



APPLICATION OF MACHINE LEARNING FOR PREDICTING ADVERSE THERMAL GRADIENTS IN REINFORCED CONCRETE BOX-GIRDER BRIDGES

Canh Minh Le*, Chuyen Duc Nguyen, Sang Thanh Le

Campus in Ho Chi Minh City, University of Transport and Communications, No. 450-451 Le Van Viet Street, Tang Nhon Phu Ward, Ho Chi Minh City, Vietnam

ARTICLE INFO

TYPE: Research Article

Received: 02/06/2025

Revised: 26/09/2025

Accepted: 13/10/2025

Published online: 15/10/2025

<https://doi.org/10.47869/tcsj.76.8.7>

* *Corresponding author*

Email: canhlm_ph@utc.edu.vn

Abstract. During the operational phase, reinforced concrete (RC) box-girder bridges are significantly influenced by environmental temperature variations. Temperature differentials among different parts of the cross-section generate considerable thermal stresses, reducing structural durability and lifespan. Recent studies highlight substantial discrepancies between vertical thermal gradients measured in the field and the standard values prescribed by the Vietnamese standard TCVN 11823:2017. This study analyzes nearly three years of continuous temperature measurement data from three representative bridges located in distinct climatic regions of Vietnam: Vinh Thinh Bridge (Northern region), Quan Hau Bridge (Central region), and Go Gang Bridge (Southern region). To enhance the accuracy of thermal effect assessments, supervised machine learning models including Linear Regression, Boosted Tree, Random Forest, XGBoost, and Artificial Neural Network (ANN) were developed to predict critical thermal gradient values. Among these models, the ANN demonstrated superior performance, achieving coefficient of determination (R^2) above 0,70 for both MAX-Gy and MIN-Gy. This result indicates strong predictive capabilities and effectiveness in capturing nonlinear relationships between meteorological conditions and the geometric characteristics of bridge cross-sections. The outcomes provide a robust scientific foundation for the design, evaluation, and maintenance of RC box-girder bridges in tropical climates, ultimately contributing to increased durability and operational efficiency.

Keywords: Thermal load, reinforced concrete box-girder, machine learning, temperature gradient, prediction thermal gradient

@2025 University of Transport and Communications



ỨNG DỤNG “MACHINE LEARNING” DỰ ĐOÁN GRADIENT NHIỆT BẤT LỢI TRÊN DẦM CẦU BTCT MẶT CẮT DẠNG HỘP

Lê Minh Cảnh*, Nguyễn Đức Chuyên, Lê Thành Sang

Phân hiệu tại Thành phố Hồ Chí Minh, Trường Đại học Giao thông vận tải, Số 450-451 Lê Văn Việt, Phường Tăng Nhơn Phú, Thành phố Hồ Chí Minh, Việt Nam

THÔNG TIN BÀI BÁO

CHUYÊN MỤC: Công trình khoa học

Ngày nhận bài: 02/06/2025

Ngày nhận bài sửa: 26/09/2025

Ngày chấp nhận đăng: 13/10/2025

Ngày xuất bản Online: 15/10/2025

<https://doi.org/10.47869/tcsj.76.8.7>

* Tác giả liên hệ

Email: canhlm_ph@utc.edu.vn

Tóm tắt. Trong giai đoạn khai thác, dầm cầu bê tông cốt thép (BTCT) mặt cắt dạng hộp chịu tác động đáng kể từ sự thay đổi nhiệt độ môi trường. Chênh lệch nhiệt độ giữa các phần của mặt cắt gây ra ứng suất nhiệt lớn, làm giảm độ bền và tuổi thọ của kết cấu. Các nghiên cứu gần đây chỉ ra rằng gradient nhiệt theo phương đứng xác định từ dữ liệu thực đo có sự sai lệch rõ rệt so với tiêu chuẩn TCVN 11823:2017. Nghiên cứu này phân tích dữ liệu nhiệt độ liên tục trong gần ba năm tại ba cầu đại diện cho các khu vực khí hậu tiêu biểu ở Việt Nam: cầu Vĩnh Thịnh (miền Bắc), cầu Quán Hàu (miền Trung) và cầu Gò Găng (miền Nam). Để nâng cao độ chính xác trong đánh giá tác động nhiệt, nghiên cứu đã phát triển các mô hình học máy có giám sát để dự báo gradient nhiệt bất lợi. Các thuật toán được áp dụng gồm: hồi quy tuyến tính (Linear Regression), cây tăng cường (Boosted Tree), rừng ngẫu nhiên (Random Forest), XGBoost và mạng nơ-ron nhân tạo (Artificial Neural Network – ANN). Kết quả cho thấy mô hình ANN đạt hiệu quả cao nhất với hệ số xác định R^2 trên 0,70 cho cả hai thông số MAX-Gy và MIN-Gy. Điều này khẳng định khả năng dự báo mạnh mẽ và hiệu quả trong việc mô hình hóa các mối quan hệ phi tuyến giữa điều kiện khí tượng và đặc trưng hình học tiết diện cầu. Kết quả của nghiên cứu cung cấp cơ sở khoa học tin cậy phục vụ công tác thiết kế, đánh giá và bảo trì cầu BTCT mặt cắt dạng hộp trong điều kiện khí hậu nhiệt đới, góp phần gia tăng tuổi thọ và hiệu quả khai thác công trình.

Từ khóa: Tải trọng nhiệt, dầm bê tông cốt thép mặt cắt dạng hộp, machine learning, gradient nhiệt, dự đoán gradient nhiệt

@ 2025 Trường Đại học Giao thông vận tải

1. ĐẶT VẤN ĐỀ

Cầu dầm hộp BTCT được áp dụng phổ biến nhờ tính ổn định, khả năng chịu tải lớn và hình thức kiến trúc hợp lý. Tuy nhiên, trong quá trình sử dụng, sự thay đổi nhiệt độ môi trường có thể tạo nên phân bố nhiệt không đồng đều trong tiết diện, dẫn đến ứng suất nhiệt và suy giảm độ bền kết cấu theo thời gian.

Tác động nhiệt trên kết cấu dầm hộp bao gồm hai dạng: phân bố đều, gây biến dạng dọc, và phân bố không đều, đặc trưng bởi gradient nhiệt, gây ra ứng suất uốn và biến dạng cục bộ. Gradient nhiệt được chia thành hai loại chính: tuyến tính và phi tuyến. Gradient tuyến tính biểu thị sự thay đổi nhiệt độ theo đường thẳng qua chiều cao hoặc chiều ngang mặt cắt, tác động trực tiếp đến ứng xử cục bộ và tổng thể kết cấu. Gradient phi tuyến phản ánh sự phân bố nhiệt dạng đường cong do sự dẫn nhiệt không đồng đều giữa lớp bê tông bề mặt và bên trong, thường gây ứng suất cục bộ cao, dẫn đến hiện tượng nứt bề mặt và giảm độ bền kết cấu.

Theo Elbadry và cộng sự [1] và Imbsen và cộng sự [2], sự chênh lệch nhiệt độ qua bề dày dầm hộp có thể tạo nên ứng suất kéo cục bộ, dẫn đến hình thành và mở rộng các vết nứt trên bản mặt và vách dầm của cầu BTCT. Một số báo cáo thực nghiệm ở châu Âu ghi nhận hiện tượng nứt rộng đến vài milimét trên cầu dầm hộp do tác động phối hợp giữa phân bố nhiệt không đều và bố trí cốt thép chưa hợp lý. Tương tự, các cầu tại Colorado (Mỹ) cũng xuất hiện nhiều vết nứt ở bản nắp trên, nắp dưới và vách dầm do tác động của gradient nhiệt theo phương đứng và phương ngang.

Các nghiên cứu gần đây tiếp tục khẳng định tác động đáng kể của gradient nhiệt đối với kết cấu cầu bê tông cốt thép có mặt cắt dạng hộp. Gu và cộng sự [3] ghi nhận rằng ứng suất sinh ra do gradient nhiệt có thể đạt tới 4–5 MPa, kèm theo chuyển vị dọc nhịp khoảng 10 mm. Shen và cộng sự [4] cũng ghi nhận rằng ứng suất kéo do gradient nhiệt ngang dao động trong khoảng 4–8 MPa, vượt quá cường độ chịu kéo của bê tông.

Tiêu chuẩn TCVN 11823:2017 [5], tham chiếu AASHTO LRFD (2007), hiện mới quy định thành phần gradient đứng và chưa có các hướng dẫn định lượng cho gradient ngang trong điều kiện khí hậu nhiệt đới, do đó cần nghiên cứu kiểm chứng bổ sung bằng số liệu các khu vực để hiệu chỉnh phù hợp với Việt Nam. Ngoài ra, tiêu chuẩn cho phép sử dụng nhiệt độ thực đo tại vị trí công trình để điều chỉnh gradient nhiệt độ trong tính toán, giúp quá trình thiết kế công trình hiệu quả hơn nhưng chưa có hướng dẫn cụ thể. Do đó, việc nghiên cứu có kiểm chứng thực nghiệm các giá trị gradient nhiệt là cần thiết nhằm nâng cao tính phù hợp và độ tin cậy của tiêu chuẩn.

Xuất phát từ những hạn chế của tiêu chuẩn hiện hành và thiếu hụt dữ liệu thực nghiệm trong nước, nghiên cứu này tập trung phân tích dữ liệu nhiệt độ và khí tượng thu thập từ thực địa và vệ tinh tại ba cầu BTCT mặt cắt hộp, đại diện cho ba khu vực khí hậu tiêu biểu. Các mô hình học máy (Machine Learning) tiên tiến được áp dụng để phát triển mô hình dự đoán gradient nhiệt tuyến tính theo phương đứng và phương ngang. Kết quả nghiên cứu cung cấp cơ sở dữ liệu và phương pháp đánh giá đáng tin cậy, góp phần hoàn thiện giải pháp thiết kế tác động nhiệt cho cầu BTCT trong điều kiện khí hậu Việt Nam.

2. XÂY DỰNG MÔ HÌNH DỰ ĐOÁN

2.1. Cơ sở lý thuyết và phương pháp

Hiện nay, gradient nhiệt trong cầu BTCT thường được xác định thông qua đo đạc thực nghiệm hoặc mô phỏng số. Tuy nhiên, các phương pháp này có chi phí cao, phạm vi áp dụng

hạn chế và độ chính xác phụ thuộc vào điều kiện biên của mô hình. Ngược lại, dữ liệu khí hậu vệ tinh như NASA [6] cung cấp thông tin chi tiết về nhiệt độ, gió và bức xạ mặt trời với độ tin cậy và độ bao phủ lớn, tạo điều kiện thuận lợi cho việc phát triển các mô hình dự đoán gradient nhiệt trong cầu BTCT.

Trên cơ sở đó, nghiên cứu này phát triển phương pháp xây dựng mô hình dự báo gradient nhiệt tuyến tính trên mặt cắt ngang dầm cầu, dựa trên đặc trưng hình học của tiết diện và dữ liệu khí tượng vệ tinh gồm nhiệt độ, tốc độ gió và bức xạ mặt trời.

Tại Việt Nam, Ngô Đăng Quang và cộng sự [7] đã tiến hành quan trắc thực địa và áp dụng hồi quy tuyến tính từng bước để dự báo nhiệt độ tại các vị trí đặc trưng trong mặt cắt dầm hộp BTCT. Kết quả cho thấy mô hình đạt độ chính xác cao, phù hợp với điều kiện khí hậu Việt Nam và có thể thay thế các phương pháp đo đạc truyền thống ở những khu vực thiếu thiết bị quan trắc chuyên dụng. Ngoài ra, Đỗ Anh Tú và cộng sự [8] đã ứng dụng mô hình ANN để dự báo nguy cơ nứt do nhiệt ở bê tông tuổi sớm tại trụ cầu, đạt sai số thấp, độ tin cậy cao và dễ dàng tích hợp vào quá trình thiết kế, thi công.

Các nghiên cứu quốc tế gần đây cũng cho thấy những kết quả tương tự. Sheng và cộng sự [9] đã phát triển mô hình dự báo gradient nhiệt trong dầm hộp bê tông bằng cách kết hợp mô phỏng phần tử hữu hạn với phân tích hồi quy, đạt sai số trung bình tuyệt đối nhỏ hơn $1\text{ }^{\circ}\text{C}$, phù hợp để áp dụng trực tiếp trong quy trình thiết kế cầu. Bên cạnh đó, Cha và cộng sự [10] chứng minh hiệu quả của việc ứng dụng mô hình học sâu (Deep Learning) trong giám sát sức khỏe kết cấu, khai thác dữ liệu khí hậu để dự báo và quản lý tình trạng kỹ thuật của cầu.

Từ những kết quả trên, nghiên cứu này mở rộng ứng dụng các mô hình học máy hiện đại nhằm nâng cao độ chính xác trong dự báo gradient nhiệt, bao gồm: hồi quy tuyến tính (Linear Regression), cây quyết định tăng cường (Boosted Tree), rừng ngẫu nhiên (Random Forest), XGBoost và ANN. Trong đó, hồi quy tuyến tính được sử dụng làm mô hình tham chiếu theo hướng dẫn của Montgomery và cộng sự [11]. Các mô hình Random Forest và Boosted Tree, do Breiman [12] và Friedman [13] phát triển, cho thấy khả năng xử lý hiệu quả dữ liệu phi tuyến và hạn chế quá khớp. XGBoost, được đề xuất bởi Chen và Guestrin [14], nổi bật với hiệu suất tính toán cao và khả năng tối ưu hóa mạnh mẽ. Trong khi đó, mạng nơ-ron nhân tạo, theo các nghiên cứu của LeCun và cộng sự [15] và Moon và cộng sự [16], thể hiện ưu thế rõ rệt trong mô hình hóa quan hệ phi tuyến sâu, đặc biệt phù hợp với tập dữ liệu lớn như tập dữ liệu về gradient nhiệt tuyến tính trong nghiên cứu này.

2.2. Cơ sở dữ liệu

Nghiên cứu sử dụng tập dữ liệu nhiệt độ thực đo thu thập tại ba cầu BTCT mặt cắt dạng hộp, đại diện cho các vùng khí hậu đặc trưng của Việt Nam. Cụ thể, bao gồm cầu Vĩnh Thịnh (Hà Nội) thuộc khu vực miền Bắc, cầu Quán Hàu (Quảng Bình) đại diện miền Trung và cầu Gò Găng (Vũng Tàu) thuộc miền Nam. Việc lựa chọn ba công trình này tuân theo phân vùng khí hậu quy định trong QCVN 02:2021/BXD [17].

Hệ thống quan trắc nhiệt độ được bố trí tại 6 trạm đo, mỗi cầu lắp đặt 2 trạm gồm một trạm ở vị trí gối và một trạm ở giữa nhịp. Mỗi trạm có 21 điểm đo được bố trí tại các vị trí và độ sâu khác nhau trong tiết diện dầm hộp. Dữ liệu được ghi nhận tự động bằng cảm biến nhiệt K type với chu kỳ 5 phút và truyền về trung tâm qua kết nối 4G, bảo đảm tính liên tục và độ chi tiết cao. Để đồng bộ với dữ liệu khí tượng và hạn chế nhiễu ngắn hạn, chuỗi số liệu được xử lý về các mốc ghi nhận theo giờ, vừa duy trì độ phân giải cần thiết cho quá trình huấn luyện mô hình học máy, vừa giảm khối lượng tính toán.

Bên cạnh dữ liệu nhiệt độ đo thực tế, nhóm nghiên cứu còn thu thập dữ liệu khí tượng tương ứng tại từng vị trí cầu từ nền tảng vệ tinh NASA POWER [6]. Các thông tin bao gồm nhiệt độ không khí trung bình, tổng bức xạ mặt trời có mây, vận tốc và hướng gió ở độ cao 10 mét, trong giai đoạn từ tháng 01/2022 đến tháng 09/2024. Đây là nhóm tham số đầu vào phản ánh điều kiện khí hậu ngoài công trình, được sử dụng trong quá trình huấn luyện và kiểm định mô hình. Tổng thể, các tham số nghiên cứu được chia thành ba nhóm chính:

- Nhóm tham số đầu vào 01: Các thông số hình học của tiết diện đầm hộp, bao gồm kích thước, góc nghiêng và hướng cầu, phản ánh khả năng hấp thụ và phân bố nhiệt theo hình học.
- Nhóm tham số đầu vào 02: Các dữ liệu khí tượng ngoài cầu, thu thập từ NASA POWER, bao gồm nhiệt độ môi trường, bức xạ mặt trời và đặc tính gió.
- Nhóm tham số đầu ra: Các giá trị gradient nhiệt độ tuyến tính theo phương đứng, gồm gradient nhiệt dương lớn nhất (MAX-Gy) và gradient nhiệt âm nhỏ nhất (MIN-Gy), được tính toán từ dữ liệu thực đo tại các điểm trong mặt cắt.

Chi tiết các tham số đầu vào và đầu ra được trình bày trong Bảng 1.

Bảng 1. Các tham số sử dụng xây dựng mô hình dự đoán gradient nhiệt tuyến tính theo phương đứng.

STT	Tham số	Mô tả	Ghi chú
01	H (m)	Chiều cao mặt cắt đầm hộp	Nhóm tham số đầu vào 01: Đặc trưng hình học mặt cắt đầm hộp cầu quan trắc nhiệt độ
02	B (m)	Chiều rộng mặt cắt đầm hộp	
03	hm (m)	Độ dày bản nắp	
04	hd (m)	Độ dày bản đáy	
05	ht (m)	Độ dày bản thành	
06	Lcs (m)	Chiều dài consol	
07	α (°)	Góc hướng cầu so với hướng Bắc	
08	β (°)	Góc nghiêng bản thành cầu so với mặt nằm ngang	
09	ALL_DWN (W/m ²)	Tổng bức xạ mặt trời có mây	Nhóm tham số đầu vào 02: Dữ liệu khí tượng được thu thập từ trạm khí tượng gần cầu quan trắc thực tế
10	Tout-Avg (°C)	Nhiệt độ trung bình bên ngoài cầu	
11	WS10M (m/s)	Tốc độ gió ở độ cao 10m bên ngoài cầu	
12	WD10M (°)	Hướng gió ở độ cao 10m bên ngoài cầu	
13	MAX-Gy (°C/m)	Gradient nhiệt dương tuyến tính theo phương đứng lớn nhất	Nhóm tham số đầu ra: Được tính toán xác định từ dữ liệu nhiệt độ quan trắc thực tế ở các mặt cắt đầm cầu
14	MIN-Gy (°C/m)	Gradient nhiệt âm tuyến tính theo phương đứng nhỏ nhất	

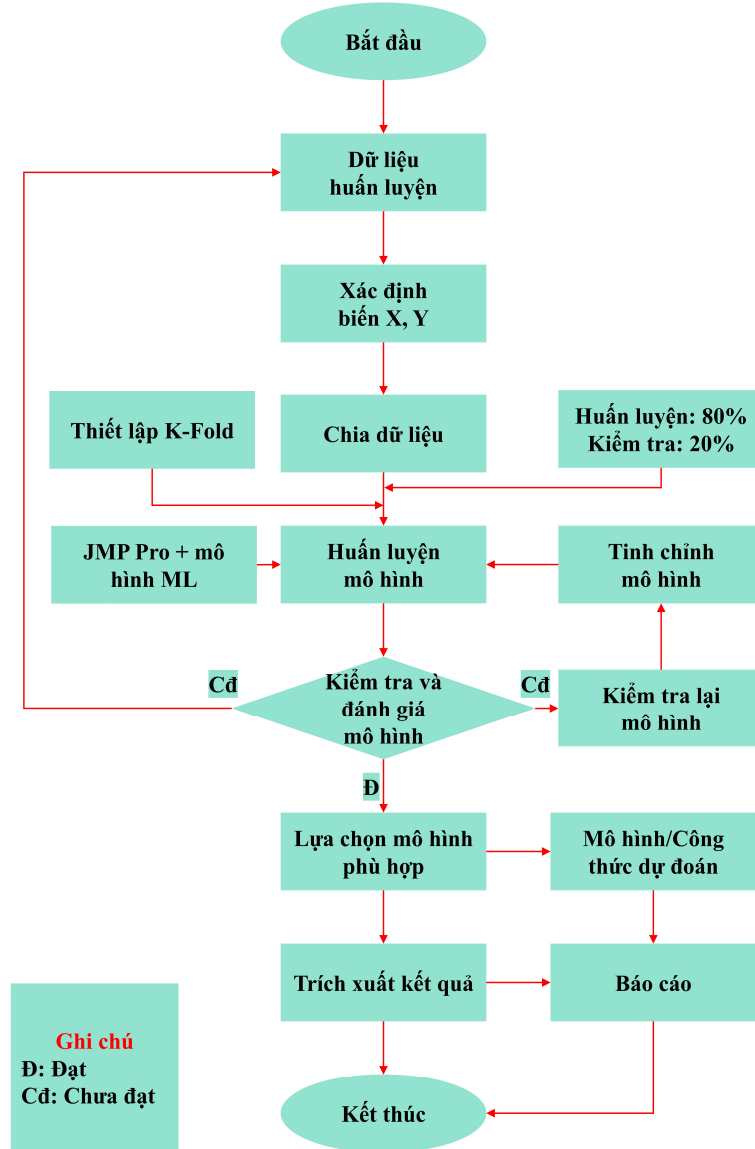
2.3. Xây dựng mô hình dự đoán

2.3.1. Quy trình triển khai

Quy trình xây dựng mô hình dự báo gradient nhiệt được thiết kế theo cấu trúc hệ thống khép kín, bao gồm các giai đoạn chính: chuẩn bị dữ liệu, huấn luyện, kiểm định và lựa chọn mô hình tối ưu. Chất lượng mô hình được kiểm soát nghiêm ngặt thông qua kỹ thuật kiểm định chéo K-Fold, cho phép đánh giá độ tin cậy và khả năng khái quát. Những mô hình chưa

đạt yêu cầu sẽ được hiệu chỉnh và đánh giá lại trong các vòng lặp huấn luyện kế tiếp. Quá trình này được hỗ trợ bởi phần mềm JMP Pro và các thuật toán học máy hiện đại, nhằm tăng cường tính ổn định và độ chính xác của kết quả dự báo.

Lưu đồ tổng quan mô tả quy trình triển khai được trình bày tại Hình 1.



Hình 1. Lưu đồ quy trình xây dựng và huấn luyện mô hình học máy.

2.3.2. Dữ liệu huấn luyện

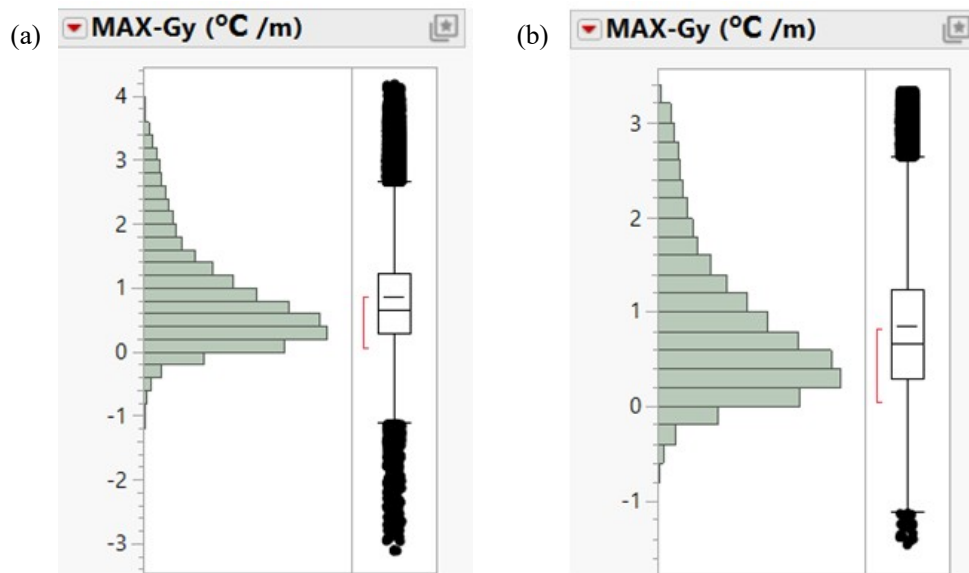
Tập dữ liệu được kiểm tra kỹ lưỡng nhằm loại bỏ các giá trị thiếu, sai số đo và những điểm bất thường, đảm bảo tính toàn vẹn và độ tin cậy cho quá trình phân tích. Các giá trị ngoại lai được nhận diện và loại trừ bằng phương pháp chuẩn hóa Z-score, một kỹ thuật thống kê phổ biến, có độ tin cậy cao khi xử lý dữ liệu quy mô lớn trong học máy.

Tính hợp lý của phương pháp này đã được xác nhận trong nhiều nghiên cứu trước, tiêu

biểu là nghiên cứu của Xu và cộng sự [18] khi giám sát nhiệt độ trên cầu dây văng. Các tác giả đã đề xuất ba tiêu chí xác định dữ liệu nhiệt độ bất thường, bao gồm: (i) chênh lệch lớn giữa nhiệt độ cực đại và cực tiểu trong một giờ vượt quá ngưỡng cho phép; (ii) độ lệch chuẩn bằng 0 – dấu hiệu cho thấy cảm biến có thể bị đóng băng; và (iii) nhiệt độ trung bình theo giờ thấp hơn ngưỡng vật lý tối thiểu.

Trong nghiên cứu này, các điểm dữ liệu có Z-score tuyệt đối vượt quá ± 3 được xác định là ngoại lai và loại bỏ, dựa trên quy tắc ba sigma trong phân phối chuẩn – theo đó 99,73% các giá trị sẽ nằm trong khoảng ± 3 nếu dữ liệu tuân theo phân phối chuẩn hoặc gần chuẩn. Phương pháp này cho phép loại bỏ hiệu quả các nhiễu loạn do cảm biến hoặc điều kiện thời tiết cực đoan, đồng thời vẫn duy trì được cấu trúc phân bố tự nhiên của dữ liệu.

Sau quá trình xử lý, 123.433 bản ghi dữ liệu hợp lệ được giữ lại, đảm bảo yêu cầu về tính đồng nhất và độ ổn định, phục vụ cho giai đoạn huấn luyện các mô hình học máy. Hình 2 trình bày phân bố gradient nhiệt độ theo phương đứng trước và sau khi loại bỏ ngoại lai bằng phương pháp Z-score.



Hình 2. Phân bố dữ liệu MAX-Gy trước khi loại bỏ giá trị nhiễu (a) và sau khi loại bỏ giá trị nhiễu (b).

2.3.3. Phân chia dữ liệu và kiểm định mô hình

Để đánh giá khách quan khả năng khái quát hóa của mô hình, tập dữ liệu được chia thành hai phần độc lập: 80% cho huấn luyện và 20% cho kiểm định. Tập huấn luyện được dùng để xây dựng và hiệu chỉnh mô hình, trong khi tập kiểm định được dùng để đánh giá hiệu suất dự báo trên các dữ liệu chưa từng được học. Tỷ lệ 80:20 là cách phân chia phổ biến trong học máy, giúp cân bằng giữa hiệu quả huấn luyện và độ tin cậy khi đánh giá. Phương pháp này đã được áp dụng thành công trong nhiều nghiên cứu, điển hình là Moon và cộng sự [16] khi dự đoán độ võng động của dầm cầu từ dữ liệu đo đạc thực tế.

Bên cạnh đó, nhằm tăng tính ổn định và giảm sai lệch do việc chia ngẫu nhiên một lần, nghiên cứu áp dụng kỹ thuật kiểm định chéo K-Fold với các giá trị K khác nhau. Cách tiếp cận này cho phép mô hình được huấn luyện và đánh giá nhiều lần trên các tập con độc lập, qua đó phản ánh chính xác hơn hiệu suất tổng thể. Kiểm định K-Fold đã được chứng minh hiệu quả trong nhiều nghiên cứu học thuật, chẳng hạn Ling và cộng sự [19] khi dự báo cường độ

độ chịu nén của bê tông trong môi trường biển. Theo khuyến nghị của Hastie và cộng sự trong “The Elements of Statistical Learning” [20], giá trị K bằng 5 hoặc 10 thường mang lại sự cân bằng giữa độ chính xác và chi phí tính toán.

Đối với mô hình ANN, nghiên cứu tiến hành kiểm thử thực nghiệm với nhiều giá trị K khác nhau, trong đó có K bằng 5, 6 và 7, nhằm xác định giá trị tối ưu cho bài toán cụ thể. Các chỉ số đánh giá như hệ số xác định, sai số tuyệt đối trung bình, sai số căn phương bình phương trung bình và sai số căn phương trung bình được tính toán và so sánh chi tiết. Kết quả thực nghiệm cho thấy, khi sử dụng K=6, mô hình ANN đạt được độ chính xác cao nhất với sai số trung bình nhỏ nhất trên cả tập huấn luyện và kiểm tra, đồng thời duy trì độ ổn định về phương sai tốt hơn so với các giá trị K khác.

Trên cơ sở các kết quả phân tích, có thể thấy rằng phân chia dữ liệu theo tỷ lệ 80:20 kết hợp với kỹ thuật kiểm định chéo K-Fold, trong đó sử dụng K bằng 5 cho mô hình XGBoost và K bằng 6 cho mô hình ANN, là hoàn toàn phù hợp với đặc điểm của bộ dữ liệu nghiên cứu. Phương pháp tổ chức dữ liệu này không chỉ tối ưu hóa độ chính xác mà còn nâng cao độ tin cậy của kết quả dự báo, góp phần đảm bảo tính khả thi và hiệu quả trong các ứng dụng thực tiễn liên quan đến dự đoán thành phần gradient nhiệt trong đầm cầu BTCT có mặt cắt dạng hộp.

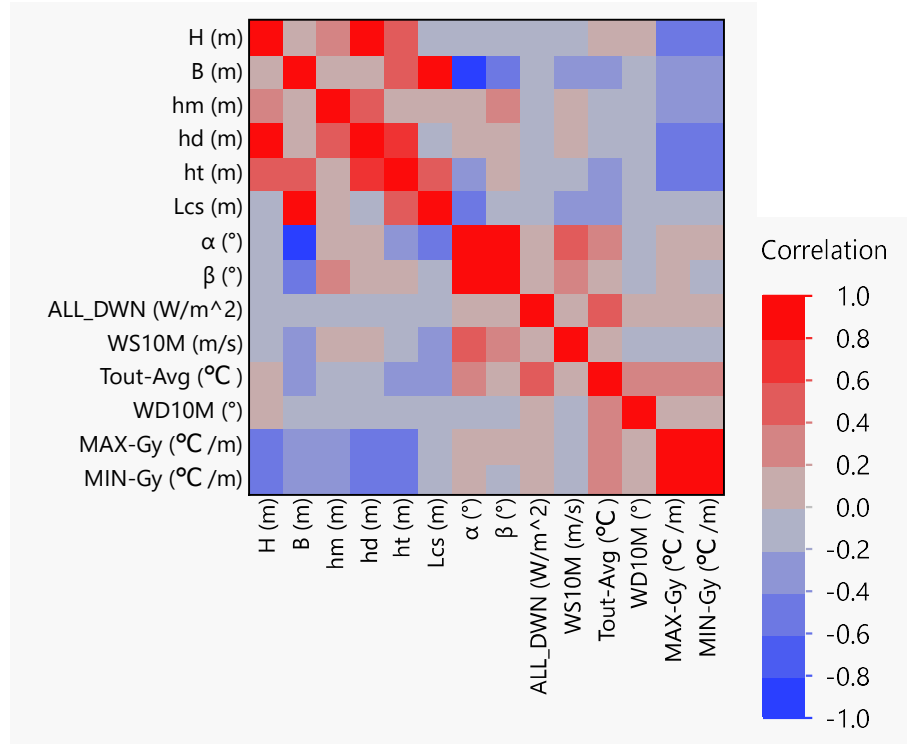
2.3.4. Mức độ ảnh hưởng của các tham số đầu vào

Dựa trên tập dữ liệu đã được xử lý và chuẩn hóa, nghiên cứu tiến hành phân tích mối tương quan tuyến tính giữa các tham số đầu vào và hai tham số đầu ra là gradient nhiệt dương tuyến tính lớn nhất theo phương đứng và gradient nhiệt âm tuyến tính nhỏ nhất theo phương đứng, lần lượt ký hiệu là MAX-Gy và MIN-Gy.

Kết quả phân tích, minh họa tại Hình 3, cho thấy:

- Yếu tố khí tượng: Nhiệt độ không khí trung bình là tham số khí tượng có ảnh hưởng mạnh nhất, với hệ số tương quan đạt 0,3853 (MAX-Gy) và 0,3720 (MIN-Gy). Hướng gió có mối tương quan dương ở mức yếu (khoảng 0,175), trong khi tổng bức xạ mặt trời và vận tốc gió không ảnh hưởng đáng kể, với hệ số tương quan tuyệt đối đều dưới 0,07.
- Yếu tố đặc trưng hình học mặt cắt: Chiều cao đầm (H) cho thấy mối tương quan âm mạnh với cả hai tham số đầu ra là -0,4582 với MAX-Gy và -0,4347 với MIN-Gy, phản ánh vai trò điều tiết của chiều cao tiết diện đối với phân bố nhiệt. Độ dày bản đáy (hd) và bản thành trên (ht) cũng thể hiện tương quan âm rõ rệt, lần lượt đạt -0,4942 và -0,4225 với MAX-Gy. Các thông số còn lại như chiều dài console (Lcs), góc α , β hay độ dày bản nắp (hm) thể hiện ảnh hưởng từ thấp đến trung bình.

Kết quả phân tích cho thấy gradient nhiệt theo phương đứng chịu ảnh hưởng đồng thời bởi điều kiện khí tượng và các tham số hình học của mặt cắt đầm, đặc biệt là chiều cao và độ dày của bản nắp, bản đáy và bản thành. Các thông số này điều tiết đáng kể biên độ nhiệt trong tiết diện, qua đó cung cấp cơ sở quan trọng cho việc lựa chọn tập tham số đầu vào phù hợp trong xây dựng và hiệu chỉnh mô hình học máy dự báo gradient nhiệt.



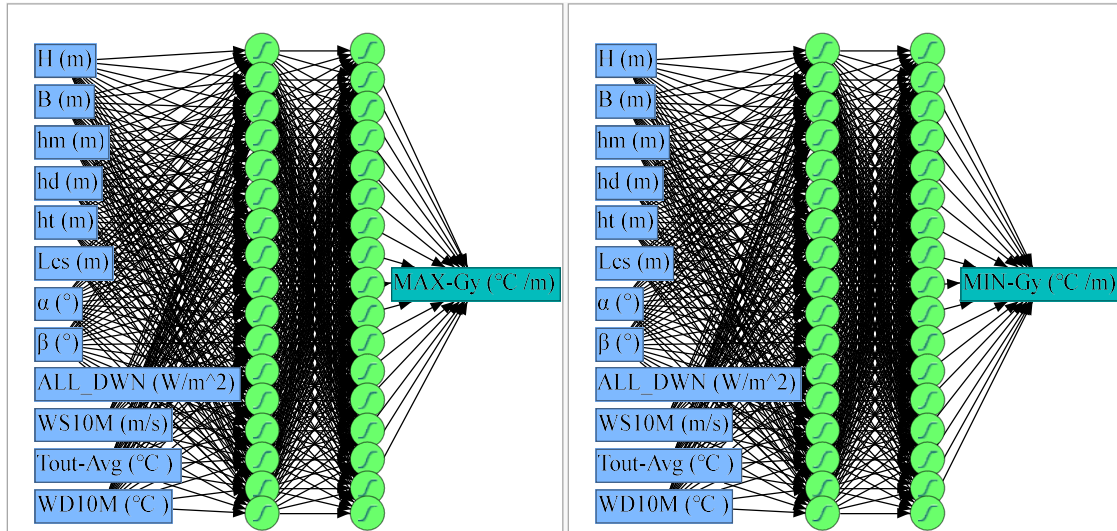
Hình 3. Ma trận phân tích tương quan giữa các tham số đầu ra và tham số đầu vào.

2.3.5. Tối ưu hóa tham số và đánh giá mô hình

Kết quả phân tích sơ bộ với các tham số mặc định cho thấy mô hình ANN đạt hiệu quả dự báo cao hơn rõ rệt so với các mô hình khác. Hệ số xác định R^2 đều vượt 0,70 đối với hai biến đầu ra MAX-Gy và MIN-Gy (xem Bảng 3), khẳng định khả năng khái quát hóa và độ chính xác cao của ANN trong mô phỏng các quan hệ phi tuyến giữa điều kiện khí tượng, thông số hình học và sự phân bố gradient nhiệt. Do đó, nghiên cứu tập trung vào quá trình tối ưu hóa tham số và đánh giá chi tiết mô hình ANN.

Sau quá trình thử nghiệm, cấu trúc ANN tối ưu được xác định gồm hai tầng ẩn, mỗi tầng có 17 neuron và sử dụng hàm kích hoạt TanH. Sơ đồ cấu trúc tối ưu được minh họa ở Hình 4. Cấu hình này được lựa chọn nhằm phát huy khả năng mô hình hóa các mối quan hệ phi tuyến phức tạp từ dữ liệu đầu vào. Hàm kích hoạt TanH không chỉ giúp chuẩn hóa đầu ra của mỗi neuron trong khoảng $[-1, 1]$ mà còn cải thiện tốc độ hội tụ và độ ổn định của quá trình huấn luyện, phù hợp với khuyến nghị trong tài liệu “Deep Learning” của Goodfellow và cộng sự [21].

Về thiết lập cấu hình huấn luyện, không sử dụng chế độ “Boosting” nhằm hạn chế sự phức tạp không cần thiết, giúp mô hình tập trung học trực tiếp thông qua thuật toán lan truyền ngược. Phương pháp tối ưu hóa được lựa chọn là “Squared”, nhằm mục tiêu tối thiểu hóa tổng bình phương sai số giữa giá trị dự đoán và giá trị thực tế. Các tùy chọn xử lý dữ liệu như “Transform Covariates” và “Robust Fit” cũng được loại bỏ nhằm duy trì trạng thái nguyên bản của dữ liệu, đảm bảo tính khách quan và hiệu quả trong việc đánh giá hiệu suất mô hình.



Hình 4. Cấu trúc ANN dự đoán thành phần gradient nhiệt tuyến tính theo phương đứng.

Bên cạnh mô hình tối ưu hóa trên, nghiên cứu đã tiến hành thử nghiệm cấu trúc ANN đơn giản hơn, chỉ sử dụng hàm kích hoạt TanH mà không tích hợp thêm các thành phần tuyến tính (Linear) hay Gaussian. Kết quả thử nghiệm cho thấy ngay cả với cấu trúc đơn giản này, mô hình ANN vẫn đạt được hệ số xác định R^2 lớn hơn 0,70, chứng minh hiệu quả cao trong việc mô hình hóa các mối quan hệ phi tuyến phức tạp [22]. Tuy nhiên, khi bổ sung thêm các thành phần Linear và Gaussian, độ chính xác của mô hình được cải thiện rõ rệt nhờ vào khả năng mô hình hóa linh hoạt các xu hướng tuyến tính và các biến thiên cục bộ. Đặc biệt, việc sử dụng thêm thành phần Gaussian đã được chứng minh là hiệu quả trong việc bắt các biến thiên nhỏ và yếu tố nhiễu tự nhiên trong dữ liệu nhiệt độ, phù hợp với các nghiên cứu đã công bố trước đây [23].

Như vậy, kết quả tối ưu hóa tham số và đánh giá chi tiết cho thấy cấu trúc ANN gồm hai tầng ẩn với 17 neuron mỗi tầng, sử dụng hàm kích hoạt TanH, kết hợp bổ sung các thành phần Linear và Gaussian là cấu trúc tối ưu nhất cho bài toán dự đoán thành phần gradient nhiệt bất lợi trên dầm cầu BTCT mặt cắt dạng hộp, đạt độ chính xác và độ tin cậy cao nhất trong các trường hợp nghiên cứu thử nghiệm.

2.4. Đánh giá và phân tích mô hình dự đoán

2.4.1. Hiệu suất và độ chính xác

Để đánh giá độ chính xác và khả năng khái quát hóa của các mô hình học máy, nghiên cứu sử dụng các chỉ số phổ biến bao gồm sai số tuyệt đối trung bình (Mean Absolute Error – MAE), sai số căn phương bình phương trung bình (Root Mean Square Error – RMSE), sai số căn phương bình quân có điều chỉnh (Root Average Squared Error – RASE) và hệ số xác định (R^2) [24]. Ý nghĩa của các chỉ số đánh giá được trình bày trong Bảng 2.

Kết quả tổng hợp trong Bảng 3 cho thấy mô hình ANN đạt hiệu suất dự đoán cao và ổn định nhất. Tuy nhiên, việc phân tích độ ảnh hưởng chi tiết của từng tham số trong mô hình ANN phức tạp cần các kỹ thuật chuyên sâu như SHAP, LIME và sẽ là hướng phát triển tiếp theo, kết quả phân tích tương quan tuyến tính ban đầu cũng đã cung cấp những gợi ý quan trọng, cho thấy các đặc trưng chiều cao, độ dày bản đáy và bản thành đóng vai trò then chốt trong việc hình thành gradient nhiệt.

Bảng 2. Các chỉ số đánh giá mô hình.

Chỉ số	Công thức	Ý nghĩa
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{k=1}^n y_k - \hat{y}_k $ (1)	Trung bình tuyệt đối sai số dự đoán; càng nhỏ càng chính xác.
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_k)^2}$ (2)	Nhấn mạnh sai số lớn; phản ánh độ lệch tổng thể của mô hình.
Root Average Squared Error (RASE)	$RASE = \sqrt{\frac{1}{n-p} \sum_{k=1}^n (y_k - \hat{y}_k)^2}$ (3)	RMSE hiệu chỉnh theo độ phức tạp mô hình; dùng để so sánh mô hình.
Hệ số xác định (R^2)	$R^2 = 1 - \frac{\sum_{k=1}^n (y_k - \hat{y}_k)^2}{\sum_{k=1}^n (y_k)^2}$ (4)	Mức độ giải thích biến thiên dữ liệu của mô hình; càng gần 1 càng tốt.

Giá trị R^2 trên cả hai tập huấn luyện và kiểm định đều vượt 0,70 đối với cả MAX-Gy và MIN-Gy. Sai số thấp và phân dư phân bố ngẫu nhiên, không tạo thành xu hướng hệ thống, cho thấy mô hình có khả năng mô phỏng tốt các mối quan hệ phi tuyến giữa đặc trưng hình học, điều kiện khí tượng và gradient nhiệt tuyến tính theo phương đứng.

Bảng 3. Kết quả so sánh hiệu suất các mô hình học máy trong dự đoán giá trị MAX-Gy và MIN-Gy.

Mô hình huấn luyện	Thông số đánh giá	MAX-Gy		MIN-Gy	
		Training	Validation	Training	Validation
Hồi quy tuyến tính (Linear Regression)	R^2	0,4434	-	0,4204	-
	RMSE	0,5748	-	0,5550	-
	MAE	-	-	-	-
	RASE	-	-	-	-
Cây tăng cường (Boosted Tree)	R^2	0,7530	0,6570	0,7360	0,6360
	RMSE	-	-	-	-
	MAE	-	-	-	-
	RASE	0,3830	0,4515	0,3746	0,4400
Rừng ngẫu nhiên (Random Forest)	R^2	0,5960	0,5730	0,5860	0,5630
	RMSE	-	-	-	-
	MAE	-	-	-	-
	RASE	0,4897	0,5035	0,4690	0,4818
XGBoost (Extreme Gradient Boosting)	R^2	0,7550	0,6950	0,7322	0,6787
	RMSE	0,3814	0,4255	0,3772	0,4132
	MAE	0,2652	0,2957	0,2624	0,2866
	RASE	-	-	-	-
Mạng nơ-ron nhân tạo (Artificial Neural Networks - ANN)	R^2	0,7229	0,7223	0,7093	0,7096
	RMSE	-	-	-	-
	MAE	-	-	-	-
	RASE	0,4487	0,4515	0,4331	0,4324

Các mô hình Boosted Tree và XGBoost cũng đạt hiệu quả tương đối tốt nhưng có dấu hiệu quá khớp nhẹ. Trong khi đó, mô hình hồi quy tuyến tính và rừng ngẫu nhiên cho kết quả

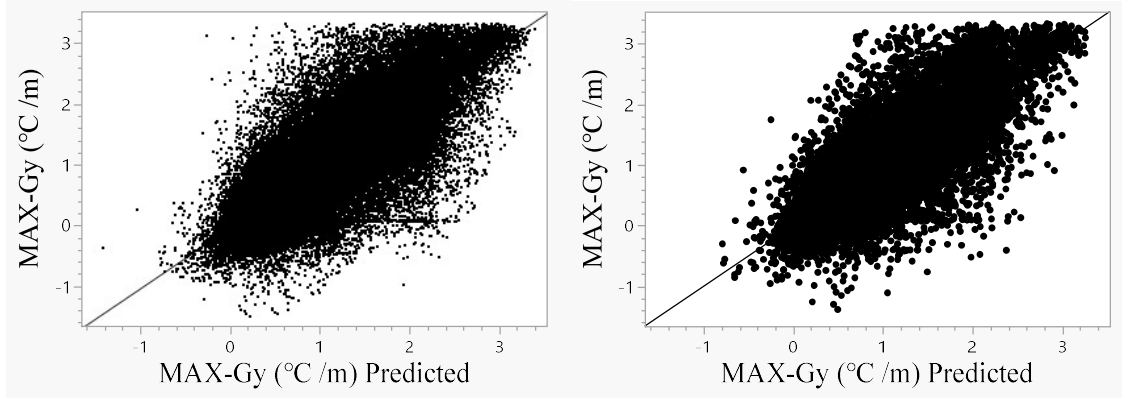
kém, với R^2 dưới 0,6, không đáp ứng yêu cầu khái quát hóa.

Ngoài ra, hiệu suất chi tiết cũng được đánh giá và trình bày thêm trong Bảng 4 có xét thêm các tham số như sai số tuyệt đối trung bình (MAD), chỉ số SSE và $-\text{LogLikelihood}$. Kết quả cũng cho thấy mô hình ANN đạt R^2 khoảng 0,72 cho cả hai tham số mục tiêu và sai số căn phương trung bình (RASE) dao động từ 0,43 đến 0,45 $^{\circ}\text{C}/\text{m}$, trong khi sai số tuyệt đối trung bình (MAD) duy trì dưới 0,31 $^{\circ}\text{C}/\text{m}$ và các chỉ số SSE và $-\text{LogLikelihood}$ ổn định giữa hai tập huấn luyện và kiểm định, khẳng định tính tổng quát hóa của mô hình.

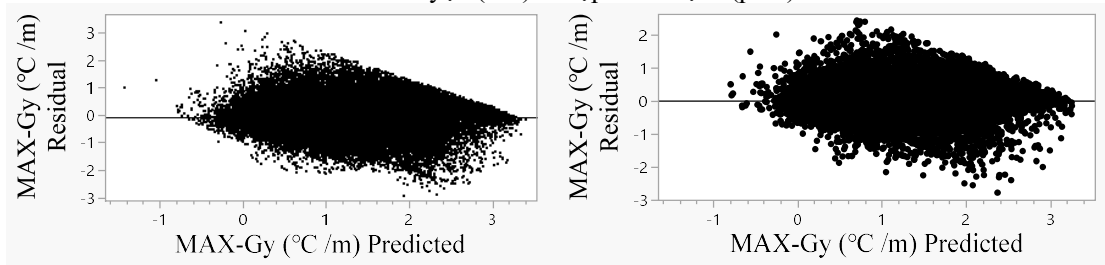
Bảng 4. Các chỉ số đánh giá hiệu suất mô hình ANN dự đoán MAX-Gy ($^{\circ}\text{C}/\text{m}$) và MIN-Gy ($^{\circ}\text{C}/\text{m}$).

Chỉ số	MAX-Gy (Huấn luyện)	MAX-Gy (Kiểm định)	MIN-Gy (Huấn luyện)	MIN-Gy (Kiểm định)
R^2	0,7229	0,7223	0,7093	0,7096
RASE	0,4487	0,4515	0,4331	0,4324
MAD	0,3073	0,3076	0,2962	0,2960
SSE	52798,3360	10580,3970	48989,1870	9789,0335
$-\text{LogLikelihood}$	20709,5490	4193,8943	19293,6050	3847,2687

Biểu đồ giá trị thực so với dự đoán (Actual by Predicted) trong Hình 5 cho thấy các điểm dữ liệu phân bố gần đường lý tưởng. Trong Hình 6 - Biểu đồ phần dư (Residual by Predicted) thể hiện phân bố ngẫu nhiên quanh trục hoành, không xuất hiện xu hướng hệ thống. Những đặc điểm này xác nhận tính ổn định và hiệu quả của mô hình ANN trong việc dự đoán gradient nhiệt tuyến tính theo phương đứng trong kết cấu dầm hộp BTCT.



Hình 5. Biểu đồ giá trị thực so với giá trị dự đoán (Actual by Predicted Plot) đối với MAX-Gy, cho tập huấn luyện (trái) và tập kiểm định (phải).



Hình 6. Biểu đồ phần dư (Residual by Predicted Plot) đối với MAX-Gy, cho tập huấn luyện (trái) và tập kiểm định (phải).

2.4.2. Công thức dự báo

Sau khi huấn luyện, mô hình ANN cho phép thiết lập một công thức dự báo gần đúng, giúp ước lượng nhanh các giá trị MAX-Gy và MIN-Gy từ các tham số hình học mặt cắt và dữ liệu khí tượng. Các công thức này không dựa trên một lý thuyết vật lý định sẵn mà được trích xuất tự động từ chính mô hình ANN sau khi được huấn luyện với bộ dữ liệu quan trắc thực tế. Công thức tổng quát có dạng:

$$+ MAX_Gy = 0,8384 + \sum_{i=1}^{17} w_i \times \tanh \left(0,5 \times \left(b_i + \sum_{j=1}^{17} a_{ij} \times \tanh \left(0,5 \times \left(c_i + \sum_{k=1}^{12} d_{jk} \times X_k \right) \right) \right) \right) \quad (5)$$

$$+ MIN_Gy = 1,6431 + \sum_{i=1}^{17} w_i \times \tanh \left(0,5 \times \left(b_i + \sum_{j=1}^{17} a_{ij} \times \tanh \left(0,5 \times \left(c_i + \sum_{k=1}^{12} d_{jk} \times X_k \right) \right) \right) \right) \quad (6)$$

Trong đó:

- X_k : là các tham số đầu vào (H, B, ht, hb, hm, Lcs, α , β , ALL_DWN, WS10M, WD10M, Tout-Avg).
- $\tanh$: là hàm kích hoạt trong mô hình ANN.
- Các hệ số d_{jk} , a_{ij} , w_i , b_i , c_i : được trích xuất tự động từ công cụ phần mềm JMP Pro sau khi hoàn tất quá trình huấn luyện mạng nơ-ron.

Các hệ số trong mô hình là những giá trị được mạng nơ-ron tự động tối ưu hóa trong quá trình học từ dữ liệu. Nhằm đảm bảo tính minh bạch và tạo điều kiện cho việc tái lập nghiên cứu, toàn bộ các hệ số chi tiết đã được cung cấp trong Phụ lục của bài báo [25].

Công thức này hỗ trợ hiệu quả trong việc dự báo nhanh gradient nhiệt tuyến tính theo phương đứng, hỗ trợ trong công tác thiết kế và phân tích đánh giá giải pháp bảo trì kết cấu dầm cầu BTCT mặt cắt dạng hộp, phù hợp với điều kiện khí hậu đặc trưng cho từng khu vực địa lý cụ thể.

3. KẾT LUẬN

Nghiên cứu xây dựng mô hình học máy nhằm ước lượng tải trọng nhiệt bất lợi tác động lên dầm cầu bê tông cốt thép mặt cắt hộp, tập trung vào gradient nhiệt tuyến tính theo phương đứng MAX-Gy và MIN-Gy. Tập dữ liệu được tổng hợp từ ba cầu đại diện cho các khu vực khí hậu đặc trưng của Việt Nam, kết hợp với dữ liệu khí tượng vệ tinh NAS, sau đó được xử lý, chuẩn hóa và chia thành các bộ dữ liệu huấn luyện và kiểm định.

Các kết quả của nghiên cứu bao gồm:

- Mô hình ANN đạt hiệu suất hiệu quả nhất với $R^2 = 0,7223$ (MAX-Gy) và $0,7096$ (MIN-Gy), mức chấp nhận được cho các bài toán phi tuyến phức tạp. Mặc dù vẫn còn khoảng 28% phương sai chưa được giải thích, nhưng với một hệ thống vật lý phức tạp như kết cấu cầu, vốn chịu tác động của nhiều yếu tố ngẫu nhiên và khó đo lường, việc đạt được hệ số xác định trên 0,70 là một kết quả rất đáng khích lệ, cho thấy mô hình đã nắm bắt được phần lớn các quy luật chi phối quan trọng.
- Mô hình XGBoost cũng cho thấy hiệu quả dự báo cao, với R^2 xấp xỉ 0,70, chứng tỏ tiềm năng ứng dụng trong các phân tích thực tế quy mô lớn.
- Kết quả phân tích tương quan đã xác định rõ vai trò quan trọng của các yếu tố khí

tượng (nhiệt độ không khí trung bình, hướng gió) và các đặc trưng hình học (độ dày bản thành và bản đáy) đối với sự biến thiên gradient nhiệt.

Kết quả nghiên cứu khẳng định tiềm năng ứng dụng thực tiễn của mô hình ANN, cung cấp giải pháp hiệu quả trong dự đoán tải trọng nhiệt cho cầu BTCT. Mô hình ANN còn là công cụ hỗ trợ ước lượng nhanh các giá trị gradient nhiệt, giúp kỹ sư có thêm cơ sở trong thiết kế và đánh giá tải trọng nhiệt, qua đó nâng cao mức độ phù hợp của các giải pháp thiết kế tại Việt Nam.

TÀI LIỆU THAM KHẢO

- [1]. M.M. Elbadry, A. Ghali, Temperature variations in concrete bridges, *Journal of Structural Engineering*, 109 (1983) 2355–2374. [https://doi.org/10.1061/\(ASCE\) 0733-9445\(1983\)109:10\(2355\)](https://doi.org/10.1061/(ASCE) 0733-9445(1983)109:10(2355))
- [2]. R.A. Imbsen, D.E. Vandershaf, R.A. Schamber, R.V. Nutt, Thermal effects in concrete bridge superstructures, NCHRP Program Report 276, Transportation Research Board, Washington D.C., 63–67, 1985.
- [3]. B. Gu, F.Y. Zhou, W. Gao, F.Z. Xie, L.H. Le, Temperature gradient and its effect on long-span prestressed concrete box girder bridge, *Advances in Civil Engineering*, (2020). <https://doi.org/10.1155/2020/5956264>
- [4]. Q. Shen, J. Chen, C. Yue, H. Cao, C. Chen, W. Qian, Investigation on the through-thickness temperature gradient and thermal stress of concrete box girders, *Buildings*, 13 (2023) 2882. <https://doi.org/10.3390/buildings13112882>
- [5]. TCVN 11823:2017, Thiết kế cầu đường bộ – Phần 3: Tải trọng và hệ số tải trọng, 2017.
- [6]. NASA Langley Research Center, POWER Data Access Viewer, NASA. Truy cập tại: <https://power.larc.nasa.gov/data-access-viewer/>
- [7]. Đ.Q. Ngô, H.C. Nguyễn, Monitoring and analysis of temperature distribution in reinforced concrete bridge box girders in Vietnam, *Case Studies in Construction Materials*, 20 (2024) e02857.
- [8]. A.T. Đỗ, B.A. Lê, Machine learning approach for predicting early-age thermal cracking potential in concrete bridge piers, *Finite Elements in Mechanics*, (2024). <https://doi.org/10.1016/j.finmec.2024.100297>
- [9]. X. Sheng, T. Zhou, S. Huang, C. Cai, T. Shi, Prediction of vertical temperature gradient on concrete box-girder considering different locations in China, *Case Studies in Construction Materials*, 16 (2022) e01026. <https://doi.org/10.1016/j.cscm.2022.e01026>
- [10]. Y.J. Cha, R. Ali, J. Lewis, O. Büyükoztürk, Deep learning-based structural health monitoring, *Automation in Construction*, 161 (2024) 105328. <https://doi.org/10.1016/j.autcon.2024.105328>
- [11]. D. C. Montgomery, E. A. Peck, G. G. Vining, *Introduction to Linear Regression Analysis*, 6th edition. Hoboken, New Jersey: Wiley, 2021.
- [12]. L. Breiman, Random forests, *Machine Learning*, 45 (2001) 5–32. <https://doi.org/10.1023/A:1010933404324>
- [13]. J.H. Friedman, Greedy function approximation: a gradient boosting machine, *Annals of Statistics*, 29 (2001). <https://doi.org/10.1214/aos/1013203451>
- [14]. T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, (2016) 785–794. <https://doi.org/10.1145/2939672.2939785>
- [15]. Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature*, 521 (2015) 436–444. <https://doi.org/10.1038/nature14539>
- [16]. H. S. Moon, Y. K. Hwang, M. K. Kim, Kang, H. T. Lim, Application of artificial neural network to predict dynamic displacements from measured strains for a highway bridge under traffic loads, *Journal of Civil Structural Health Monitoring*, 12 (2021) 117–126.
- [17]. Bộ Xây dựng, QCVN 02: 2021/BXD, Quy Chuẩn Kỹ Thuật Quốc Gia Sổ Liệu Điều Kiện Tự Nhiên Dùng Trong Xây Dựng, 2021.

- [18]. Y.L.Xu, B.Chen, C.L.Ng, K.Y.Wong, W.Y.Chan, Monitoring temperature effect on a long suspension bridge, *Structural Control and Health Monitoring*, 17 (2010) 632–653.
- [19]. H. Ling, C. Qian, W. Kang, C. Liang, H. Chen, Combination of Support Vector Machine and K-Fold cross validation to predict compressive strength of concrete in marine environment, *Construction and Building Materials*, 206 (2019) 355–363. <https://doi.org/10.1016/j.conbuildmat.2019.02.071>
- [20]. T. Hastie, R. Tibshirani, J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer, 2009. [Online]. Available: <https://www.abebooks.it/9780387848570/Elements-Statistical-Learning-Data-Mining-0387848576/plp>.
- [21]. I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*. Cambridge, MA: MIT Press, 2016. [Online]. Available: <https://www.deeplearningbook.org/>
- [22]. D. Singh, Mastering Tanh: A deep dive into balanced activation for machine learning, *AI-Enthusiast*, [Online]. Available: <https://medium.com/ai-enthusiast/mastering-tanh-a-deep-dive-into-balanced-activation-for-machine-learning-4734ec147dd9>
- [23]. A. I. Maverick, The role of the Gaussian distribution in machine learning, *Medium*, [Online]. Available: <https://samanemami.medium.com/the-role-of-the-gaussian-distribution-in-machine-learning-5b1c04f87967>.
- [24]. M.V. Chiến, Dự đoán sức kháng cắt của dầm FRP-RC bằng mô hình học máy CatBoost tối ưu hóa, *Tạp chí Khoa học và Công nghệ Giao thông vận tải*, 4 (2024) 13–27. <https://doi.org/10.58845/jstt.utt.2024.vn.4.3.13-27>
- [25]. M.C. Lê, Phụ lục – Các hệ số chi tiết của mô hình mạng nơ-ron, Tài liệu Google Drive, 2025. Truy cập tại: <https://tinyurl.com/2b9h3frj>