



Tạp chí điện tử  
Khoa học và Công nghệ Giao thông  
Trang website: <https://jstt.vn/index.php/vn>



## Application of Gradient Boosting combined with metaheuristic algorithms to predict the compressive strength of concrete using manufactured sand

### Article info

#### Type of article:

Original research paper

#### DOI:

<https://doi.org/10.58845/jstt.utt.2024.vn.4.3.1-12>

#### \*Corresponding author:

Email address:

[anhnh89@utt.edu.vn](mailto:anhnh89@utt.edu.vn)

**Received:** 30/06/2024

**Revised:** 28/07/2024

**Accepted:** 30/07/2024

Huu-Anh Nguyen

University of Transport Technology, 54 Trieu Khuc, Thanh Xuân, Ha Noi, 100000, Vietnam

**Abstract:** This study applies the Gradient Boosting (GB) algorithm combined with the Honey Badger Optimization (HBA) algorithm to predict the compressive strength of concrete using manufactured sand. Utilizing a dataset of 298 experimental samples, the GB-HBA model was developed, and used to analyze factors such as cement, curing age, maximum aggregate size (D<sub>max</sub>), aggregate content, sand fineness modulus, water/binder ratio, water/cement ratio, water content, sand content, and slump. The results indicate that the GB-HBA model accurately predicts compressive strength of concrete using manufactured sand, significantly enhancing the efficiency and durability of the concrete. This research introduces a novel approach for applying machine learning technology in the construction industry, contributing to sustainable development.

**Keywords:** Machine learning; compressive strength; concrete using manufactured sand; Gradient Boosting; metaheuristic algorithm.



**Thông tin bài viết**

**Dạng bài viết:**

Bài báo nghiên cứu

**DOI:**

<https://doi.org/10.58845/jstt.utt.2024.vn.4.3.1-12>

**\*Tác giả liên hệ:**

Địa chỉ Email:

[anhnh89@utt.edu.vn](mailto:anhnh89@utt.edu.vn)

**Ngày nộp bài:** 30/06/2024

**Ngày nộp bài sửa:** 28/07/2024

**Ngày chấp nhận:** 30/07/2024

## Ứng dụng thuật toán Gradient Boosting kết hợp tối ưu hóa để dự đoán cường độ nén của bê tông sử dụng cát nhân tạo

Nguyễn Hữu Anh

Trường Đại học Công nghệ Giao thông vận tải, 54 Triều Khúc, Thanh Xuân, Hà Nội, Việt Nam

**Tóm tắt:** Nghiên cứu này áp dụng thuật toán Tăng cường độ dốc (Gradient Boosting-GB) kết hợp với thuật toán tối ưu hóa Honey Badger (HBA) để dự đoán cường độ nén của bê tông sử dụng cát nhân tạo. Sử dụng cơ sở dữ liệu gồm 298 mẫu thử nghiệm, mô hình GB-HBA đã được phát triển và qua đó, đã được sử dụng phân tích các yếu tố như xi măng, tuổi bảo dưỡng,  $D_{max}$  của đá dăm, hàm lượng bột đá, mô đun độ mịn của cát, tỷ lệ nước/chất kết dính, tỷ lệ nước/xi măng, lượng nước, hàm lượng cát và độ sụt. Kết quả cho thấy mô hình GB kết hợp HBA dự đoán chính xác cường độ nén, cải thiện đáng kể hiệu quả và độ bền của bê tông. Nghiên cứu này mở ra hướng đi mới cho việc áp dụng công nghệ học máy trong ngành xây dựng, góp phần vào phát triển bền vững.

**Từ khóa:** Máy học, cường độ nén, bê tông cát nhân tạo, thuật toán tăng cường độ dốc, thuật toán tối ưu hóa.

### 1. Đặt vấn đề

Việc sử dụng cát nhân tạo trong bê tông, thay thế cho cát tự nhiên, đã thu hút sự chú ý ngày càng tăng do sự thiếu hụt cát tự nhiên phù hợp ở hầu hết các khu vực trên thế giới [1]. Cát nhân tạo thường được tạo ra từ đá cứng bằng cách nghiền, với các đặc điểm độc đáo như hình dạng lập phương và các cạnh được mài mòn, mang lại lợi thế cho các ứng dụng trong lĩnh vực kỹ thuật dân dụng [2,3]. Đặc biệt, cát nhân tạo có thể giảm thiểu các phản ứng kiềm-silica không mong muốn trong bê tông, vốn có thể gây ra các vấn đề nghiêm trọng như giãn nở và suy giảm cấu kiện bê tông [4]. Ngoài ra, việc sử dụng cát nhân tạo còn cải thiện độ bền của bê tông nhờ vào mức độ liên kết hạt cao hơn [3].

Cường độ nén, phản ánh khả năng chịu lực nén, là một thông số quan trọng xác định hiệu suất

của bê tông trong điều kiện sử dụng. Nhiều nghiên cứu [5-7] đã được thực hiện để nghiên cứu ảnh hưởng so sánh của cát nhân tạo và cát tự nhiên đến độ bền nén của bê tông. Các nghiên cứu này phần lớn cho thấy hiệu suất vượt trội của bê tông sử dụng với cát nhân tạo so với sử dụng cát tự nhiên. Bên cạnh đó, các nghiên cứu của Shen et al. [2,7] đã xác định ảnh hưởng của hình dạng cát nhân tạo và độ thô ráp lên cường độ nén của bê tông dựa trên phân tích hình ảnh số và các thử nghiệm nén thực nghiệm. Các kết quả chỉ ra rằng bột đá của cát nhân tạo có ảnh hưởng đáng kể đến hiệu suất của bê tông. Điều này cũng đã được chứng minh trong nhiều nghiên cứu thực nghiệm khác rằng bê tông cát nhân tạo có khả năng nén tốt hơn so với bê tông sử dụng cát sông tự nhiên [8,9]. Tuy nhiên, những nghiên cứu trong phòng thí nghiệm này thường phức tạp, tốn kém và mất thời

gian. Hơn nữa, các thử nghiệm trong phòng thí nghiệm thường bị giới hạn trong khoảng thời gian 90-180 ngày [10]. Mặc dù đã có nhiều nỗ lực, nhưng về mặt thực tế không thể xác định đầy đủ tất cả các sự kết hợp khác nhau của các thành phần hỗn hợp (tức là xi măng, cốt liệu thô, nước, cát nhân tạo, v.v.) đến cường độ nén của bê tông.

Trong bối cảnh đó, các thuật toán học máy (machine learning – ML) đã được áp dụng rộng rãi trong các ứng dụng kỹ thuật dân dụng, mang lại cơ hội hứa hẹn cho sự phát triển bền vững trong ngành xây dựng. Các mô hình ML có khả năng mạnh mẽ trong việc phát hiện các mối quan hệ phi tuyến và phức tạp giữa các thành phần hỗn hợp và cường độ nén của bê tông. Việc sử dụng các mô hình học máy giúp cải thiện đáng kể độ chính xác của dự đoán, tiết kiệm thời gian và chi phí so với các phương pháp thử nghiệm truyền thống. Ngoài ra, các mô hình học máy còn có khả năng tự động hóa quá trình phân tích và đưa ra dự đoán nhanh chóng dựa trên dữ liệu đầu vào, góp phần nâng cao hiệu quả và độ tin cậy.

Tuy nhiên, các thuật toán học dựa trên cây chưa được điều tra đầy đủ trong việc dự đoán cường độ nén của bê tông cát nhân tạo. Các thuật toán này, như thuật toán tăng cường độ (Gradient Boosting-GB), có tiềm năng lớn trong việc giải quyết các bài toán phức tạp nhờ khả năng kết hợp sức mạnh của nhiều cây quyết định để cải thiện độ chính xác của dự đoán. Kết hợp với các thuật toán tối ưu hoá như Honey Badger Algorithm (HBA), việc điều chỉnh các siêu tham số của mô hình có thể được thực hiện hiệu quả hơn, giúp tối ưu hóa kết quả dự đoán.

Do đó, mục tiêu chính của nghiên cứu này là nghiên cứu khả năng áp dụng thuật toán GB kết hợp với thuật toán tối ưu hoá HBA để dự đoán cường độ nén của bê tông sử dụng cát nhân tạo. Phần tiếp theo sẽ giới thiệu ngắn gọn về các thuật toán sử dụng trong nghiên cứu, tiếp theo là phần trình bày về tập dữ liệu bê tông. Kết quả dự đoán được đưa ra ở phần sau, cuối cùng là các kết luận và triển vọng nghiên cứu.

## 2. Mô hình dự báo và cơ sở dữ liệu

### 2.1. Mô hình dự báo

*Thuật toán tăng cường độ dốc (Gradient Boosting-GB)*

Thuật toán Tăng cường độ dốc (Gradient Boosting-GB) là một kỹ thuật mạnh mẽ trong học máy, được sử dụng để xây dựng mô hình dự đoán bằng cách kết hợp nhiều mô hình dự đoán yếu (weak learners), thường là cây quyết định (decision trees). Mục tiêu chính của GB là cải thiện hiệu suất của mô hình bằng cách giảm lỗi dự đoán một cách liên tục. Thuật toán hoạt động theo các bước sau: Bắt đầu với một mô hình dự đoán đơn giản, chẳng hạn như một cây quyết định nhỏ hoặc giá trị trung bình của đầu ra mục tiêu. Sau đó, sai số (hoặc phần dư) của mô hình hiện tại được tính bằng cách lấy chênh lệch giữa giá trị dự đoán và giá trị thực tế. Một mô hình mới được huấn luyện để dự đoán sai số này, và mô hình hiện tại được cập nhật bằng cách cộng thêm mô hình mới vào, thường được điều chỉnh bằng một hệ số học (learning rate) để kiểm soát mức độ ảnh hưởng của mô hình mới. Quá trình này được lặp lại cho đến khi số lượng mô hình mong muốn được xây dựng hoặc khi sai số đạt đến mức chấp nhận được.

GB có nhiều ưu điểm, bao gồm hiệu suất cao, khả năng tạo ra các mô hình chính xác bằng cách kết hợp nhiều mô hình yếu, và linh hoạt khi có thể áp dụng cho nhiều loại dữ liệu khác nhau và giải quyết cả bài toán hồi quy và phân loại. Thuật toán này cũng có khả năng xử lý các đặc trưng quan trọng một cách tự động và kiểm soát overfitting thông qua việc sử dụng các tham số như số lượng cây quyết định, độ sâu của cây và hệ số học. Tuy nhiên, Gradient Boosting cũng có một số nhược điểm như thời gian huấn luyện lâu do tính chất tuần tự của quá trình huấn luyện, nhạy cảm với tham số và dễ bị overfitting nếu không được điều chỉnh cẩn thận.

*Thuật toán tối ưu hoá Honey Badger (HBA)*

Thuật toán Honey Badger (HBA) là một thuật toán tối ưu hóa mới được đề xuất bởi Fatma A.

Hashim vào năm 2021 [11], lấy cảm hứng từ hành vi tìm kiếm thức ăn thông minh của loài lửng mật. Thuật toán này mô phỏng các chiến lược thích nghi của lửng mật trong việc xác định nguồn thức ăn, bao gồm hai chiến lược bổ sung lẫn nhau: tìm kiếm bằng khứu giác và theo dõi chim dẫn đường. Trong cả hai trường hợp, có hai giai đoạn riêng biệt: giai đoạn đào và giai đoạn mật ong, mỗi giai đoạn đóng góp đặc biệt vào quá trình tìm kiếm tổng thể.

Trong giai đoạn đào, lửng mật sử dụng khả năng khứu giác xuất sắc của mình để ước tính vị trí của con mồi. Chúng sử dụng khả năng ngửi mùi con mồi và sau đó di chuyển quanh con mồi để xác định vị trí tối ưu cho việc bắt và đào bới. Giai đoạn này đặc trưng bởi quỹ đạo di chuyển của lửng mật theo hình dạng của một cardioid. Ngược lại, trong giai đoạn mật ong, lửng mật dựa vào chim dẫn đường để xác định trực tiếp tổ ong. Khi phát hiện tổ ong, chim dẫn đường gọi lửng mật theo sau, và lửng mật, với móng vuốt mạnh mẽ của mình, hợp tác với chim dẫn đường để đột nhập vào tổ ong và lấy thức ăn cho lợi ích chung.

Cách tiếp cận thích nghi của thuật toán Honey Badger, phản ánh hành vi tìm kiếm thức ăn thông minh của lửng mật, mang lại nhiều ưu điểm. Nó cân bằng hiệu quả giữa việc thăm dò và khai thác, cho phép điều hướng hiệu quả các không gian tìm kiếm phức tạp và hội tụ đến các giải pháp chất lượng cao. Ngoài ra, việc tích hợp các chiến lược tìm kiếm đa dạng và hành vi hợp tác làm tăng tính bền bỉ và linh hoạt của thuật toán, giúp nó phù hợp với nhiều bài toán tối ưu hóa khác nhau trong nhiều lĩnh vực.

Tuy nhiên, giống như bất kỳ thuật toán tối ưu hóa nào, HBA có thể yêu cầu điều chỉnh tham số cẩn thận để đạt hiệu suất tối ưu, và hiệu quả của nó có thể bị ảnh hưởng bởi các yếu tố như kích thước quần thể và tiêu chí lựa chọn.

## 2.2. Cơ sở dữ liệu

Trong nghiên cứu này, mục tiêu đầu ra là cường độ nén của bê tông cát nhân tạo ở các tuổi bảo dưỡng khác nhau từ 3 đến 388 ngày. Bộ cơ sở dữ liệu gồm 298 mẫu thí nghiệm được thu thập

từ các thử nghiệm có sẵn trong tài liệu [12-15]. Trong bộ dữ liệu này, cả cường độ nén khối lập phương và xi lanh đều được thu thập. Một hệ số 0.82 đã được áp dụng để chuyển đổi các giá trị này nhằm sử dụng cho việc phát triển mô hình máy học (ML).

Các mô hình ML được phát triển để dự đoán cường độ nén của bê tông cát nhân tạo sử dụng mười một thuộc tính được ghi nhận làm thông số đầu vào, bao gồm: cường độ nén của xi măng (MPa), cường độ kéo của xi măng (MPa), tuổi bảo dưỡng (ngày),  $D_{max}$  của đá nghiền (mm), hàm lượng bột đá (%), mô đun độ mịn của cát (FM), tỷ lệ nước/chất kết dính, tỷ lệ nước/xi măng, nước ( $kg/m^3$ ), hàm lượng cát (%) và độ sụt (mm). Lượng bột đá thích hợp trong cát nhân tạo giúp cải thiện tính công tác của MSC. Điều này là do độ đặc của hồ xi măng được tăng cường bởi sự hấp thụ nước cao của bột đá và thể tích lớn hơn của hồ xi măng chứa bột đá nghiền. Cường độ nén của bê tông dạng khối lập phương là đầu ra của mô hình. Bảng 1 tóm tắt thông tin thống kê của các thông số đầu vào và đầu ra sử dụng trong nghiên cứu. Bên cạnh đó, Hình 1 minh họa sự phân bố dữ liệu của các thông số và tần suất xuất hiện (số lượng mẫu) của chúng trong bộ dữ liệu được sử dụng.

## 2.3. Đánh giá khả năng dự báo của các mô hình ML

Để đánh giá hiệu suất của các mô hình học máy trong việc dự đoán cường độ nén của bê tông cát nhân tạo, các tiêu chí thống kê phổ biến như  $R^2$ , RMSE và MAE thường được sử dụng. Mỗi tiêu chí cung cấp một góc nhìn khác nhau về độ chính xác và hiệu quả của mô hình. Hệ số xác định, hay  $R^2$ , đo lường mức độ mà mô hình dự đoán giải thích được biến thiên của dữ liệu thực tế.  $R^2$  có giá trị từ 0 đến 1, trong đó  $R^2 = 1$  nghĩa là mô hình dự đoán hoàn hảo, tất cả các điểm dữ liệu thực tế đều nằm trên đường dự đoán, còn  $R^2 = 0$  nghĩa là mô hình không giải thích được bất kỳ biến thiên nào của dữ liệu thực tế, dự đoán trung bình của mô hình không tốt hơn việc dự đoán bằng giá trị trung bình của dữ liệu. Căn của sai số bình phương trung

bình (RMSE), đo lường độ lệch chuẩn của sai số dự đoán, tức là mức độ mà các giá trị dự đoán của mô hình phân tán xung quanh các giá trị thực tế. RMSE càng nhỏ, mô hình càng chính xác. Sai số tuyệt đối trung bình (MAE) đo lường trung bình của

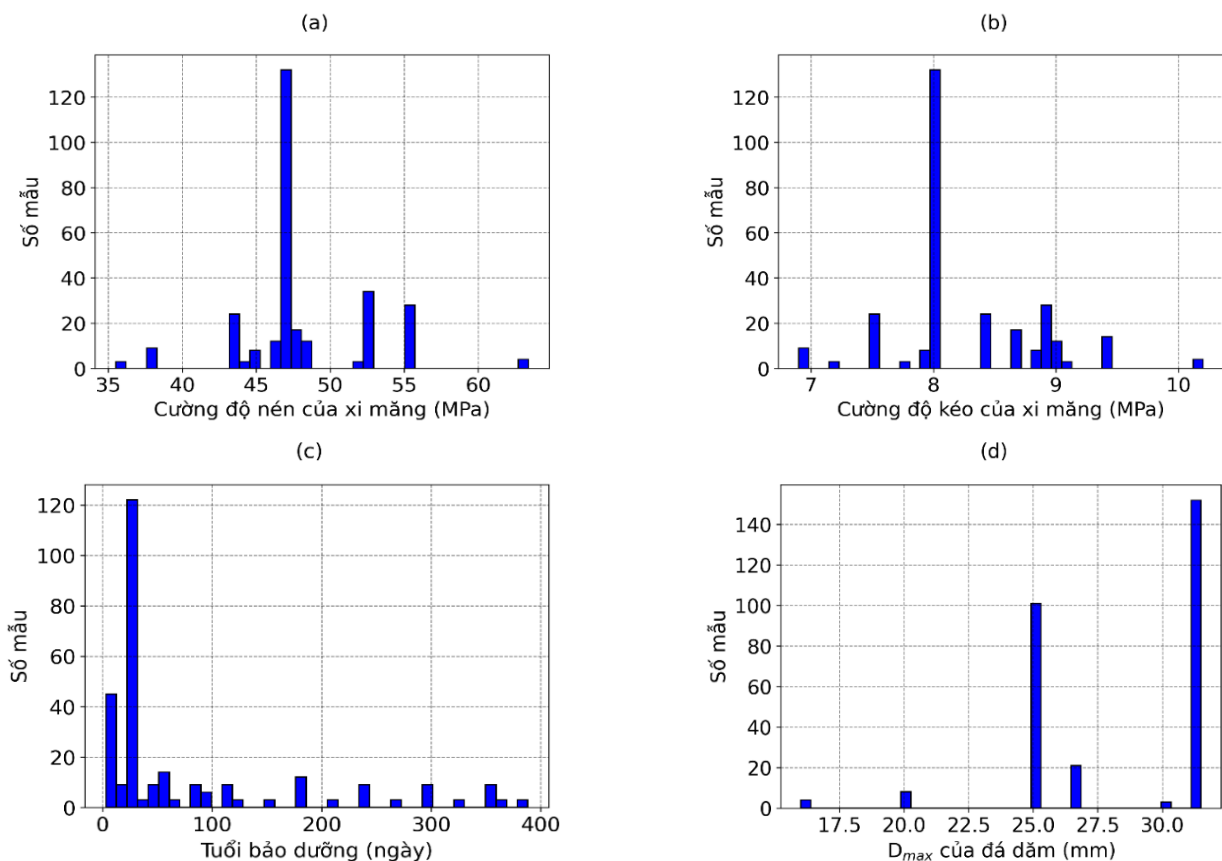
các sai số tuyệt đối giữa các giá trị dự đoán và giá trị thực tế, cung cấp một cách đánh giá dễ hiểu về độ chính xác của mô hình dự đoán. Công thức xác định  $R^2$ , RMSE, MAE được trình bày trong tài liệu tham khảo [16].

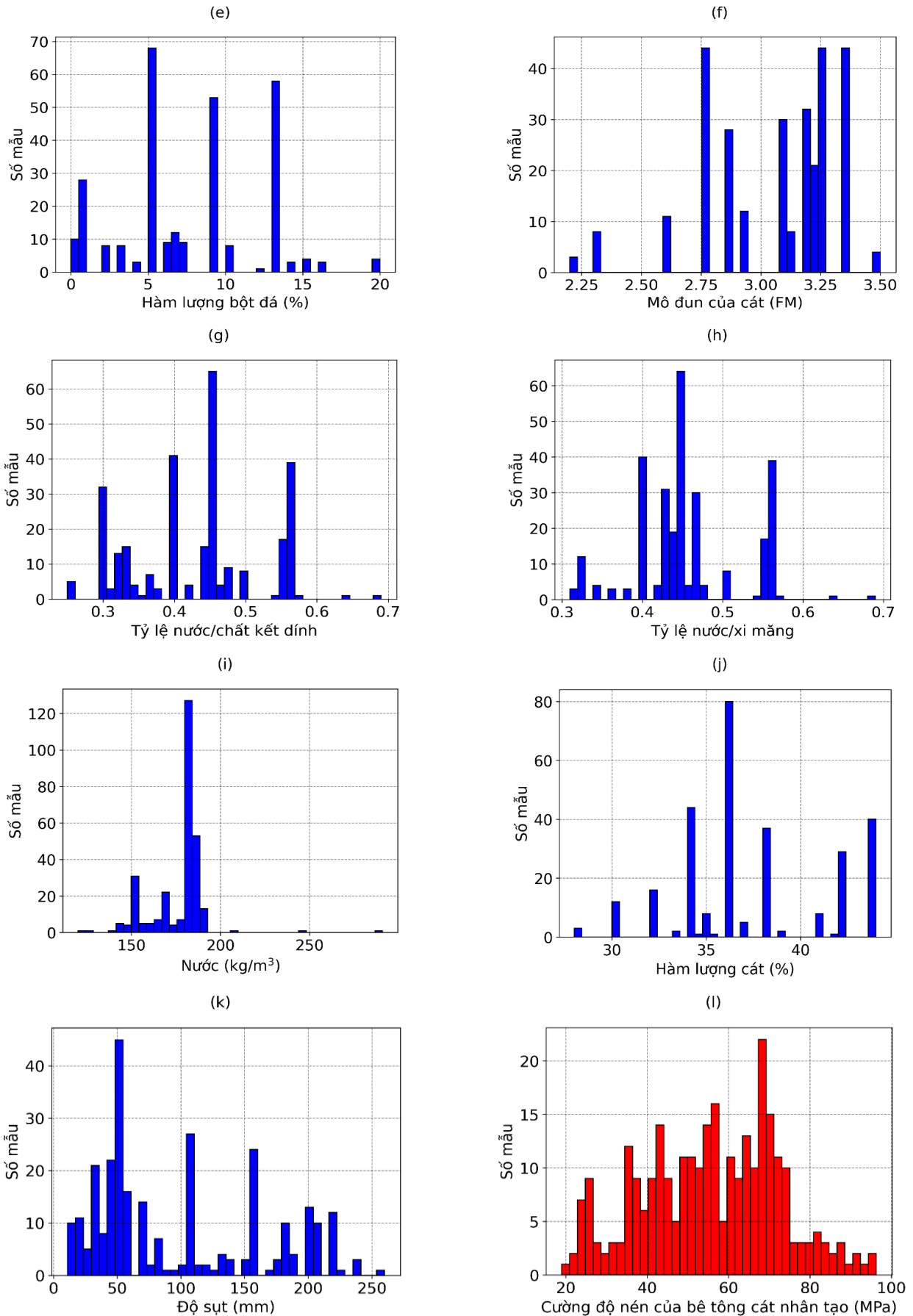
**Bảng 1.** Phân tích thống kê của cơ sở dữ liệu

Thông số	Cường độ nén xi măng	Cường độ kéo xi măng	Tuổi bảo dưỡng	$D_{max}$ đá nghiền	Hàm lượng bột đá	Mô đun độ mịn cát
Đơn vị	MPa	MPa	ngày	mm	%	[FM]
Ký hiệu	$l_1$	$l_2$	$l_3$	$l_4$	$l_5$	$l_6$
Min	35.50	6.90	3.00	16.00	0.00	2.20
Trung bình	47.95	8.25	80.93	28.31	7.54	3.06
Trung vị	46.80	8.00	28.00	31.50	6.60	3.15
Max	63.40	10.20	388.00	31.50	20.00	3.50
Độ lệch chuẩn	4.29	0.60	102.36	3.68	4.48	0.27

Parameter	Tỷ lệ nước/chất kết dính	Tỷ lệ nước/xi măng	Nước	Hàm lượng cát	Độ sụt	Cường độ nén
Đơn vị	[-]	[-]	kg/m <sup>3</sup>	%	mm	MPa
Ký hiệu	$l_7$	$l_8$	$l_9$	$l_{10}$	$l_{11}$	$Y$
Min	0.25	0.31	120.00	28.00	11.00	19.00
Trung bình	0.43	0.46	175.49	37.23	98.34	55.80
Trung vị	0.45	0.45	180.00	36.00	70.00	56.45
Max	0.69	0.69	291.00	44.00	260.00	96.30
Độ lệch chuẩn	0.09	0.07	15.16	4.00	66.64	16.70





Hình 1. Biểu đồ phân phối của các thông số trong nghiên cứu

### 3. Kết quả và thảo luận

#### 3.1. Tối ưu hoá siêu tham số mô hình GB

Trong nghiên cứu này, mô hình học máy dựa trên thuật toán độ dốc tăng cường (GB) được đề xuất để ước lượng cường độ nén của bê tông cát nhân tạo. Thực tế, mỗi mô hình học máy có nhiều siêu tham số, và việc lựa chọn các siêu tham số này ảnh hưởng đáng kể đến hiệu suất của các mô hình học máy cũng như khả năng học hỏi và tổng quát hóa từ dữ liệu. Do đó, tối ưu hóa các siêu tham số là điều cần thiết để nâng cao hiệu suất của các mô hình ML. Các kỹ thuật như tìm kiếm lưới (grid search), tìm kiếm ngẫu nhiên (random search), tối ưu hóa Bayesian hoặc các thuật toán tối ưu hóa thường được sử dụng để xác định bộ siêu tham số tốt nhất cho một vấn đề cụ thể.

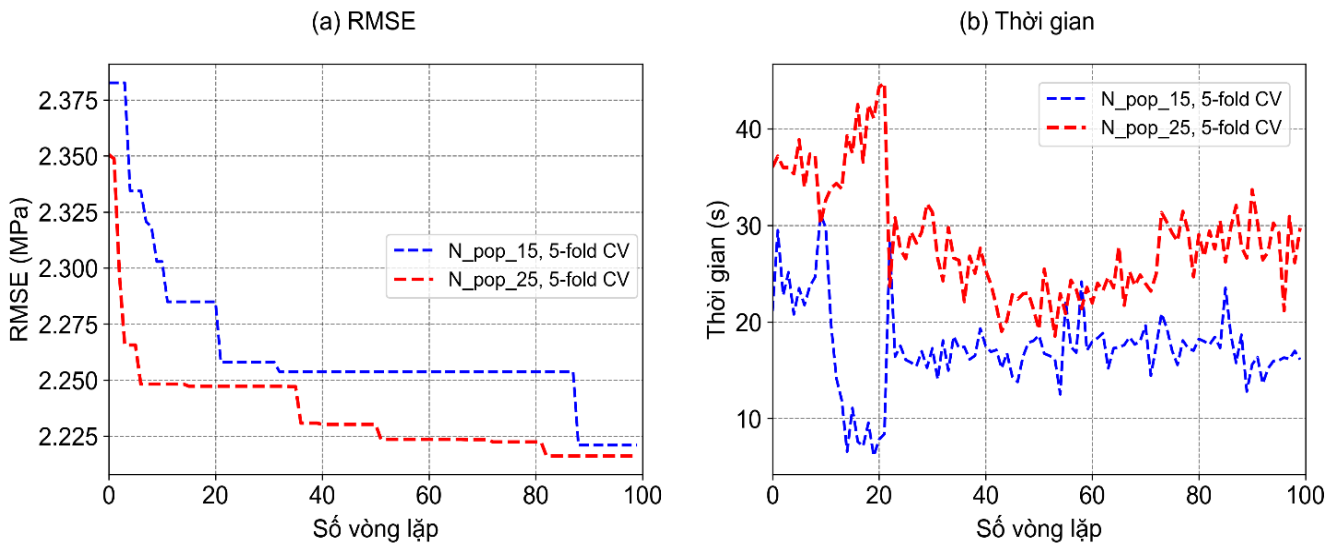
Trong nghiên cứu này, Thuật toán Honey Badger (HBA) được sử dụng để điều chỉnh các siêu tham số và cải thiện độ chính xác của mô hình. Ba siêu tham số được chứng minh có ảnh hưởng đến mô hình GB được chọn để điều chỉnh, cụ thể là "n\_estimators", "learning rate", và "max\_depth". Trong đó, "n\_estimators" là tham số quyết định số lượng các cây quyết định được tạo ra trong mô hình, "learning rate" là tham số quyết định độ lớn của các bước đi trong quá trình tối ưu hóa. Nó kiểm soát độ lớn của giá trị được thêm vào mô hình ở mỗi bước để cải thiện chất lượng dự đoán, "max\_depth" là tham số quyết định độ sâu tối đa của các cây quyết định trong mô hình. Không gian tìm kiếm cho các siêu tham số này được định nghĩa như sau: "n\_estimators" từ 10 đến 300, "learning\_rate" từ 0.01 đến 0.3, và "max\_depth" từ 3 đến 8. Để tránh trường hợp quá khớp, phương pháp xác thực chéo 5 lần được áp dụng trong quá trình điều chỉnh siêu tham số. Đầu tiên, tập dữ liệu được chia thành các tập dữ liệu huấn luyện và kiểm chứng, trong đó, tập dữ liệu huấn luyện chiếm 70% dữ liệu trong bộ dữ liệu, và tập dữ liệu kiểm chứng chiếm 30% dữ liệu còn lại. Sự phân chia này đảm bảo rằng mô hình được huấn luyện trên một phần đủ lớn của tập dữ liệu để nắm bắt các mẫu và mối quan hệ cơ bản, đồng thời được kiểm

tra trên dữ liệu chưa từng thấy để đánh giá khả năng tổng quát hóa của nó. Sau đó, xác thực chéo 5 lần được thực hiện dựa trên ba bước sau: (1) Tập dữ liệu huấn luyện được chia ngẫu nhiên thành 5 phần bằng nhau, (2) Đào tạo mô hình bằng cách sử dụng 4 phần và xác nhận mô hình dựa trên phần còn lại, (3) Lặp lại bước (2) 5 lần để thu được 5 số chỉ số hiệu suất. Hiệu suất cuối cùng của mô hình được lấy làm giá trị trung bình của 5 chỉ số hiệu suất. Tập dữ liệu xác thực tạo ra từ kỹ thuật xác thực chéo 5 lần được sử dụng trong quá trình điều chỉnh siêu tham số và đánh giá mô hình. Trong nghiên cứu này, mục tiêu của các bài toán tối ưu hóa là giảm sai số giữa kết quả thực nghiệm và kết quả dự đoán. Do đó, sai số RMSE được sử dụng để đánh giá hiệu suất của mô hình GB trong quá trình điều chỉnh siêu tham số. Bên cạnh đó, để nâng cao hơn nữa hiệu quả của quá trình tối ưu hóa bằng cách sử dụng thuật toán tối ưu HBA, cần đưa ra sự lựa chọn những tham số cho thuật toán tối ưu này. Thông thường với các thuật toán tối ưu hóa, hai tham số quan trọng nhất là kích thước quần thể tham gia quá trình dự báo ( $N_{pop}$ ) và số vòng lặp dừng để điều chỉnh siêu tham số. Trong nghiên cứu này, các giá trị cho kích thước quần thể được quét trong phạm vi 15 và 25 để chọn giá trị phù hợp. Số lượng vòng lặp được chọn là 100. Kết quả mô phỏng với các điều kiện khác nhau được thể hiện ở Hình 2. Hơn nữa, Bảng 2 cung cấp các giá trị tiêu chí RMSE chi tiết trên tập dữ liệu được xác thực chéo 5 lần, thời gian tối ưu hóa và các giá trị siêu tham số tốt nhất tương ứng cho mỗi mô hình ML.

Có thể thấy rằng, khi số vòng lặp thay đổi từ 1 đến 100 thì hàm sai số RMSE có sự thay đổi, rõ rệt. Đến khoảng vòng lặp thứ 82, kết quả RMSE hội tụ về giá trị 2.216 tương ứng với  $N_{pop} = 25$ , trong khi đó với trường hợp  $N_{pop} = 15$ , RMSE hội tụ về giá trị 2.221 sau khoảng 85 vòng lặp. Mặc dù thời gian thực hiện điều chỉnh siêu tham số lâu hơn đối với  $N_{pop} = 25$  so với  $N_{pop} = 15$ , hiệu suất dự báo tốt hơn với giá trị RMSE thấp hơn chứng tỏ rằng việc tăng số lượng cá thể trong quá trình tìm

kiểm mang lại kết quả dự đoán chính xác hơn. Điều này cho thấy rằng, mặc dù có sự gia tăng thời gian tính toán, việc chọn  $N_{pop} = 25$  vẫn đáng giá vì

mang lại độ chính xác cao hơn cho mô hình. Do đó, bộ siêu tham số tốt nhất được chọn cho mô hình GB tương ứng với  $N_{pop} = 25$ .



**Hình 2.** Kết quả tối ưu hoá mô hình GB (a) đánh giá theo tiêu chí RMSE, và (b) thời gian thực hiện mô phỏng theo số vòng lặp

**Bảng 2.** Giá trị tiêu chí RMSE trên bộ dữ liệu xác thực chéo 5 lần, thời gian thực hiện tối ưu hóa và các giá trị siêu tham số được tinh chỉnh tương ứng với mô hình GB

Model	RMSE (MPa)	Thời gian (s)	n_estimators	learning_rate	max_depth
$N_{pop} = 15$	2.221	1727.34	262	0.122	4
$N_{pop} = 25$	2.216	2847.73	300	0.082	4

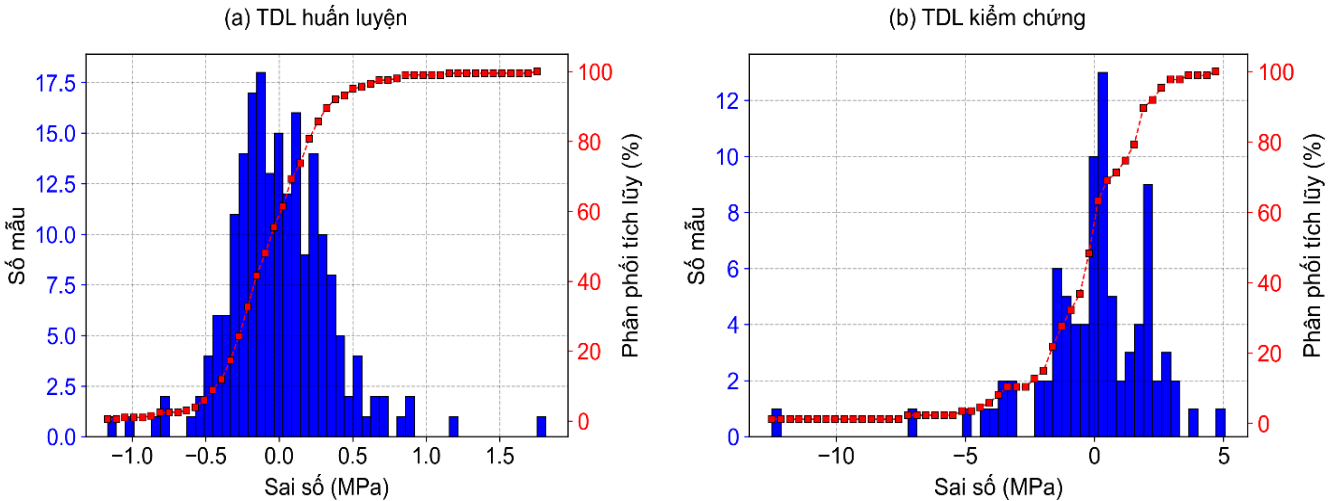
**3.2. Hiệu suất dự báo cường độ nén của bê tông cát nhân tạo theo mô hình GB tốt nhất**

Trình bày kết quả điển hình của mô hình GB tốt nhất (ký hiệu là GB\_25) để dự báo cường độ nén của bê tông cát nhân tạo. Biểu đồ phân bố tích lũy và tần suất của sai số khi dự đoán bằng mô hình GB\_25 được biểu diễn trên Hình 3. Với tập dữ liệu huấn luyện (Hình 3a), có thể thấy phần lớn sai số nằm trong phạm vi từ -0.5 đến 0.5 MPa (khoảng 91.38%), và chỉ có 8.16% số lượng mẫu nằm ngoài phạm vi này với sai số lớn nhất là xấp xỉ 2 MPa. Tập dữ liệu kiểm chứng có phạm vi sai số lớn hơn, với các mẫu có sai số phần lớn trong phạm vi từ -5 đến 5MPa (chiếm 96.63%), và chỉ có 3.37% số lượng mẫu có sai số nằm ngoài phạm vi này với sai số lớn nhất có trị số khoảng 12 MPa.

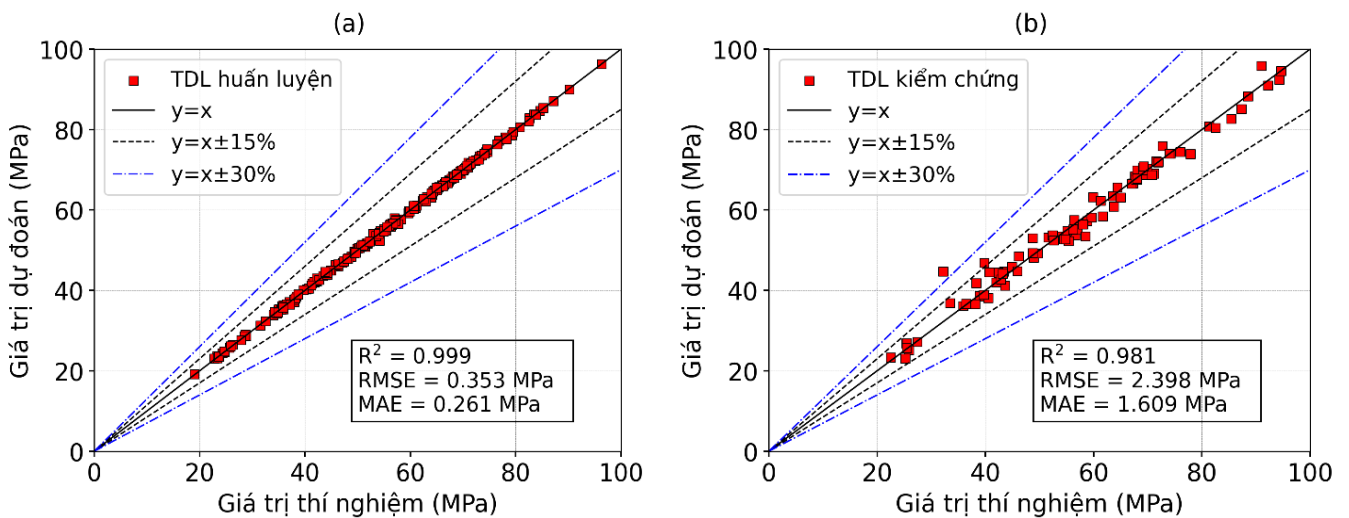
Mối tương quan giữa các giá trị dự đoán do

mô hình tạo ra và kết quả thử nghiệm thực tế được hiển thị trên biểu đồ hồi quy, như được nêu trong Hình 4. Trục x biểu thị kết quả thử nghiệm và trục y biểu thị các giá trị dự đoán thu được từ mô hình GB\_25. Trong Hình 4, đường liền nét màu đen biểu thị sự liên kết gần như hoàn hảo giữa kết quả mong đợi và kết quả thực tế, cho thấy dự đoán của mô hình có độ chính xác cao. Ngoài ra, các đường đứt nét màu đen và xanh lần lượt đại diện cho đường viền của sai số  $\pm 15\%$  và  $\pm 30\%$ , cung cấp thông tin chi tiết về hiệu suất của mô hình trong phạm vi biên độ lỗi có thể chấp nhận được. Có thể thấy rõ rằng mô hình GB\_25 thể hiện hiệu suất vượt trội trên các tập dữ liệu huấn luyện và kiểm chứng. Hầu hết các điểm dữ liệu thể hiện sự phân cụm chặt chẽ xung quanh đường  $y=x$ , cho thấy mức độ phù hợp cao giữa giá trị dự đoán và giá trị thực tế.





**Hình 3.** Biểu đồ sai số giữa giá trị dự đoán và giá trị thực nghiệm



**Hình 4.** Biểu đồ hồi quy giữa giá trị dự đoán và giá trị thực nghiệm

**3.3. Phân tích ảnh hưởng của các thông số đến cường độ nén của bê tông cát nhân tạo**

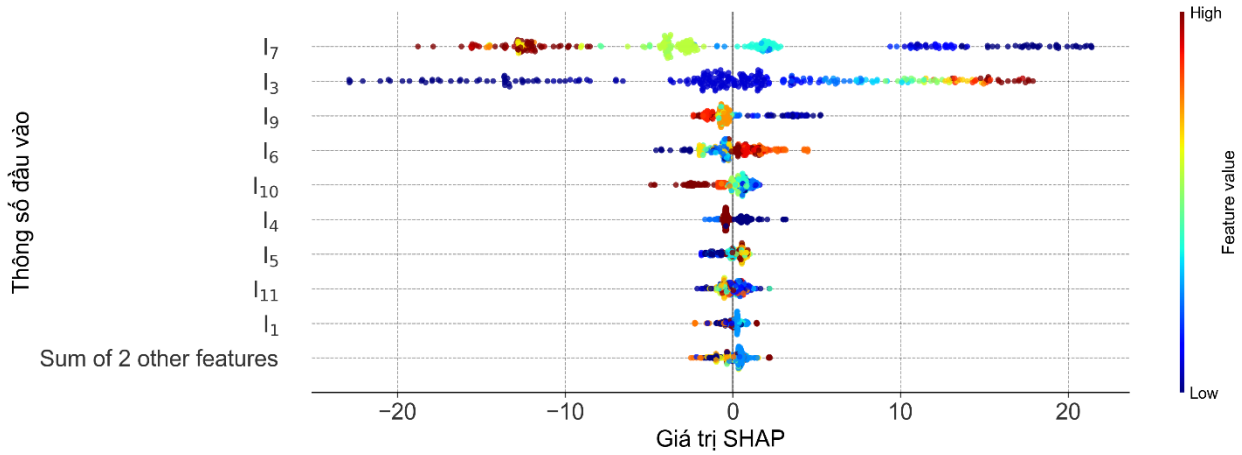
Để đánh giá mức độ ảnh hưởng của các thông số đầu vào đến cường độ nén của bê tông cát nhân tạo, nghiên cứu này sử dụng phân tích SHAP (SHapley Additive exPlanations). SHAP là một phương pháp giải thích mô hình học máy bằng cách tính toán mức độ đóng góp của từng đặc trưng đầu vào đối với dự đoán của mô hình. Giá trị SHAP càng cao (dương hoặc âm) thì đặc trưng đó càng có ảnh hưởng lớn đến kết quả dự đoán. Hình 5 trình bày biểu đồ phân tích SHAP, trong đó trục X biểu thị giá trị SHAP và màu sắc biểu diễn giá trị của các thông số đầu vào. Giá trị trung bình tuyệt đối của giá trị SHAP trên toàn bộ tập dữ liệu được sử dụng để đánh giá tầm quan trọng của từng đặc trưng, và được sắp xếp theo thứ tự từ trên xuống

dưới (Hình 5).

Các thông số đầu vào được biểu diễn từ  $I_1$  đến  $I_{11}$ . Những thông số này có thể đại diện cho các biến khác nhau được sử dụng trong mô hình dự đoán (Bảng 1). Phân tích SHAP cho thấy, biến  $I_7$  có nhiều giá trị SHAP dương và màu sắc chủ yếu là đỏ và xanh lá, điều này có nghĩa rằng các giá trị cao của biến  $I_7$  (màu đỏ) có xu hướng làm tăng giá trị dự đoán của mô hình. Ngược lại, các giá trị thấp của biến  $I_7$  (màu xanh) thường làm giảm giá trị dự đoán. Biến  $I_3$  chủ yếu có giá trị SHAP âm, đặc biệt với màu xanh, chỉ ra rằng các giá trị thấp của biến  $I_3$  có xu hướng làm giảm giá trị dự đoán của mô hình. Biến  $I_9$  có xu hướng giảm giá trị dự đoán khi giá trị của nó tăng lên (giá trị cao màu đỏ bên phần giá trị SHAP âm). Các biến  $I_6$  và  $I_{10}$  này cho thấy ảnh hưởng đa chiều, với cả giá trị SHAP dương và

âm, điều này có nghĩa rằng ảnh hưởng của chúng đến dự đoán phụ thuộc vào các giá trị cụ thể của chúng. Cuối cùng, các biến còn lại có giá trị SHAP phân bố quanh trục 0, chỉ ra rằng chúng có ảnh hưởng tương đối thấp hoặc trung bình đến giá trị

dự đoán của mô hình. Biểu đồ SHAP cung cấp một cách trực quan để hiểu rõ hơn về sự đóng góp của từng thông số đầu vào đối với dự đoán của mô hình và giúp xác định các yếu tố quan trọng nhất ảnh hưởng đến kết quả dự báo.



Hình 5. Phân tích giá trị SHAP và sắp xếp tầm quan trọng của các thông số đầu vào

### 3.4. GUI



Hình 6. Giao diện GUI được phát triển trong nghiên cứu

Hình 6 cung cấp một giao diện người dùng (GUI) được thiết kế để nhập các thông số đầu vào và dự đoán cường độ nén của bê tông sử dụng cát nhân tạo. Giao diện này bao gồm các thanh trượt (sliders) cho phép người dùng điều chỉnh các giá trị của từng thông số đầu vào. Người dùng có thể dễ dàng điều chỉnh các giá trị thông số bằng cách kéo các thanh trượt. Điều này giúp kiểm tra nhanh các kịch bản khác nhau và thấy được ảnh hưởng của từng thông số đầu vào đến cường độ nén của bê tông.

#### 4. Kết luận và kiến nghị

Trong nghiên cứu này, chúng tôi đã áp dụng các thuật toán học máy, đặc biệt là thuật toán Tăng cường độ dốc (Gradient Boosting-GB) kết hợp với thuật toán tối ưu hóa Honey Badger (HBA), để dự đoán cường độ nén của bê tông sử dụng cát nhân tạo. Kết quả cho thấy rằng mô hình kết hợp này có thể dự đoán chính xác cường độ nén của bê tông với độ tin cậy cao.

Cụ thể, các yếu tố như cường độ nén và kéo của xi măng, tuổi bảo dưỡng,  $D_{max}$  của đá dăm, hàm lượng bột đá, mô đun độ mịn của cát, tỷ lệ nước/chất kết dính, tỷ lệ nước/xi măng, lượng nước, hàm lượng cát và độ sụt đều được chứng minh là có ảnh hưởng đáng kể đến cường độ nén của bê tông. Mô hình GB kết hợp HBA không chỉ cung cấp dự đoán chính xác mà còn giúp tối ưu hóa các thông số đầu vào, góp phần nâng cao hiệu quả và độ bền của bê tông cát nhân tạo.

Nghiên cứu này mở ra hướng đi mới trong việc áp dụng công nghệ học máy và thuật toán tối ưu hóa vào ngành xây dựng, đặc biệt là trong việc cải thiện và dự đoán chất lượng của bê tông sử dụng cát nhân tạo. Với các kết quả tích cực đạt được, việc áp dụng rộng rãi các công nghệ này sẽ góp phần quan trọng vào sự phát triển bền vững của ngành xây dựng. Các doanh nghiệp xây dựng và các kỹ sư nên xem xét việc áp dụng mô hình dự đoán cường độ nén sử dụng cát nhân tạo vào thực tiễn để tối ưu hóa quy trình sản xuất bê tông, giảm thiểu chi phí và nâng cao chất lượng công trình. Cần tiếp tục phát triển và hoàn thiện giao diện

người dùng để dễ sử dụng hơn, cung cấp thêm các tính năng như dự đoán theo thời gian thực và phân tích độ nhạy, giúp người dùng có cái nhìn toàn diện hơn về ảnh hưởng của các thông số đầu vào.

#### Tài liệu tham khảo

- [1]. P. Nanthagopalan, M. Santhanam. (2011). Fresh and hardened properties of self-compacting concrete produced with manufactured sand. *Cement and Concrete Composites*, 33, 353-358.
- [2]. W. Shen, Z. Yang, L. Cao, L. Cao, Y. Liu, H. Yang, Z. Lu, J. Bai. (2016). Characterization of manufactured sand: Particle shape, surface texture and behavior in concrete. *Construction and Building Materials*, 114, 595-601.
- [3]. T. Ji, C.-Y. Chen, Y.-Z. Zhuang, J.-F. Chen. (2013). A mix proportion design method of manufactured sand concrete based on minimum paste theory. *Construction and Building Materials*, 44, 422-426.
- [4]. R.N. Swamy. (1992). The alkali-silica reaction in concrete, Blackie Glasgow. <https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780203036631&type=googlepdf> (accessed June 12, 2024).
- [5]. M. Guan, Z. Lai, Q. Xiao, H. Du, K. Zhang. (2019). Bond behavior of concrete-filled steel tube columns using manufactured sand (MS-CFT). *Engineering Structures*, 187, 199-208.
- [6]. B. Li, J. Wang, M. Zhou. (2009). Effect of limestone fines content in manufactured sand on durability of low-and high-strength concretes. *Construction and Building Materials*, 23, 2846-2850.
- [7]. W. Shen, Y. Liu, Z. Wang, L. Cao, D. Wu, Y. Wang, X. Ji. (2018). Influence of manufactured sand's characteristics on its concrete performance. *Construction and Building Materials*, 172, 574-583.
- [8]. H. Donza, O. Cabrera, E.F. Irassar. (2002). High-strength concrete with different fine aggregate. *Cement and Concrete Research*,

- 32, 1755-1761.
- [9]. S. Park. (2012). Study on the Fluidity and Strength Properties of High Performance Concrete Utilizing Crushed Sand. *International Journal of Concrete Structures and Materials*, 6, 231-237.
- [10]. S.L. Mak, K. Torii. (1995). Strength development of high strength concretes with and without silica fume under the influence of high hydration temperatures. *Cement and Concrete Research*, 25, 1791-1802.
- [11]. F.A. Hashim, E.H. Houssein, K. Hussain, M.S. Mabrouk, W. Al-Atabany. (2022). Honey Badger Algorithm: New metaheuristic algorithm for solving optimization problems. *Mathematics and Computers in Simulation*, 192, 84-110.
- [12]. X. Ding, C. Li, Y. Xu, F. Li, S. Zhao. (2016). Experimental study on long-term compressive strength of concrete with manufactured sand. *Construction and Building Materials*, 108, 67-73.
- [13]. X. Ding, C. Li, Y. Xu, F. Li, S. Zhao. (2016). Dataset of long-term compressive strength of concrete with manufactured sand. *Data in Brief* 6, 959-964.
- [14]. S. Zhao, X. Ding, M. Zhao, C. Li, S. Pei. (2017). Experimental study on tensile strength development of concrete with manufactured sand. *Construction and Building Materials*, 138, 247-253.
- [15]. S. Zhao, F. Hu, X. Ding, M. Zhao, C. Li, S. Pei. (2017). Dataset of tensile strength development of concrete with manufactured sand. *Data in Brief*, 11, 469-472.
- [16]. H.-B. Ly, E. Monteiro, T.-T. Le, V.M. Le, M. Dal, G. Regnier, B.T. Pham. (2019). Prediction and sensitivity analysis of bubble dissolution time in 3D selective laser sintering using ensemble decision trees. *Materials*, 12, 1544.