3.3>

* Tác giả liên hệ

Email: hohung.thuyluc@tlu.edu.vn

Tóm tắt. Chiều dài xoáy cuộn của nước nhảy có vị trí quan trọng trong việc xác định kích thước và hiệu quả của các công trình tiêu năng phía hạ lưu đập tràn, cống ngầm. Vì vậy, việc ước tính chính xác thông số này là yêu cầu cấp thiết trong thiết kế các công trình thủy lợi. Do đó, nghiên cứu này đã phát triển và đánh giá tám mô hình mạng nơ ron nhân tạo (artificial neural networks – ANN) nhằm dự báo chiều dài khu xoáy của nước nhảy trên kênh lăng trụ mặt cắt chữ nhật, đáy bằng. Bốn biến đầu vào và một biến đầu ra của mô hình được lựa chọn theo định lý π -Buckingham nhằm đảm bảo tính đại diện và không thứ nguyên, việc này chưa được nghiên cứu kỹ trước đây. Hơn nữa, các mô hình học máy đã được ứng dụng để đánh giá tác động của trọng lực, lực cản và chiều rộng lòng dẫn đến chiều dài khu xoáy giúp tăng cường hiệu suất dự báo. Các kết quả kiểm định xác nhận rằng tất cả mô hình ANN đều có sai số thấp hơn các phương trình thực nghiệm truyền thống, với hệ số Nash–Sutcliffe (NSE) cao hơn 0,975. Đặc biệt, ANN hai lớp ẩn với 16 nơ-ron mỗi lớp đạt giá trị NSE cao nhất (0,992) và sai số tương đối nhỏ nhất (4,32%), khẳng định cấu hình này là tối ưu cho mô hình dự báo và có thể thay thế các phương pháp thực nghiệm hiện hành.

Từ khóa: chiều dài xoáy cuộn, hiện tượng nước nhảy, mô hình học máy, thủy lực, ANN.

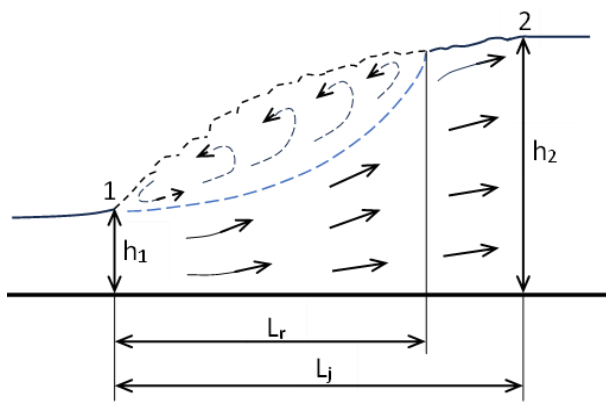
@ 2026 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

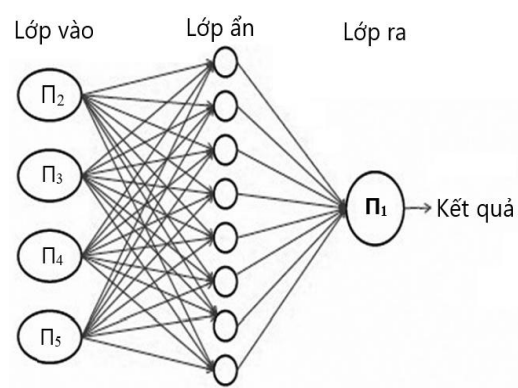
Nước nhảy xuất hiện trong dòng chảy khi độ sâu tầng đột ngột, vận tốc giảm đột ngột, dòng nước chuyển từ chảy xiết với Froude lớn hơn 1 sang chảy êm với Froude nhỏ hơn 1 [1]. Nước nhảy có các đặc trưng hình học sau đây: các độ sâu liên hiệp của nước nhảy, độ sâu

cuối xoáy cuộn, chiều dài xoáy cuộn, chiều dài toàn bộ nước nhảy. Các đặc trưng này cần được xác định khi tính toán thủy lực và thiết kế công trình thủy lợi, giao thông [2]. Chiều dài khu xoáy L_r là một yếu tố rất rõ ràng (xem Hình 1), có thể quan sát trong thực tế và đo đạc chính xác trong phòng thí nghiệm, dùng để tính chiều dài của bề tiêu năng phía hạ lưu tràn xả lũ, do vậy cần được tính toán chính xác.

Từ trước đến nay, các phương trình thực nghiệm thường được áp dụng để ước tính chiều dài khu xoáy của nước nhảy trong kênh hình chữ nhật đáy bằng, nếu biết độ sâu và số Froude trước nước nhảy [3], [4], hoặc biết số Froude trước nước nhảy và hai độ sâu liên hiệp [5]. Một số nhà khoa học như Safranez, Pietrkowski, Schroder, Sarma-Newnham và Hager [4] đã đưa ra các phương trình để tính chiều dài xoáy cuộn theo số Froude trước nước nhảy. Ngoài ra, Peterka (1984) [6] và Bretz (1988) đã trình bày kết quả nghiên cứu trong các máng kính mặt cắt chữ nhật [7]. Các công thức thực nghiệm thường có phạm vi ứng dụng khác nhau, cho kết quả không đồng nhất và chưa xét đến đầy đủ các yếu tố thủy lực, nên sai số còn khá lớn. Vì vậy, để phù hợp với thực tế và khắc phục các hạn chế của phương pháp thực nghiệm, bài báo này đề xuất phương pháp mô hình toán nhằm đối chiếu với các công thức đang sử dụng và bổ sung giải pháp giúp ước tính chiều dài xoáy cuộn được chính xác và thuận tiện hơn.



Hình 1. Các độ sâu liên hiệp h_1 , h_2 , chiều dài xoáy cuộn L_r , chiều dài toàn bộ nước nhảy L_j .



Hình 2. Cấu trúc của mạng ANN có ba lớp với số nơ-ron là 4-8-1.

Các thuật toán trí tuệ nhân tạo (AI) và học máy (ML) hiện được áp dụng phổ biến trong nhiều lĩnh vực nhằm giải quyết các vấn đề thực tiễn phức tạp. Đặc biệt, các thuật toán ML liên quan đến ML có giám sát đã được triển khai rộng rãi trong nghiên cứu thủy lực. Từ những năm 2012, 2013, mô hình mạng nơ-ron nhân tạo (ANN) đã được ứng dụng để xác định chiều dài toàn bộ nước nhảy trong kênh mặt cắt chữ nhật [8] và kênh chữ U [9]. Cho đến năm 2021, các thuật toán như adaptive neuro-fuzzy inference system (ANFIS), ANFIS-PSO (particle swarm optimization), least absolute shrinkage and selection operator cũng được dùng để ước tính tỷ số hai độ sâu liên hiệp trong kênh mặt cắt chữ nhật [10]. Trên thế giới và ở Việt Nam, chiều dài toàn bộ nước nhảy đã được dự báo bằng các mô hình: Rừng ngẫu nhiên (Random Forest), Cây bổ sung (Extra Trees), Tăng cường độ dốc (Gradient Boosting), Tăng cường phân loại (Categorical Boosting), Tăng cường thích ứng (Adaptive Boosting), Máy Vector hỗ trợ (Support Vector Machine) [11], [12]. Hơn thế nữa, các mô hình trên còn dự báo chiều dài xoáy cuộn [13] và tỷ số của hai độ sâu liên hiệp [14]. Các công bố trước đây cho thấy, các mô hình ML cho kết quả chính xác và dự báo tốt hơn nhiều so với các công thức thực nghiệm thông thường.

Như vậy, các nghiên cứu trước đây thường ứng dụng ANN để ước lượng chiều dài toàn bộ nước nhảy dựa trên một bộ dữ liệu với tập kiểm định chỉ chiếm 25%. Nghiên cứu này mở rộng bằng cách áp dụng ANN trên nhiều bộ dữ liệu tổng hợp, vừa tính toán chiều dài khu xoáy vừa phân tích tác động của các biến đầu vào. Khi phạm vi kiểm định tăng lên, kết quả dự báo đáng tin hơn và có sức thuyết phục cao hơn.

Nghiên cứu này sử dụng Python 3.7, Keras và Matplotlib để xây dựng và đánh giá tám mô hình ANN nhằm mục đích dự báo chiều dài xoáy cuộn của nước nhảy trong kênh hình chữ nhật đáy bằng. Các mô hình ANN với một hoặc hai lớp ẩn và số lượng nơ-ron thay đổi đã được xây dựng nhằm phân tích ảnh hưởng của cấu trúc mạng đến mức độ chính xác trong dự báo. Kết quả dự báo được so sánh với nhau và đối chiếu với hai công thức thực nghiệm hiện có, qua đó xác định mô hình tối ưu.

2. PHƯƠNG PHÁP NGHIÊN CỨU VÀ CÁC DỮ LIỆU

2.1. Phương pháp ứng dụng các công thức thực nghiệm

Hiện tại, chiều dài xoáy cuộn chỉ được ước tính bằng các công thức thực nghiệm do chưa có công thức lý thuyết thuần túy. Nghiên cứu này tổng hợp các công thức thực nghiệm dùng để ước lượng chiều dài xoáy cuộn trong nước nhảy trên kênh mặt cắt chữ nhật đáy bằng (xem Bảng 1). Giá trị L_r được tính dựa trên số Froude trước nước nhảy, ký hiệu là Fr_1 . Tuy nhiên, các công thức này bỏ qua độ nhám tương đối, chiều rộng kênh cũng như độ nhót của nước, với giả thiết rằng các yếu tố này không đáng kể so với vai trò chủ đạo của số Fr_1 .

Bảng 1. Các phương trình thực nghiệm tính chiều dài xoáy cuộn.

TT	Tác giả	Số Froude Fr_1	Công thức tính
1	Safranez (1929)	1,72 – 19,1	$\frac{L_r}{h_1} = 6Fr_1$
2	Pietrkowski (1932)	5,5 – 19,8	$\frac{L_r}{h_1} = 5,9Fr_1$
3	Schroder (1963)	1,83 – 9,93	$\frac{L_r}{h_1} = 40 \times erf \left\{ \frac{1}{16} (1 + 8Fr_1^2)^{1/2} - 3 \right\}$
4	Sarma - Newnham (1973)	1,21 – 3,79	$\frac{L_r}{h_1} = 6,73(Fr_1 - 1)$
5	Hager (1990)	2,5 – 8	$\frac{L_r}{h_1} = 8(Fr_1 - 1,5)$
		$Fr_1 > 8$	$\frac{L_r}{h_1} = -12 + 160 \tanh(Fr_1 / 20) : \frac{h_1}{b} < 0,1$
		$Fr_1 > 8$	$\frac{L_r}{h_1} = -12 + 100 \tanh(Fr_1 / 12,5) : 0,1 < \frac{h_1}{b} < 0,7$

Trong Bảng 1: b là chiều rộng kênh chữ nhật; h_1 và V_1 là độ sâu và vận tốc trước nước nhảy; số Fr_1 tính theo công thức (1) với g là gia tốc trọng trường.

$$Fr_1 = \frac{V_1}{\sqrt{gh_1}} \tag{1}$$

Theo Bảng 1, công thức Safranez và Hager có phạm vi áp dụng rộng, trong khi Sarma–Newnham và một số công thức khác bị giới hạn hơn. Kết quả của Safranez thường cao hơn

Pietrkowski dù cấu trúc tương tự. Điểm khác biệt của Hager là xét đến chiều rộng đáy kênh khi số Froude Fr_1 lớn hơn 8. Nói chung, các công thức thực nghiệm dễ áp dụng và khá tiện lợi, nhưng chúng chỉ phù hợp với một số phạm vi giá trị Fr_1 nhất định. Độ chính xác của chúng cũng có thể thay đổi tùy theo từng trường hợp tính toán.

Tập dữ liệu phục vụ kiểm định ANN, trong bài báo này có số Fr_1 tương tự Safranez và Hager, không phù hợp hoàn toàn cho Pietrkowski, Schroder và Sarma–Newnam. Do đó, kết quả nghiên cứu của Safranez và Hager đã được sử dụng làm cơ sở để đối chiếu với các sản phẩm từ những mô hình ANN.

2.2. Phương pháp phân tích thứ nguyên

Kết quả áp dụng các phương trình thủy động lực học cho đoạn dòng chảy trong khu vực nước nhảy (tham khảo Hình 1) cho thấy, chiều dài xoáy cuộn L_r là hàm số của các biến sau: vận tốc V_1 , độ sâu h_1 , chiều rộng b và độ nhám e của bề mặt kênh, hệ số nhớt μ và khối lượng riêng ρ của nước, gia tốc rơi tự do g , như phương trình (2).

$$L_r = f(V_1, h_1, b, \rho, \mu, e, g) \quad (2)$$

Định lý Π -Buckingham chỉ ra rằng có thể dùng năm hàm Π không thứ nguyên để thay cho tám biến số thuộc phương trình (2); các hàm này được liệt kê ở hệ phương trình (3).

$$\Pi_1 = L_r/h_1; \Pi_2 = Fr_1; \Pi_3 = Re_1^*; \Pi_4 = e/h_1; \Pi_5 = h_1/b \quad (3)$$

Trong đó: Re_1^* là số Reynold trước nước nhảy được hiệu chỉnh theo phương trình (4) với ν là hệ số nhớt động học của nước.

$$Re_1^* = \frac{V_1 h_1}{\nu} \quad (4)$$

Như vậy, chiều dài xoáy cuộn phụ thuộc vào nhiều yếu tố, như mô tả ở phương trình (5), với bốn biến đầu vào ($\Pi_2, \Pi_3, \Pi_4, \Pi_5$) và một biến đầu ra (Π_1). Việc xác định chính xác chiều dài này đòi hỏi xem xét tổng thể các tham số đặc trưng của dòng chảy.

$$\frac{L_r}{h_1} = \Phi \left(Fr_1, Re_1^*, \frac{e}{h_1}, \frac{h_1}{b} \right) \quad (5)$$

2.3. Phương pháp mô hình toán ứng dụng ML

2.3.1. Mô hình tăng cường phân loại (CB)

Thuật toán CB thuộc hệ ML tiên tiến, nằm trong phân hệ tăng cường độ dốc, nổi bật bởi năng lực xử lý các biến phân biệt trực tiếp mà không cần qua các thao tác mã hóa thủ công như one-hot hoặc biến đổi số hóa riêng biệt. CB cho phép phân giải các biến dạng phân loại mà không yêu cầu bước trung gian chuyển đổi, tiết kiệm thời gian và giảm rủi ro sai lệch trong bước tiền xử lý dữ liệu [15]. Ưu thế nổi bật của CB chính là khả năng tự động chuyển hóa các biến phân loại sang dạng số, dựa vào những phương thức mã hóa tân tiến như mã hóa trung bình cộng, giúp tinh gọn giai đoạn tiền xử lý dữ liệu và tiết kiệm đáng kể thời gian triển khai mô hình. Ngoài ra, CB còn tận dụng các kỹ thuật khắc phục tình trạng dữ liệu bị khuyết rất linh hoạt, góp phần nâng cao độ xác tín và gia tăng phạm vi ứng dụng của mô hình trong các vấn đề phức tạp. Mô hình dự báo của CB được diễn giải như sau:

$$\hat{y}(x) = \sum_{i=1}^K \alpha_i \cdot f_i(x) \quad (6)$$

Trong đó: K là số lượng cây; α_i là tốc độ học tập của cây thứ i ; x là giá trị của véc tơ đầu vào; f_i là hàm dự báo của cây thứ i .

2.3.2. Mô hình tăng cường độ dốc cực đại (XGB)

XGB nổi trội trong dự báo giá trị số nhờ khả năng vận hành song song và huấn luyện nhanh. Mô hình này là lựa chọn thay thế hiệu quả cho các mô hình cũ, giúp giải quyết quan hệ phi tuyến giữa biến dự báo và các thuộc tính đầu vào và tối ưu khả năng tận dụng tài nguyên [16].

XGB vận dụng thuật toán ML có giám sát, phối hợp đa mô hình giản hóa như cây quyết định dưới dạng chuỗi qua nhiều lần lặp. Mỗi cấu trúc mới được thiết lập nhằm triệt tiêu sai số từ mô hình trước thông qua cực tiểu hóa hàm tổn thất, sử dụng thuật toán gradient descent theo phương trình (7). Ưu điểm dễ nhận thấy của XGB là năng lực xử lý chuỗi số liệu phức tạp, kiểm soát tốt trạng thái quá khớp (overfitting) và hỗ trợ tính toán đa luồng, tối ưu hóa tốc độ xử lý cùng với độ tin cậy của dự báo.

$$\hat{y}(x) = \sum_{b=1}^K \lambda f_b(x) \quad (7)$$

Trong đó: x là thành phần vectơ nhập liệu; $f_b(x)$ đại diện cho hàm dự báo của mô hình b trong một chuỗi mô hình; λ là tham biến định lượng; K là số lượng các cây.

2.3.3. Mô hình mạng nơ-ron nhân tạo (ANN)

ANN là một phương pháp trong lĩnh vực ML, lấy cảm hứng từ cách hệ thần kinh sinh học truyền và xử lý thông tin. Một ANN điển hình bao gồm lớp đầu vào, một hoặc nhiều lớp ẩn và lớp đầu ra; mỗi lớp được cấu thành từ các nút (nơ-ron) kết nối với nhau bằng các giá trị trọng số và độ lệch (bias). Dữ liệu được lan truyền xuôi qua mạng, trong đó mỗi nơ-ron thực hiện phép tổng có trọng số và biến đổi thông qua hàm kích hoạt phi tuyến (ReLU, sigmoid, tanh), giúp mô hình học và biểu diễn các quan hệ phi tuyến phức tạp giữa các đầu vào và đầu ra của nó [17].

Quá trình huấn luyện ANN thông thường sử dụng thuật toán lan truyền ngược sai số kết hợp với các thuật toán tối ưu như gradient descent hoặc các biến thể của nó (Adam, RMSProp) nhằm cực tiểu hóa hàm mất mát. Nhờ khả năng xấp xỉ hàm phổ quát, ANN đặc biệt hiệu quả trong các bài toán hồi quy, phân loại, nhận dạng mẫu và dự báo số liệu theo thời gian thực [8]. ANN có năng lực tiếp thu đặc trưng của chuỗi số liệu phức tạp và phát hiện xu thế tiềm ẩn, góp phần thúc đẩy việc xử lý những bài toán học búa. Thành tựu của mô hình này được quyết định bởi kiến trúc mạng, mật độ nơ-ron, phương án chọn siêu tham số và chất lượng của chuỗi số liệu nhập vào. Bởi vậy, các biện pháp như regularization, early stopping và dropout thường xuyên được triển khai nhằm hạn chế hiện tượng quá khớp cũng như củng cố năng lực khái quát hóa của mô hình.

2.4. Phương pháp xác định sai số của mô hình

Trong công trình này, các chỉ tiêu như Sai số tuyệt đối bình quân (MAE), Sai số căn bậc hai bình quân (RMSE), Sai số phần trăm tuyệt đối bình quân (MAPE) và hệ số Nash-Sutcliffe (NSE) được sử dụng như những công cụ định lượng hiệu năng dự báo của các thuật toán ML. Các đại lượng này giúp đánh giá mức độ chính xác và khả năng khái quát hóa xu hướng dữ liệu cũng như tiềm năng ứng dụng của mô hình toán. MAE, RMSE, MAPE thấp và NSE gần 1 phản ánh mô hình dự báo dữ liệu hiệu quả. Các công thức tính toán các chỉ số này cùng khoảng giá trị tối ưu có trong Bảng 2.

Bảng 2. Các chỉ số đánh giá và công thức tính toán.

Chỉ số đánh giá	Công thức tính	Khoảng giá trị tối ưu
<i>NSE</i>	$1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$	$0,5 \leq NSE \leq 1,0$
<i>MAE</i>	$\frac{\sum_{i=1}^n y_i - \hat{y}_i }{n}$	$MAE \approx 0$
<i>RMSE</i>	$\sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$	$RMSE \approx 0$
<i>MAPE</i>	$\left(\frac{1}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{y_i} \right) \times 100\%$	$MAPE \approx 0$

Trong các công thức trên, y_i và \hat{y}_i là số liệu thực đo và số liệu dự báo thứ i ; \bar{y}_i là trị số bình quân của các giá trị thực đo; n là số lần đo đạc.

2.5. Dữ liệu thực nghiệm

Tập hợp dữ liệu dùng cho lớp vào ANN cũng như phục vụ kiểm định đã được tổng hợp từ các bài báo của Hager cùng đồng nghiệp [4], [18], và nghiên cứu của Bretz (1988) [7]. Các nghiên cứu này đã tiến hành năm chuỗi thí nghiệm trong những máng kính hình chữ nhật nằm ngang với bề rộng thay đổi từ 98 mm đến 500 mm. Đáy máng cùng tường bên trái được chế tạo từ nhựa, tường đối diện sử dụng kính trong suốt nhằm hỗ trợ việc quan sát hiện tượng nước nhảy một cách trực diện và chi tiết. Kết quả là độ nhám lòng kênh rất thấp, chỉ đạt khoảng 5×10^{-6} m [18]. Phòng thí nghiệm là môi trường kín, với nhiệt độ luôn được kiểm soát ổn định trong khoảng từ 16 đến 18 °C, khi đó nước có hệ số nhớt là $1,1 \times 10^{-6}$ m²/s. Các thiết kế mẫu thí nghiệm đều bám sát thực tế, phản ánh chính xác các đặc tính của nước nhảy như số Fr_1 , số Re_1^* và vận tốc V_1 , tương thích với các công trình thủy lợi tại Việt Nam. Bảng 3 nêu rõ kích thước máng và các tham số đặc trưng cho từng thí nghiệm; ở đó, số Fr_1 biến thiên trong khoảng từ 2,26 đến 15,96.

Bảng 3. Các thông số cơ bản của các thí nghiệm.

Chiều rộng b (mm)	h_1/b	e/h_1	$Re_1^* \times 10^{-5}$	Fr_1
500	0,012-0,078	0,0008-0,0001	0,09-0,95	3,33-15,31
98	0,1-0,65	0,0005-0,0001	0,09-2,00	2,45-10,87
500	0,055-0,109	0,0002-0,0001	0,52-3,12	2,26-11,12
500	0,01-0,072	0,001-0,0001	0,05-1,04	2,38-15,96
500	0,036-0,237	0,034-0,278	0,04-5,16	3,31-7,82

Sau khi loại bỏ sai số và nhiễu, dữ liệu đã được xử lý rồi đồng nhất hóa. Tổng số 156 bộ dữ liệu thực nghiệm đã được triển khai cho ANN. Thông thường, dữ liệu được phân chia theo các tỷ lệ 80/20, 75/25 hoặc 70/30 giữa huấn luyện và kiểm định. Nhằm nâng cao độ chuẩn xác và mở rộng khả năng ứng dụng mô hình, trong bài báo này, 109 mẫu đã được dùng cho huấn luyện (chiếm gần 70%) và 47 mẫu cho kiểm định (trên 30%). Nhờ vào các thuật toán ML, 47 bộ số liệu kiểm định đã được chọn lựa hoàn toàn ngẫu nhiên để dùng chung cho mọi mô hình. Việc này nhằm bảo đảm tính minh bạch, khách quan và giảm thiểu tối đa chủ quan khi chọn lựa.

3. ẢNH HƯỞNG CỦA CÁC BIẾN ĐẦU VÀO ĐẾN KẾT QUẢ DỰ BÁO

3.1. Tối ưu các siêu tham số mô hình CB và XGB

Siêu tham số trong thuật toán ML giữ vị trí trọng yếu, tác động trực tiếp đến mức độ chính xác của kết quả. Quá trình điều chỉnh các tham số này thường được tiến hành thông qua ba phương thức nổi bật: quét lưới tổ hợp (grid search), sàng lọc ngẫu nhiên (random search) và tối ưu hóa theo trường xác suất Bayes (Bayesian optimization).

Trong công trình này, kỹ thuật tối ưu hóa Bayes được lựa chọn để tìm kiếm bộ tham số tối ưu cho hai mô hình CB và XGB. Để đạt được sự ổn định và hạn chế hiện tượng quá khớp, quy trình kiểm định chéo (cross-validation) đã được áp dụng với dữ liệu được chia thành 10 nhóm, lặp lại 5 lần. Hai mô hình CB và XGB đều được kiểm thử nhiều lần nhằm tìm ra bộ tham số tối ưu, đồng thời phát triển mô hình hiệu quả. Bảng 4 thể hiện phạm vi và các giá trị của siêu tham số cho mỗi mô hình, trong đó các tham số có ảnh hưởng lớn nhất được sắp xếp ở hàng đầu tiên. Cấu trúc của ANN được chọn sơ bộ gồm hai lớp ẩn, mỗi lớp có 16 nơ ron. Chi tiết về siêu tham số của ANN có ở mục 4.

Bảng 4. Các siêu tham số chính của hai mô hình học máy.

TT	Mô hình	Tham số chính	Khoảng giá trị	Giá trị chọn
1	CB	n_estimators	100, 200, 300, 500, 800	5004
		depth	2, 3, 4, 5, 6	
2	XGB	learning_rate	0,1; 0,2; 0,3; 0,4	0,1
		max_depth	3, 4, 5, 6, 7, 8	8
		objective	Default	reg:squarederror

3.2. Đánh giá ảnh hưởng của các biến đầu vào

Cách tiếp cận thăm định mức độ quan trọng của thuộc tính thông qua cơ chế ML (vận dụng chỉ số gain importance) cung cấp chỉ số định lượng, giúp nhận diện rõ rệt từng biến đầu vào. Phương pháp này được xem là một giải pháp điển hình, hiện diện rộng rãi trong lĩnh vực khai phá dữ liệu và đã được lý giải cặn kẽ trong công trình của Gregorutti cùng các cộng sự năm 2017 [19]. Các mô hình ML có thể ước tính tầm quan trọng của các đầu vào thông qua điểm số, thể hiện mức độ ảnh hưởng của từng biến đến kết quả dự báo. Biến nào có điểm số cao hơn sẽ có ảnh hưởng lớn hơn tới mô hình. Cách tiếp cận này giúp quá trình đánh giá trở nên khách quan và đồng nhất hơn.

Kết quả đánh giá độ nhạy của từng biến đầu vào được trình bày trong Bảng 5. Các mô hình CB, XGB và ANN đã đưa ra điểm số cho từng biến, phản ánh sự nhất quán trong dự báo. Phân tích số liệu từ ba mô hình cho thấy: số Fr_1 có ảnh hưởng lớn nhất tới chiều dài xoáy cuộn, hơn hẳn các đặc trưng khác. Tiếp theo, độ nhám là yếu tố có tác động lớn thứ hai. Ngược lại, số Re_1^* và tỷ số h_1/b chỉ có tác động nhỏ, thể hiện vai trò hạn chế trong việc dự báo. Kết quả này phản ánh đúng thực tế: nước nhảy thường mang khối lượng nước lớn nhưng chỉ diễn ra trên một đoạn kênh ngắn, do đó ảnh hưởng của trọng lực là chủ đạo, còn lực cản chỉ đóng vai trò thứ yếu. Theo nghiên cứu của Hager [4], [18], nước nhảy xảy ra ở khu thành nhám thủy lực, nên lực cản phụ thuộc vào độ nhám và số Reynolds.

Kết quả so sánh các mô hình (Bảng 5) chỉ ra rằng, các mô hình dựa trên cây quyết định (CB và XGB) thể hiện ưu thế nổi bật so với mô hình ANN. Cả CB và XGB đều xác định rằng biến Fr_1 đóng vai trò chi phối hàng đầu, phản ánh đúng đặc điểm và bản chất vật lý của nước nhảy. Ngược lại, mô hình ANN phân bố mức độ quan trọng tương đối đồng đều cho các biến, làm giảm khả năng nhận diện biến chủ đạo. Điều này thể hiện rằng, mô hình cây quyết định đạt ưu điểm hơn so với ANN trong việc chọn lọc đặc trưng.

Bảng 5. Điểm mức độ quan trọng của các biến đầu vào.

TT	Mô hình	Fr_1	e/h_1	Re_1^*	h_1/b
1	CB	0,8386	0,0611	0,0560	0,0442
2	XGB	0,9543	0,0245	0,0115	0,0097
3	ANN	0,2798	0,2710	0,2305	0,2186

4. CẤU TRÚC CỦA MẠNG NƠ-RON NHÂN TẠO

Cho đến nay có nhiều loại ANN với cấu hình khác nhau được sử dụng hiệu quả trong dự báo kỹ thuật, trong đó multilayer perceptron (MLP) là một ANN phổ biến nhất [9]. MLP học từ dữ liệu thông qua một lớp vào gồm nhiều nơ-ron, mỗi nơ-ron tương thích với một đầu vào. Sau khi nhập số liệu, việc xử lý xảy ra tại các lớp ẩn, số lượng lớp ẩn có thể tăng hoặc giảm phụ thuộc mức độ phức tạp của bài toán. Kết quả giải quyết tiếp tục được chuyển đến lớp ra, nơi các nơ-ron đại diện cho số lượng kết quả mong muốn. Để huấn luyện mô hình MLP, cần ứng dụng đa dạng các thuật toán tối ưu tùy theo các mục tiêu phân tích.

Công trình này đã xây dựng hai mô hình mạng MLP với số lớp ẩn khác biệt (một hoặc hai lớp ẩn), nhằm đánh giá và so sánh hiệu năng của từng mô hình. Mạng 1 gồm ba lớp: một lớp vào, một lớp ẩn và một lớp ra. Mạng 2 gồm bốn lớp: một lớp vào, hai lớp ẩn và một lớp ra. Bốn nơ-ron ở lớp vào biểu diễn cho bốn hàm Π đầu vào. Lớp ra chỉ có một nơ-ron, đại diện cho một hàm Π đầu ra. Hàm kích hoạt ở lớp ẩn giúp điều chỉnh các tín hiệu đầu vào truyền tới từng nơ-ron, góp phần tăng cường hiệu năng xử lý thông tin của cả mạng lưới. Số lượng nơ-ron trong mỗi lớp ẩn được giữ nhất quán, dao động từ 8 đến 64 theo cấp số nhân của hai. Số lượng nơ-ron cần thiết cho từng lớp ẩn được quyết định nhờ việc kiểm thử liên tục. Vì dữ liệu đầu ra luôn dương, hàm kích hoạt tại lớp ẩn được chọn là hàm phi tuyến mượt ReLU, trong khi đó, lớp ra ứng dụng hàm tuyến tính. Hàm tổn thất là Bình quân sai số bình phương (MSE), thường áp dụng cho các nghiên cứu thủy văn – thủy lực [20]. Thuật toán tối ưu Adam được lựa chọn nhờ hiệu quả cao và tính phổ biến hiện nay. Số vòng lặp (epochs) tối đa được đặt là 10.000. Tuy nhiên, việc huấn luyện sẽ được kết thúc sớm nếu hàm tổn thất trên tập kiểm định không có sự cải thiện sau 200 vòng lặp liên tiếp (kỹ thuật early stopping nhằm tránh hiện tượng quá khớp). Thuật toán Adam dùng tốc độ học mặc định 0,001 cho mọi mô hình. Bên cạnh đó, các kỹ thuật nâng cao như regularization, model checkpoint (lưu tham số mô hình) cũng được áp dụng giúp lưu lại các thông tin trọng yếu khi huấn luyện. Dữ liệu đầu vào được chuẩn hóa sau khi tiền xử lý. Hình 2 minh họa kiến trúc của Mạng 1 với cấu trúc ba lớp có số nơ-ron là 4-8-1.

5. KẾT QUẢ NGHIÊN CỨU VÀ THẢO LUẬN

5.1. Kết quả kiểm định mô hình

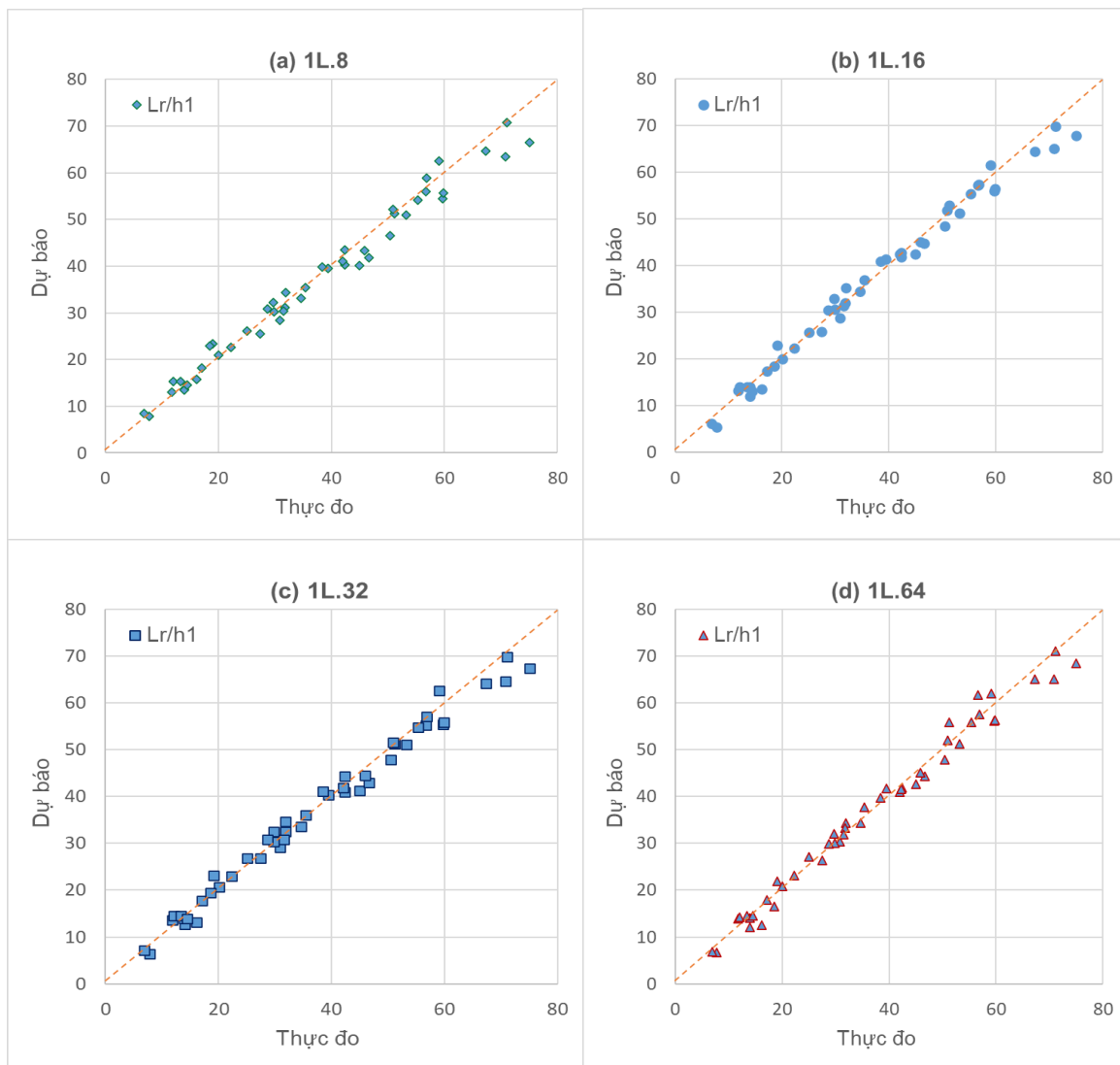
Khi quá trình huấn luyện mô hình hoàn tất, chỉ số L_r/h_1 sẽ được dùng để kiểm định kết quả dự báo. Bảng 6 trình bày hiệu quả dự báo của các mô hình dựa trên bốn thông số đánh giá đã có trong mục 2.4. Kết quả kiểm định chỉ ra rằng hai mô hình 2L.16 và 2L.32 đạt giá trị NSE trên 0,99, bốn mô hình khác có NSE lớn hơn 0,98 và hai mô hình còn lại đạt 0,977. Những số liệu này thể hiện rằng, mô hình dự báo đạt mức độ chính xác cao khi đối chiếu với dữ liệu thực tế, đồng thời nhấn mạnh độ khả tín cao và tính bền vững của mô hình.

Xét về sai số tương đối, hai mô hình 2L.16 và 2L.32 cho kết quả MAPE thấp nhất, đều dưới 5%, trong khi những mô hình khác có MAPE dao động trong khoảng từ 5,7% đến 6,6%.

Các chỉ số MAE, RMSE thể hiện xu hướng biến động tương tự và phù hợp với hệ số NSE, tức là mô hình có sai số càng nhỏ thì NSE càng lớn. Đáng chú ý, mô hình 2L.16 đạt được cả giá trị MAE và RMSE thấp nhất cùng NSE cao nhất, khẳng định hiệu quả dự báo tốt nhất.

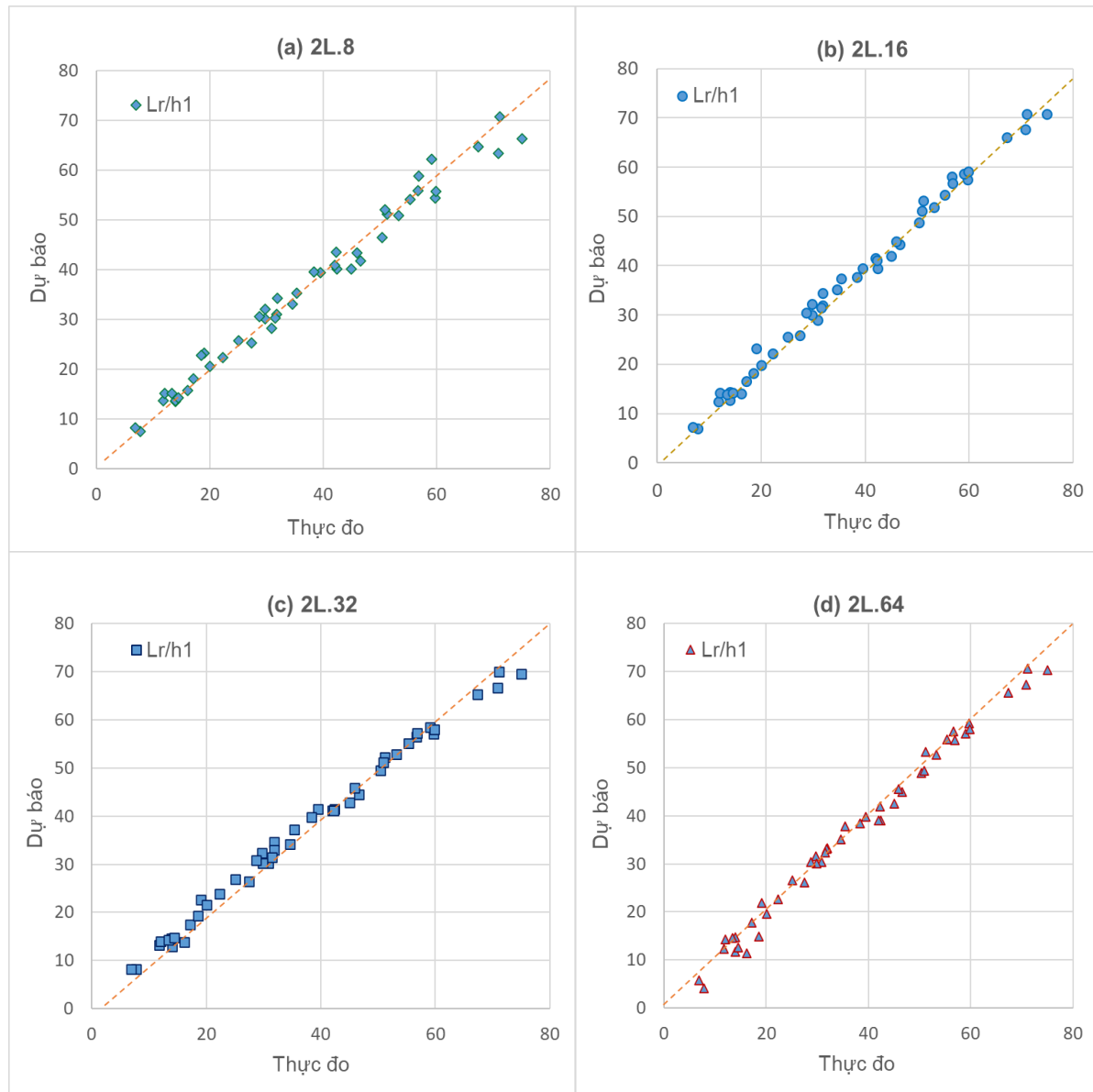
Bảng 6. Các tham số, kết quả kiểm định các mô hình ANN và công thức thực nghiệm.

TT	Mô hình	Số lớp ẩn	Số nơ-ron	Số lần lặp	MAE	RMSE	NSE	MAPE
1	1L.8	1	8	8000	2,074	2,799	0,977	6,54%
2	1L.16	1	16	5110	1,666	2,224	0,986	5,69%
3	1L.32	1	32	5000	1,663	2,095	0,987	5,79%
4	1L.64	1	64	5000	1,828	2,351	0,984	5,87%
5	2L.8	2	8+8	4595	2,059	2,801	0,977	6,49%
6	2L.16	2	16+16	4410	1,261	1,659	0,992	4,32%
7	2L.32	2	32+32	3860	1,372	1,758	0,991	4,98%
8	2L.64	2	64+64	3183	1,583	1,965	0,989	6,61%
9	Công thức		Safranez		3,678	4,302	0,975	17,44%
10			Hager		2,103	2,740	0,980	7,04%



Hình 3. So sánh thực đo và tính toán khi kiểm định bốn mô hình một lớp ẩn.

Hình 3 minh họa mối liên hệ giữa số liệu thực nghiệm và dự báo từ bốn ANN một lớp ẩn của Mạng 1. Phần lớn các điểm quy tụ sát tuyến đẳng trị 45° , cho thấy sự tương liên tuyến tính có cường độ trội kèm theo biên độ sai lệch tối thiểu. Ngoại trừ mô hình 1L.8, những mô hình còn lại đều trình bày năng lực dự báo tốt. Đặc biệt, 1L.32 đạt hiệu suất cao nhất trong nhóm mô hình một lớp ẩn, chứng minh năng lực khái quát hóa vững chắc trên tập kiểm định.

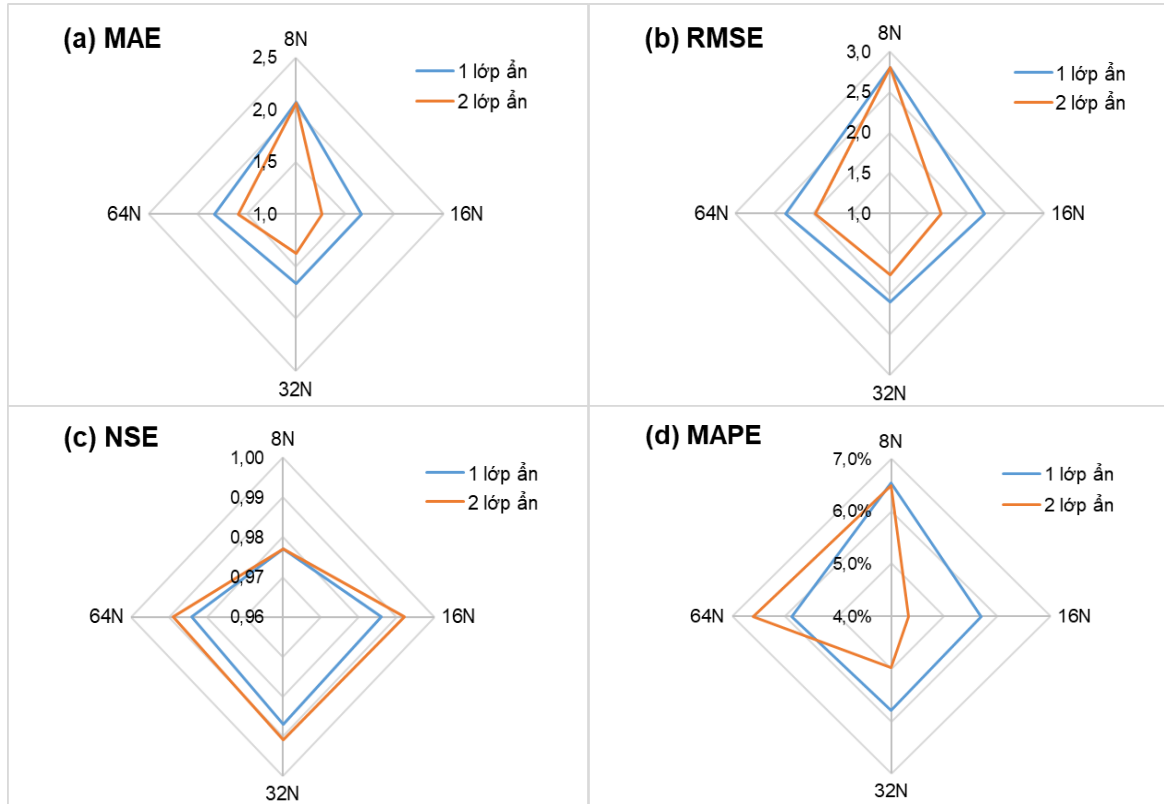


Hình 4. So sánh số liệu thí nghiệm với dự báo khi kiểm định bốn mô hình hai lớp ẩn.

Hình 4 phác họa sự đối chiếu giữa số liệu thực tế và kết quả ước tính từ bốn ANN hai lớp ẩn thuộc Mạng 2. Nhìn chung, các điểm dữ liệu hội tụ cận kề tuyến đẳng trị 45° , minh chứng cho mối liên hệ tuyến tính rõ nét và mức sai lệch thấp. Trong nhóm này, mô hình 2L.16 nổi bật nhất về độ chuẩn xác và hiệu năng dự báo, tiếp theo là 2L.32, hiệu quả hơn các ANN còn lại.

5.2. So sánh các mô hình một lớp ẩn với hai lớp ẩn

Nhằm phân tích so sánh giữa ANN một lớp ẩn với hai lớp ẩn, Hình 5 trình bày kết quả thẩm định hiệu năng của tám mô hình khác nhau. Các biểu đồ thể hiện rằng, các sai số như MAE và RMSE của nhóm mô hình một lớp ẩn đều lớn hơn so với nhóm hai lớp ẩn, trong khi hệ số NSE lại thấp hơn. Giá trị của MAPE cũng thể hiện xu hướng tương tự MAE và RMSE, ngoại trừ mô hình 1L.64 có giá trị MAPE thấp hơn 2L.64. Từ kết quả so sánh có thể suy luận rằng ANN với hai lớp ẩn sở hữu năng lực khái quát hóa và dự báo tốt hơn ANN một lớp ẩn.

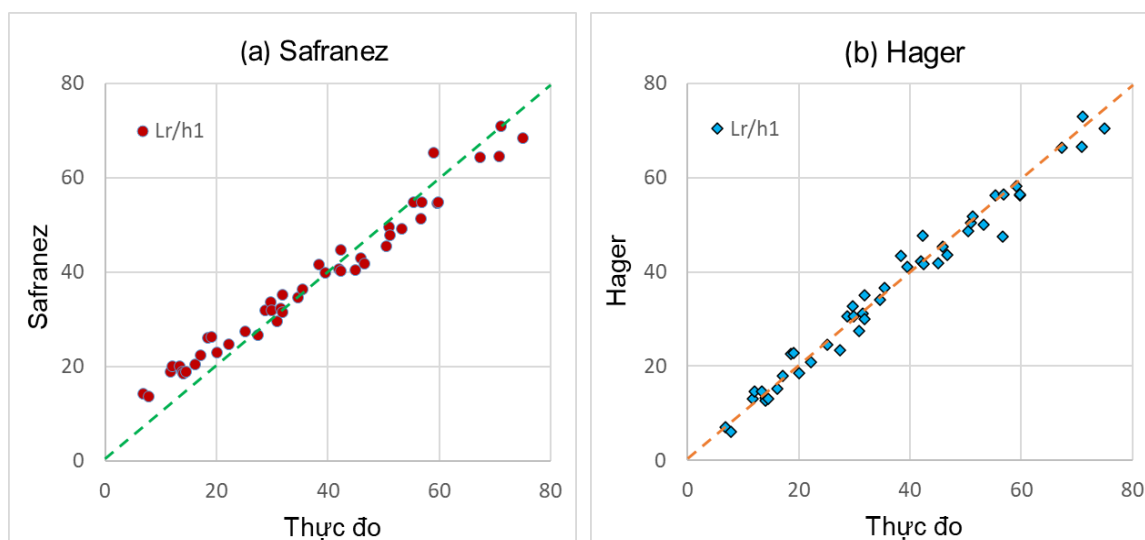


Hình 5. So sánh kết quả dự báo của các mô hình một lớp ẩn và hai lớp ẩn.

5.3. Kết quả so sánh các mô hình ANN với các công thức thực nghiệm

Phân tích dưới đây nhằm đối chiếu khả năng ước tính chiều dài xoáy cuộn giữa các mô hình ANN với hai phương trình thực nghiệm thông dụng. Hình 6 trình bày so sánh giữa số liệu thực đo và tính toán từ hai công thức nêu trên, cho thấy cả hai phương trình này đều đạt độ chính xác khá cao. Hơn thế nữa, kết quả so sánh dữ liệu trong Bảng 6 chứng tỏ công thức Hager đạt độ chuẩn xác cao hơn công thức Safranez ở tất cả các tiêu chí. Đáng chú ý, giá trị MAPE của công thức Hager thấp hơn rất nhiều, qua đó thể hiện khả năng dự báo hiệu quả hơn của công thức này khi tính toán chiều dài xoáy cuộn.

Số liệu ở Bảng 6 cũng chỉ ra rằng độ chính xác của công thức Hager tương đương với hai mô hình 1L.8 và 2L.8, nhưng vẫn thấp hơn sáu ANN khác. Tổng quát chung, các mô hình ANN có năng lực dự báo tốt hơn hẳn các phương trình thực nghiệm, biểu thị qua giá trị sai số nhỏ hơn và chỉ số NSE lớn hơn. Cụ thể, công thức Hager đạt NSE 0,98 và MAPE 7,04%, trong khi các mô hình ANN như 2L.16 và 2L.32 đạt NSE trên 0,99 và MAPE dưới 5%, thể hiện sự tương quan rất mạnh giữa dự báo với thực đo. Kết quả của các mô hình luôn ổn định qua nhiều lần chạy, đảm bảo tính khách quan trong đánh giá.



Hình 6. So sánh giá trị thực đo với kết quả tính toán kiểm định công thức.

5.4. Thảo luận

Mô hình 2L.16 đạt sai số rất thấp so với các phương trình thực nghiệm (tham khảo Hình 7), qua đó chứng minh ANN là lựa chọn đầy tiềm năng để tính toán thủy lực nhờ ưu điểm về tính chuẩn xác và ổn định cao cùng với năng lực khái quát hóa dữ liệu tốt.

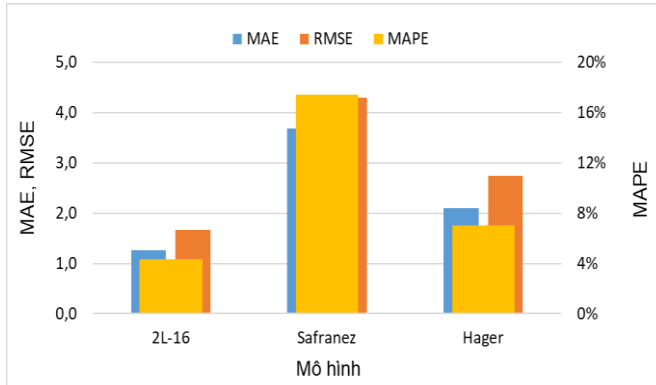
Xét về tốc độ, các mô hình hai lớp ẩn thường có thời gian tính toán ngắn hơn, cần số vòng lặp huấn luyện ít hơn so với các mô hình chỉ có một lớp ẩn, đồng thời hiệu suất xử lý dữ liệu tốt hơn. Cụ thể, tổng thời gian chạy của các ANN hai lớp ẩn chỉ dao động trong khoảng 60–90 giây, trong khi các mô hình một lớp ẩn lại kéo dài tới 100–160 giây cho mỗi lần dự báo. Số vòng lặp cũng như thời gian tính có chiều hướng giảm khi số nơ-ron trong lớp ẩn tăng (xem Hình 8). Như vậy, các mô hình hai lớp ẩn có hiệu năng dự báo tốt hơn các ANN một lớp ẩn.

Tuy nhiên, kết quả kiểm định chỉ ra rằng, khi số nơ-ron trong mỗi lớp ẩn vượt quá 32, mô hình trở nên phức tạp hơn nhưng độ chính xác không hề được cải thiện. Ngược lại, mô hình 2L.64 có sai số MAPE cao hơn so với các mô hình có số nơ-ron ít hơn. Điều này thể hiện rằng mô hình có thể bị quá khớp do mức độ phức tạp vượt quá so với quy mô của bộ dữ liệu (156 mẫu). Khi số lượng nơ-ron tăng lên, ANN trở nên nhạy cảm hơn với nhiễu và các biến động nhỏ trong tập dữ liệu huấn luyện, dẫn đến giảm hiệu suất tổng quát hóa trên các dữ liệu mà mô hình chưa từng gặp trước đó. So với mô hình 2L.64, 2L.16 cân bằng tốt hơn giữa độ phức tạp và khả năng tổng quát hóa, từ đó đạt hiệu suất dự báo cao hơn.

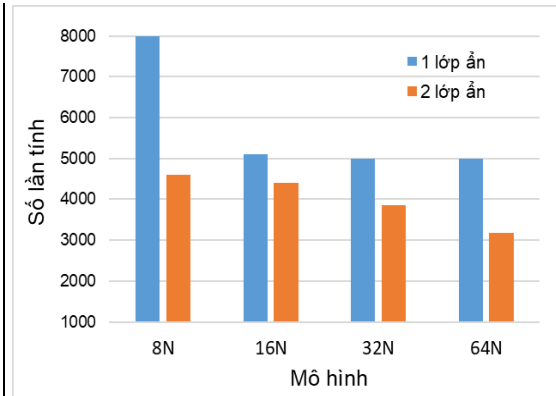
Như vậy, công trình này đã tìm ra cấu hình ưu trội của mô hình là ANN với hai lớp ẩn, mỗi lớp có 16 nơ-ron. Mặc dù vậy, các mô hình vẫn duy trì hiệu quả hoạt động khi lượng nơ-ron trong mỗi lớp ẩn dao động từ 16 đến 32. Do đó, khi lựa chọn cấu trúc mô hình, cần cân nhắc kỹ dựa trên yêu cầu của từng bài toán để đạt được sự cân bằng tốt nhất giữa mức độ chính xác và chi phí tính toán.

Nghiên cứu này tiếp tục khẳng định những ưu điểm nổi bật của các mô hình học máy, đặc biệt là ANN, bao gồm mức độ chuẩn xác cao, năng lực tích hợp đồng thời nhiều biến đầu vào, tính linh hoạt trong cập nhật khi được bổ sung số liệu, thời gian tính toán ngắn và tiềm năng ứng dụng rộng rãi khi dự báo nước nhảy trong kênh mặt cắt hình chữ nhật đáy bằng. Trong đó, mức độ chuẩn xác và năng lực cập nhật thích ứng được xem là hai lợi thế đáng kể

so với các phương trình thực nghiệm truyền thống. Trong thời gian tới, khi lượng dữ liệu thực nghiệm được bổ sung dồi dào hơn, các mô hình có thể được tái huấn luyện nhằm gia tăng năng lực và cải thiện hiệu quả dự báo.



Hình 7. So sánh mô hình 2L.16 với các công thức.



Hình 8. Số lần lặp của các mô hình ANN.

Mặc dù vậy, mô hình ANN cũng tồn tại một vài nhược điểm nhất định. Quá trình huấn luyện thường kéo dài hơn các mô hình ML khác. Sản phẩm đầu ra giữa các lần chạy có thể không đồng nhất do tính ngẫu nhiên trong khởi tạo và tối ưu hóa. Hiệu năng của mô hình cũng phụ thuộc nhiều vào chất lượng chuỗi số liệu đầu vào cũng như khả năng lựa chọn siêu tham số của người vận hành. Ngoài ra, sai số dự báo của ANN có thể thay đổi thất thường, nên cần tiến hành huấn luyện mô hình nhiều lần nhằm đạt kết quả tốt nhất.

6. KẾT LUẬN

Nghiên cứu này đã kiểm định độ chính xác trong dự báo chiều dài xoáy cuộn của tám mô hình ANN. Mục tiêu chính là xây dựng, phân tích và đánh giá hiệu năng của các mô hình dựa trên hai công thức thực nghiệm, nhằm khám phá cấu trúc tối ưu cho ANN. Các kết quả đã được phân tích cả về định tính cũng như định lượng nhằm làm sáng tỏ hiệu năng dự báo của từng mô hình. Những kết luận chính được tóm tắt như sau:

- Với các bộ dữ liệu nhỏ, việc chọn ngẫu nhiên 30% số liệu làm tập kiểm định là phương pháp hữu ích giúp tăng cường mức độ chuẩn xác và đảm bảo của kết quả dự báo.
- Năng lực dự báo của mô hình ANN lệ thuộc vào cấu trúc mạng và việc điều chỉnh siêu tham số. Mô hình XGB và XB thẩm định mức độ ảnh hưởng của các đặc trưng thủy lực tốt hơn ANN.
- Số Froude trước nước nhảy có ảnh hưởng mạnh nhất tới kết quả dự báo, tiếp đến là độ nhám tương đối và số Reynolds; độ sâu tương đối trước nước nhảy ít tác động nhất.
- Cấu hình ANN tối ưu gồm hai lớp ẩn với 16 nơ-ron/lớp; tăng lên 32 nơ-ron/lớp vẫn hiệu quả, nhưng lên 64 nơ-ron sai số tăng khi bộ số liệu nhỏ.

Kết quả kiểm định chỉ ra rằng, khi số Fr_1 dao động từ 2,26 đến 15,96, mô hình ANN hai lớp ẩn với 16 nơ-ron mỗi lớp là lựa chọn tốt nhất để dự báo chiều dài xoáy cuộn của nước nhảy trong lòng dẫn hình chữ nhật đáy bằng.

TÀI LIỆU THAM KHẢO

- [1]. W. H. Hager, Energy Dissipators and Hydraulic Jump, Volume 8. Dordrecht, The Netherlands: Kluwer Academic Publishers, Water Science and Technology Library, 1992.
- [2]. W. H. Hager, Hydraulic jump, Energy dissipators, (1995) 43–60.

<https://doi.org/10.1201/9780203757512-4>.

- [3]. K. V. N. Sarma and D. A. Newnham, Surface profiles of hydraulic jump for Froude numbers less than four, *Water Power*, 25 (1973) 139142.
- [4]. W. H. Hager, R. Bremen, and N. Kawagoshi, Length of roller Recess hydraulique classique: Longueur du rouleau, *J. Hydraul. Res.*, 28 (1990) 591–608. <https://doi.org/10.1080/00221689009499048>.
- [5]. H. E. Schulz, A. L. A. Simões, and J. D. Nóbrega, Roller lengths, sequent depths, surface profiles for pre-design of dissipation basins, *Iwhs* 2015 (2015) 81. <https://doi.org/10.13140/RG.2.1.3027.0889>.
- [6]. A. J. Peterka, *Hydraulic Design of Stilling Basins and Energy Dissipators*, Monogr. E, Ed. A water Resour. Tech. Publ. USBR, 25 (1984) 240.
- [7]. N. V. Bretz, *Hydraulic Jump Forced by Sill*. Ecole Polytechnique Fédérale de Lausanne, 1988.
- [8]. M. Naseri and F. Othman, Determination of the length of hydraulic jumps using artificial neural networks, *Adv. Eng. Softw.*, 48 (2012) 27–31. <https://doi.org/10.1016/j.advengsoft.2012.01.003>.
- [9]. L. Houichi, N. Dechemi, S. Heddami, and B. Achour, An evaluation of ANN methods for estimating the lengths of hydraulic jumps in U-shaped channel, *J. Hydroinformatics*, 15 (2013) 147–154. <https://doi.org/10.2166/hydro.2012.138>.
- [10]. S. Baharvand, A. Jozaghi, R. Fatahi-Alkouhi, S. Karimzadeh, R. Nasiri, and B. Lashkar-Ara, Comparative Study on the Machine Learning and Regression-Based Approaches to Predict the Hydraulic Jump Sequent Depth Ratio, *Iran. J. Sci. Technol. - Trans. Civ. Eng.*, 45 (2021) 2719–2732. <https://doi.org/10.1007/s40996-020-00526-2>.
- [11]. N. M. Ngọc. Lê Văn Nghi, Ứng dụng thuật toán SVM dự báo chiều dài nước nhảy trên kênh hình thang cân, *Tạp chí khoa học và công nghệ thủy lợi*, vol. số chuyên đề, (2023) 43–49.
- [12]. H. V. Hùng, Ứng dụng phương pháp học máy tính toán chiều dài nước nhảy trong kênh lăng trụ mặt cắt hình chữ nhật, *Tạp chí khoa học và công nghệ thủy lợi*, 84 (2024) 76–86.
- [13]. H. V. Ho, Determination of the surface roller length of hydraulic jumps in horizontal rectangular channels using the machine learning method, *Stoch. Environ. Res. Risk Assess.*, 38 (2024) 2539–2562. <https://doi.org/10.1007/s00477-024-02697-0>.
- [14]. H. V. Hùng, Ứng dụng các thuật toán học máy xác định độ sâu sau nước nhảy trong kênh chữ nhật có xét đến ảnh hưởng của lực ma sát, *Tạp Chí Khoa Học Giao Thông Vận Tải*, 76 (2025) 172–186. <https://doi.org/10.47869/tcsj.76.2.5>.
- [15]. L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, Catboost: Unbiased boosting with categorical features, *Adv. Neural Inf. Process. Syst.*, 2018 (2018) 6638–6648.
- [16]. T. Chen and C. Guestrin, XGBoost: A scalable tree boosting system, *Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, 13-17-Aug (2016) 785–794.
- [17]. M. F. Sauda, Prediction of hydraulic jump length downstream of multi-vent regulators using Artificial Neural Networks, *Ain Shams Eng. J.*, 7 (2016) 819–826.
- [18]. W. H. Hager and R. Bremen, Sequent depths: Le ressaut hydraulique classique: étude des hauteurs conjuguées, *J. Hydraul. Res.*, 27 (1989) 565–585. <https://doi.org/10.1080/00221688909499111>.
- [19]. B. Gregorutti, B. Michel, and P. Saint-Pierre, Correlation and variable importance in random forests, *Stat. Comput.*, 27 (2017) 659–678. <https://doi.org/10.1007/s11222-016-9646-1>.
- [20]. H. V. Ho, Calculation of the hydraulic jump length in horizontal rectangular channels using artificial neural network models, *J. Water Resour. Environ. Eng.*, 92 (2024) 54–61.