

# Dự đoán cường độ nén của bê tông tro bay bằng mô hình máy học kết hợp dựa trên thuật toán cây quyết định

■ TS. MAI THỊ HẢI VÂN

Trường Đại học Công nghệ Giao thông vận tải

**TÓM TẮT:** Bê tông tro bay (Fly ash concrete - FAC) hiện được sử dụng rộng rãi do có nhiều ưu điểm so với bê tông truyền thống như cải thiện cường độ, độ bền, tăng khả năng làm việc, giảm nứt. Hơn nữa, sử dụng tro bay để thay thế một phần chất kết dính trong bê tông sẽ tận dụng được một lượng lớn tro bay (là một trong các chất thải công nghiệp) góp phần giảm ô nhiễm môi trường, tiết kiệm chi phí. Trong số các đặc tính cơ học của FAC thì cường độ chịu nén là một trong những đặc tính quan trọng nhất. Nghiên cứu này phát triển mô hình máy học kết hợp dựa trên thuật toán cây quyết định và sử dụng kỹ thuật tăng cường để dự báo cường độ chịu nén của FAC. Hiệu suất dự báo của mô hình EBT được thể hiện thông qua 3 tiêu chí đánh giá là hệ số tương quan  $R = 0,946$ , căn của sai số toàn phương trung bình  $RMSE = 4,281$  MPa và sai số tuyệt đối trung bình  $MAE = 3,053$  MPa. Kết quả cho thấy mô hình EBT là một công cụ hiệu quả trong việc dự đoán cường độ chịu nén của FAC, giúp các kỹ sư tiết kiệm thời gian, giảm chi phí thí nghiệm.

**TỪ KHÓA:** Mô hình máy học kết hợp, cây quyết định (DT), tro bay (FA), cường độ chịu nén (CS).

**ABSTRACT:** Fly ash concrete (FAC) is increasingly commonly utilized due to its several advantages over conventional concrete, including greater strength, durability, workability, and crack resistance. Additionally, by using fly ash to replace the binder in concrete partially, a substantial amount of fly ash (industrial waste) will be utilized, helping to environmental pollution reduction and economic savings. Compressive strength is one of the most critical mechanical qualities of FAC. This study aims to construct an Ensemble model based on the Decision trees method and use the Ensemble boosted trees (EBT) technique to predict the compressive strength of FAC. The predictive performance of the EBT model is shown through 3 evaluation criteria: correlation coefficient ( $R=0.946$ ), root mean square error ( $RMSE=4.281$  MPa) and mean absolute error ( $MAE=3.053$  MPa). The results indicate that the EBT

model is a useful tool for estimating the compressive strength of the FAC, helping engineers save time and reduce testing costs.

**KEYWORDS:** Ensemble machine learning, Decision trees (DT), Fly ash (FA), Compressive strength (CS).

## 1. ĐẶT VẤN ĐỀ

Ngày nay, nền công nghiệp phát triển đã tạo ra một lượng lớn chất thải công nghiệp, gây ảnh hưởng nghiêm trọng đến môi trường, trong đó, phải kể đến một trữ lượng lớn tro bay (FA). FA là một loại bột mịn, là sản phẩm phụ của quá trình đốt than nghiền thành bột trong các nhà máy điện và được ứng dụng trong công nghiệp sản xuất bê tông. Việc tận dụng những sản phẩm phụ của ngành công nghiệp như FA sẽ góp phần giảm thiểu ô nhiễm môi trường, tiết kiệm chi phí sản xuất bê tông. Quan trọng hơn, với hàm lượng tro bay thay thế phù hợp có thể có thể cải thiện đáng kể các đặc tính của FAC như làm tăng cường độ nén, tăng khả năng làm việc, giảm nứt. Với những ưu điểm này, FAC hiện nay được nghiên cứu và sử dụng rộng rãi trong các công trình xây dựng dân dụng, thủy lợi và đặc biệt là trong xây dựng đường giao thông. Để sử dụng FAC trong xây dựng công trình, việc xác định CS rất quan trọng.

Hiện nay, CS của FAC được xác định dựa trên các thí nghiệm [1]. Tuy nhiên, quy trình thí nghiệm thường rất tốn kém cả về thời gian và chi phí. Các công thức thực nghiệm có thể cho kết quả một cách nhanh chóng nhưng là kết quả gần đúng và nó chỉ có thể tính toán được với một số lượng giới hạn các tham số khi xây dựng các hàm quan hệ. Do đó, một cách tiếp cận mới cần được phát triển để giảm tiêu tốn thời gian cũng như chi phí thử nghiệm và đặc biệt là trong trường hợp nghiên cứu có nhiều tham số đầu vào.

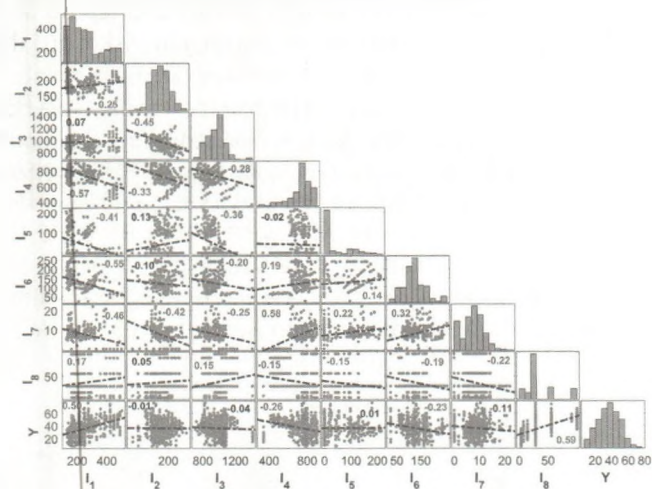
Trong những thập kỷ gần đây, phương pháp tiếp cận trí tuệ nhân tạo (AI) đã được sử dụng rộng rãi để dự đoán nhiều vấn đề trong các lĩnh vực khoa học và kỹ thuật. Đặc biệt, để dự đoán cường độ chịu nén của bê tông tro bay đã có các nghiên cứu [2, 3]. Tuy nhiên, các nghiên cứu này đều xây dựng mô hình dự báo dưới dạng mô hình đơn lẻ, chưa có nghiên cứu nào sử dụng mô hình máy học kết hợp, không sử dụng kỹ thuật xác thực chéo trong quá trình huấn luyện mô hình để nâng cao độ tin cậy của kết quả dự báo.

Nghiên cứu này phát triển mô hình máy học kết hợp (Ensemble) dựa trên thuật toán cây quyết định và sử dụng kỹ thuật tăng cường (Ensemble Boosted Trees - EBT) để dự báo cường độ chịu nén của FAC, sử dụng 622 dữ liệu thử nghiệm với 8 thông số đầu vào. Đóng góp của bài báo là đã sử dụng thuật toán Ensemble - một phương pháp kết hợp các thuật toán yếu để tạo ra một thuật toán mạnh hơn nhằm nâng cao hiệu suất dự đoán và tính ổn định của mô hình. Tiếp theo, kỹ thuật xác thực chéo cũng được áp dụng nhằm hạn chế hiện tượng quá khớp (overfitting) trong quá trình đào tạo mô hình.

**2. CƠ SỞ DỮ LIỆU**

Để phát triển mô hình dự báo cường độ CS của FAC, nghiên cứu đã sử dụng một bộ dữ liệu gồm 622 kết quả thí nghiệm được thu thập từ 4 bài báo khoa học được đăng trên các tạp chí uy tín [4-7]. Mô hình EBT được xây dựng với 8 thông số đầu vào tương ứng với ký hiệu từ  $I_1$  đến  $I_8$  là các yếu tố ảnh hưởng đến CS của FAC, 1 thông số đầu ra là CS (MPa), ký hiệu là Y. Trong đó,  $I_1$  là hàm lượng xi măng ( $kg/m^3$ ),  $I_2$  là hàm lượng nước ( $kg/m^3$ ),  $I_3$  là hàm lượng cốt liệu thô ( $kg/m^3$ ),  $I_4$  là hàm lượng cốt liệu mịn ( $kg/m^3$ ),  $I_5$  là hàm lượng xỉ lò cao ( $kg/m^3$ ),  $I_6$  là hàm lượng tro bay ( $kg/m^3$ ),  $I_7$  là hàm lượng phụ gia siêu dẻo ( $kg/m^3$ ),  $I_8$  tuổi của mẫu bê tông (ngày). Hình 2.1 thể hiện phân bố của các biến đầu vào và đầu ra, đồng thời thể hiện mối tương quan giữa các biến đầu vào với nhau.

Dựa vào giá trị  $r_s$  nhận thấy rằng, mối tương quan giữa các biến vào với nhau và giữa các biến đầu vào với biến đầu ra tương đối thấp, ngoại trừ có một vài tương quan ở mức độ vừa phải như tương quan giữa  $I_1$  với Y ( $r_s = 0,5$ ), tương quan giữa  $I_1$  với  $I_6$  ( $r_s = 0,55$ ). Do đó, 8 thông số đầu vào của bộ dữ liệu có thể coi là các biến độc lập. Vì vậy, trong nghiên cứu này, tất cả các biến sẽ được xem xét để tăng mức độ chính xác và tổng quát của mô hình dự báo.



Hình 2.1: Ma trận tương quan giữa các biến đầu vào và biến đầu ra

**3. PHƯƠNG PHÁP MÁY HỌC**

**3.1. Phương pháp cây quyết định (Decision tree)**

Cây quyết định (DT) là một phương pháp máy học phổ biến có thể được áp dụng để giải quyết nhiều vấn đề phức tạp do Quinlan đề xuất. Nguyên tắc chính của DT là sử dụng một loạt các quy tắc để xác định các vùng có các biến đầu

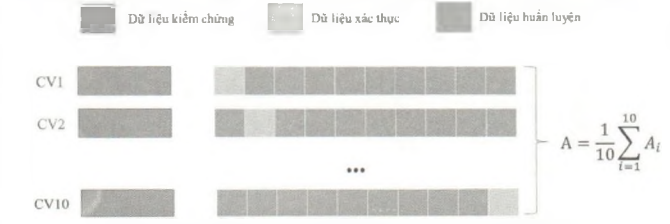
ra đồng nhất với các biến đầu vào mà trên đó xác định một hằng số phù hợp với từng vùng. DT có các ưu điểm như đại diện cho thông tin một cách trực quan và dễ hình dung, mạnh mẽ để khai thác các hiệu ứng phi tuyến và tương tác giữa các biến phụ thuộc và độc lập, không yêu cầu giả định toán học giữa đầu ra và các biến đầu vào, có khả năng xử lý các giá trị bị thiếu và các giá trị ngoại lai. Mặt khác, DT cũng còn một số hạn chế như: khó mô hình hóa các hàm, cấu trúc cây nhạy cảm với dữ liệu mẫu vì chỉ cần thay đổi nhỏ trong dữ liệu huấn luyện cũng có thể cho kết quả rất khác biệt, có phương sai cao và độ lệch thấp.

**3.2. Mô hình máy học kết hợp dựa trên thuật toán cây quyết định dùng kỹ thuật tăng cường**

Để tăng hiệu suất dự báo và khả năng ổn định của mô hình, người ta có thể sử dụng kết hợp các mô hình lại với nhau gọi là mô hình máy học kết hợp (Ensemble Learning) [8]. Đồng thời, để khắc phục các nhược điểm của DT, nhiều kỹ thuật đã được đề xuất để nâng cao khả năng dự báo của mô hình như Ensemble Bagged Trees và Ensemble Boosted Trees. Trong nghiên cứu này, tác giả sử dụng mô hình Ensemble Boosted Trees (EBT). Đây là một mô hình hồi quy cộng tính, tận dụng lợi thế của các phương pháp cây quyết định và khắc phục nhược điểm của các mô hình cây đơn như có thể chọn các biến có liên quan để phù hợp với các hàm chính xác, có thể phù hợp với lượng dữ liệu khác nhau bằng cách sử dụng tăng cường ngẫu nhiên, có thể giảm cả độ lệch bằng cách phù hợp theo từng giai đoạn về phía trước và bằng cách lấy trung bình phương sai của mô hình.

**3.3. Xác thực chéo (Cross Validation)**

Trong lĩnh vực máy học, xác thực chéo là một kỹ thuật phổ biến được sử dụng trong quá trình đào tạo và chỉnh sửa mô hình, nhằm khắc phục hiện tượng quá khớp. Toàn bộ dữ liệu được chia làm hai phần, tập dữ liệu đào tạo và tập dữ liệu kiểm tra. Khi đó, tập dữ liệu đào tạo sẽ được chia ngẫu nhiên thành K phần bằng nhau, mỗi lần đào tạo sẽ chọn 1 phần làm dữ liệu xác thực và (K-1), phần còn lại làm dữ liệu huấn luyện. Quá trình đào tạo mô hình sẽ được thực hiện trong K lần lặp lại. Kết quả đánh giá mô hình cuối cùng sẽ là trung bình cộng kết quả đánh giá của K lần đào tạo. Việc chọn K phải phù hợp vì nếu K quá lớn, tập dữ liệu huấn luyện sẽ lớn hơn nhiều so với tập dữ liệu kiểm chứng, và kết quả đánh giá sẽ không phản ánh đúng bản chất của phương pháp máy học, đặc biệt là với các tập dữ liệu lớn. Trong nghiên cứu này chọn K = 10. Sự lựa chọn này trùng khớp với sự lựa chọn của nhiều nghiên cứu quốc tế [9]. Sơ đồ kỹ thuật xác thực chéo 10 lần được mô tả như Hình 3.1.



Hình 3.1: Minh họa kỹ thuật xác nhận chéo 10 lần

**3.4. Các chỉ số đánh giá hiệu suất dự báo của mô hình học máy**

Hiệu suất dự báo của mô hình được đánh giá thông qua

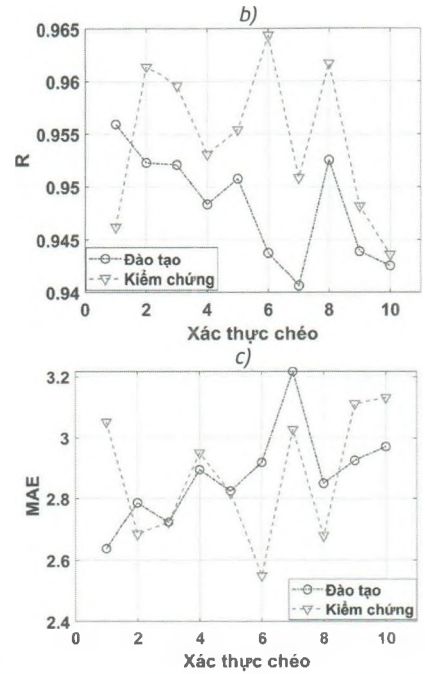
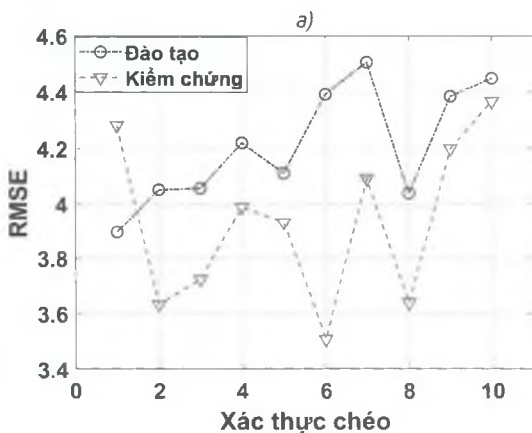
3 chỉ số là hệ số tương quan (R), căn của sai số toàn phương trung bình (RMSE), sai số tuyệt đối trung bình (MAE). Trong đó, R biểu thị mối tương quan giữa giá trị dự đoán và giá trị thực tế, có giá trị trong khoảng [-1, 1]. Mô hình sẽ có độ chính xác cao khi giá trị tuyệt đối của R tiến tới 1. Ngược lại, RMSE và MAE càng thấp cho thấy khả năng dự đoán của mô hình càng chính xác. Giá trị lý tưởng của RMSE và MAE là 0. Chỉ số RMSE tính trung bình bình phương chênh lệch giữa giá trị dự đoán và giá trị thực tế, trong khi MAE là phép đo sai số trung bình giữa chúng. Công thức xác định R, RMSE, MAE tham khảo trong tài liệu [10].

**4. KẾT QUẢ VÀ THẢO LUẬN**

**4.1. Xây dựng mô hình máy học EBT**

Nội dung phần này mô tả quá trình xây dựng mô hình máy học EBT, bản chất của quá trình này là đi xác định các siêu tham số. Toàn bộ dữ liệu được chia ngẫu nhiên thành hai phần: phần dữ liệu đào tạo (chiếm 70% tổng số dữ liệu) và phần dữ liệu kiểm chứng (chiếm 30% tổng số dữ liệu). Phần dữ liệu kiểm chứng được để riêng, không sử dụng trong quá trình xây dựng mô hình, có nghĩa là các siêu tham số của mô hình EBT được lựa chọn chỉ dựa trên phần dữ liệu đào tạo. Tiếp theo, quá trình xác thực chéo 10 lần được thực hiện bằng cách chia ngẫu nhiên phần dữ liệu đào tạo thành 10 phần bằng nhau. Mỗi lần đào tạo sẽ chọn 1 phần làm dữ liệu xác thực và 9 phần còn lại làm dữ liệu huấn luyện. Hiệu suất dự báo của mô hình được xác định là trung bình cộng sau 10 lần mô phỏng và kết quả được trình bày ở Hình 4.1.

Hình 4.1a, 4.2b, 4.3c lần lượt mô tả kết quả hiệu suất dự báo của mô hình EBT dựa trên các tiêu chí là RMSE, R và MAE. Quan sát trên Hình 4.1 cho thấy, mô hình EBT có khả năng đào tạo tương đối tốt thể hiện ở các giá trị R tương đối cao, các giá trị MAE, RMSE thấp ( $R \approx 0,95$ ,  $RMSE \approx 4,2$ ,  $MAE \approx 2,9$ ) và tương đối ổn định thể hiện ở biên độ dao động cho mỗi giá trị R, RMSE, MAE giữa các lần xác thực chéo tương đối thấp (biên độ dao động của  $RMSE \approx 0,6$ ,  $R \approx 0,016$ ,  $MAE \approx 0,57$ ). Tương tự, ở giai đoạn kiểm chứng, các chỉ tiêu R, RMSE, MAE cũng được tính dựa trên tập dữ liệu kiểm chứng. Kết quả cho thấy, ở giai đoạn kiểm chứng, khả năng dự báo của mô hình đề xuất cũng rất chính xác ( $R \approx 0,955$ ,  $RMSE \approx 4,1$ ,  $MAE \approx 3,0$ ) và ổn định (biên độ của giá trị  $RMSE \approx 0,88$ ,  $R \approx 0,021$ ,  $MAE \approx 0,57$ ). Các kết quả trên cho thấy sau 10 lần xác thực chéo, mô hình EBT được xây dựng có khả năng dự báo tương đối chính xác và ổn định.

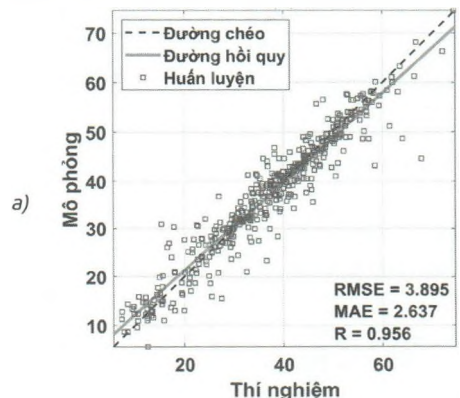


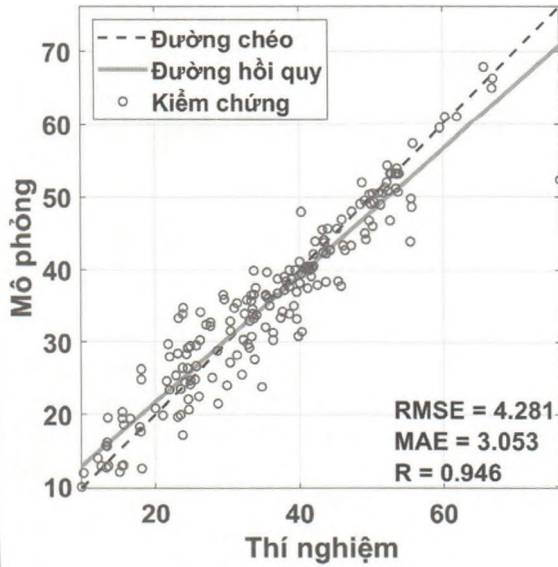
Hình 4.1: Kết quả hiệu suất dự báo của mô hình EBT thể hiện qua các chỉ số thống kê: (a) RMSE; (b) R và (c) MAE

**4.2. Kết quả dự báo tiêu biểu của mô hình EBT**

Trong phần này, kết quả dự báo điển hình của mô hình EBT được trình bày thông qua biểu đồ hồi quy như Hình 4.2. Mô hình hồi quy thể hiện tương quan kết quả giữa giá trị cường độ chịu nén dự đoán (mô phỏng) theo mô hình EBT và giá trị cường độ chịu nén thực tế (thu được từ thí nghiệm) cho tập dữ liệu huấn luyện (Hình 4.2a) và tập dữ liệu kiểm chứng (Hình 4.2b). Trong đó, trục hoành đại diện cho kết quả của thí nghiệm được thu thập, trục tung đại diện cho kết quả được dự đoán theo mô hình đề xuất. Quan sát cho thấy, giá trị thu được từ mô hình đề xuất cho tập dữ liệu huấn luyện và tập dữ liệu kiểm chứng rất gần với kết quả thí nghiệm cho thấy khả năng dự báo chính xác của mô hình.

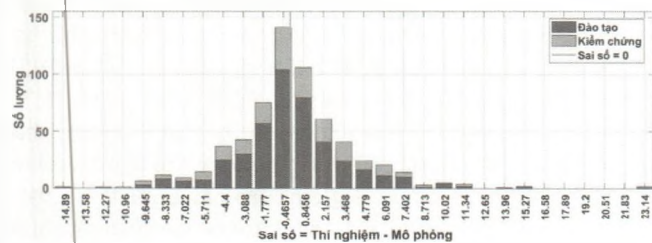
Hơn nữa, hiệu suất của mô hình còn được đánh giá bởi các tiêu chí thống kê RMSE, MAE, R. Trong trường hợp kết quả dự báo tốt nhất đối với tập dữ liệu huấn luyện là  $RMSE = 3,895$  Mpa,  $MAE = 2,637$  Mpa,  $R = 0,956$  và đối với tập dữ liệu kiểm chứng là  $RMSE = 4,281$  Mpa,  $MAE = 3,053$  Mpa,  $R = 0,946$ . Hệ số tương quan R cao kết hợp với sai số thấp chứng tỏ rằng mô hình EBT được đề xuất có khả năng dự báo chính xác và thể hiện hiệu suất tổng quát hóa trong việc dự báo CS của FAC.





Hình 4.2: Biểu đồ hồi quy của mô hình EBT cho (a) tập dữ liệu huấn luyện, (b) tập dữ liệu kiểm chứng

Tiếp theo, biểu đồ phân bố sai số giữa kết quả thí nghiệm và kết quả mô phỏng của mô hình EBT cho cả tập dữ liệu huấn luyện và tập dữ liệu kiểm chứng được trình bày trên Hình 4.3. Quan sát cho thấy, với cả hai tập dữ liệu đào tạo và tập dữ liệu kiểm chứng thì phần lớn việc dự đoán các mẫu đều có sai số nhỏ, xấp xỉ bằng 0 (tức là giá trị thí nghiệm bằng giá trị mô phỏng). Cụ thể, có khoảng 180 trường hợp trong tổng số 435 dữ liệu đào tạo và khoảng 80 trường hợp trong tổng số 187 dữ liệu kiểm chứng có sai số nhỏ xấp xỉ 0 MPa. Chỉ có một vài trường hợp có sai số lớn (lớn hơn 10 MPa). Điều này chứng tỏ khả năng mô phỏng tuyệt vời của mô hình được đề xuất khi dự báo CS của FAC.



Hình 4.3: Biểu đồ phân bố sai số của mô hình EBT cho tập dữ liệu đào tạo và tập dữ liệu kiểm chứng

## 5. KẾT LUẬN

Trong nghiên cứu này, để dự đoán cường độ chịu nén của bê tông tro bay, một mô hình kết hợp dựa trên thuật toán cây quyết định sử dụng kỹ thuật tăng cường (EBT) đã được phát triển. Mô hình kết hợp EBT tận dụng được lợi thế của phương pháp cây quyết định và khắc phục nhược điểm của các mô hình cây đơn để nâng cao hiệu suất và khả năng dự báo. Kết quả cho thấy EBT là một mô hình rất tốt trong dự báo CS của FAC với hiệu suất dự đoán cao và ổn định thể hiện ở các giá trị  $R = 0,946$ ,  $RMSE = 4,281$  Mpa,  $MAE = 3,053$  MPa cho dữ liệu kiểm chứng. Kết quả nghiên cứu giúp giảm bớt được các thí nghiệm, từ đó tiết kiệm được chi phí và thời gian trong việc xác định cường độ chịu nén của FAC.

## Tài liệu tham khảo

- [1]. A. Oner, S. Akyuz and R. Yildiz (Jun. 2005), *An experimental study on strength development of concrete containing fly ash and optimum usage of fly ash in concrete*, Cement and Concrete Research, vol.35, no.6, pp.1165-1171, doi: 10.1016/j.cemconres.2004.09.031.
- [2]. H. N. Muliauwan, D. Prayogo, G. Gaby and K. Harsono (Sep. 2020), *Prediction of Concrete Compressive Strength Using Artificial Intelligence Methods*, J. Phys.: Conf. Ser., vol.1625, no.1, p.012018, doi: 10.1088/1742-6596/1625/1/012018.
- [3]. İ. B. Topçu and M. Saridemir (Jan. 2008), *Prediction of compressive strength of concrete containing fly ash using artificial neural networks and fuzzy logic*, Computational Materials Science, vol.41, no.3, pp.305-311, doi: 10.1016/j.commatsci.2007.04.009.
- [4]. I.-C. Yeh (Dec. 1998), *Modeling of strength of high-performance concrete using artificial neural networks*, Cement and Concrete Research, vol.28, no.12, pp.1797-1808, doi: 10.1016/S0008-8846(98)00165-3.
- [5]. Dr. J. Pitroda (Oct. 2014), *PREDICTION OF STRENGTH FOR FLY ASH CEMENT CONCRETE THROUGH SOFT COMPUTING APPROACHES*, International Journal of Advanced Research in Engineering, Science & Management (IJARESM), ISSN: 2394-1766, vol.1, pp.1-11.
- [6]. P. Chopra, R. K. Sharma and M. Kumar (Jan. 2016), *Prediction of Compressive Strength of Concrete Using Artificial Neural Network and Genetic Programming*, Advances in Materials Science and Engineering, vol.2016, p.e7648467, doi: 10.1155/2016/7648467.
- [7]. H. Nguyen, T. Vu, T. P. Vo and H.-T. Thai (Jan. 2021), *Efficient machine learning models for prediction of concrete strengths*, Construction and Building Materials, vol.266, p.120950, doi: 10.1016/j.conbuildmat.2020.120950.
- [8]. Z.-H. Zhou (2009), *Ensemble Learning*, in Encyclopedia of Biometrics, S. Z. Li and A. Jain, Eds. Boston, MA: Springer US, pp.270-273. doi: 10.1007/978-0-387-73003-5\_293.
- [9]. R. Kohavi (Aug. 1995), *A study of cross-validation and bootstrap for accuracy estimation and model selection*, in Proceedings of the 14th international joint conference on Artificial intelligence, vol.2, San Francisco, CA, USA, pp.1137-1143.
- [10]. H.-V. T. Mai, T.-A. Nguyen, H.-B. Ly and V. Q. Tran (May 2021), *Prediction Compressive Strength of Concrete Containing GGBFS using Random Forest Model*, Advances in Civil Engineering, vol.2021, p.e6671448, doi: 10.1155/2021/6671448.

Ngày nhận bài: 01/4/2022

Ngày chấp nhận đăng: 17/5/2022

Người phản biện: TS. Hồ Sĩ Lành

TS. Lý Hải Bằng