

# Ứng dụng mô hình học máy để dự đoán cường độ chịu nén hỗn hợp chất thải mỏ quặng gia cố bằng xi măng trong trường hợp khan hiếm dữ liệu

Applying machine learning model for predicting unconfined compressive strength of cemented paste backfill on scarce data

> **TRẦN VĂN QUÂN\***, **NGÔ VIỆT CƯỜNG**

Trường Đại học Công nghệ Giao thông vận tải, \* Email: quantv@utt.edu.vn

## TÓM TẮT

Cường độ chịu nén của chất thải mỏ quặng được gia cố bằng xi măng (Cemented Paste Backfill CPB) là tính chất cơ học quan trọng trong việc đánh giá khả năng áp dụng của hỗn hợp này trong gia cố hố đào hầm mỏ. Bài báo trình bày việc sử dụng hai mô hình học máy đơn giản để dự đoán cường độ chịu nén của chất thải mỏ quặng. Do đó, hai thuật toán học máy bao gồm thuật toán tăng cường độ dốc (Gradient Boosting GB) và thuật toán vector hỗ trợ (Support Vector Regression SVR) được sử dụng để dự đoán cường độ chịu nén chất thải mỏ quặng được gia cố bằng xi măng. Để thực hiện việc xây dựng mô hình học máy, 92 dữ liệu thí nghiệm đã được thu thập từ những công bố quốc tế. Bộ dữ liệu bao gồm sáu biến đầu vào Xi măng/Chất thải mỏ C/T; Hàm lượng chất rắn (%); trọng lượng riêng G<sub>s</sub>; Kích thước lỗ sàng đạt lượng tích lũy 10% D<sub>10</sub> (μm); hệ số đồng đều C<sub>u</sub>; hệ số cấp phối C<sub>c</sub>. Hiệu suất vượt trội của mô hình học máy GB so với mô hình học máy SVR được kiểm chứng bằng 200 lần mô phỏng ngẫu nhiên Monte Carlo. Phân tích yếu tố quan trọng cho thấy sự cần thiết của các yếu tố đầu vào đến việc tăng cường hiệu suất của mô hình GB có thể sắp xếp theo thứ tự giảm dần như sau tỷ lệ xi măng/chất thải mỏ C/T > hệ số đồng đều C<sub>u</sub> > hàm lượng chất rắn (Solid content) > kích thước lỗ sàng đạt lượng tích lũy 10% D<sub>10</sub> > hệ số cấp phối C<sub>c</sub> > trọng lượng riêng G<sub>s</sub>.

**Từ khóa:** Mô hình học máy; cường độ chịu nén; tỷ lệ xi măng/chất thải mỏ; chất thải mỏ quặng gia cố bằng xi măng.

## ABSTRACT

The compressive strength of Cemented Paste Backfill CPB (Cemented Paste Backfill CPB) is an important mechanical property in evaluating the applicability of this mixture in the reinforcement of mine pits. This paper presents the use of two simple machine learning models to predict the compressive strength of CPB. Therefore, two machine learning algorithms including Gradient Boosting GB and Support Vector Regression SVR are used to predict the compressive strength of CPB. For building the machine learning model, 92 experimental data were collected from international publications. The dataset includes six input variables such as cement/tailing content C/T, solids content (%), specific gravity G<sub>s</sub>, sieve hole size with cumulative 10% D<sub>10</sub> (μm), coefficient of uniformity C<sub>u</sub>, coefficient of curvature C<sub>c</sub>. The superior performance of GB machine learning model over SVR machine learning model is verified by 200 times of Monte Carlo random simulation. Feature importance analysis shows the necessity of the inputs to enhance the performance of the GB model can be arranged in descending order as follows: C/T ratio of cement/tailing content > coefficient of uniformity C<sub>u</sub> > solid content > sieve hole size with cumulative 10% D<sub>10</sub> > coefficient of curvature C<sub>c</sub> > specific gravity G<sub>s</sub>.

**Keywords:** Machine Learning (ML); unconfined compressive strength; cement/tailing; cemented Paste Backfill (CPB).

## 1. ĐẶT VẤN ĐỀ

Chất thải mỏ quặng là một sản phẩm phụ không thể tránh khỏi trong quá trình khai thác tài nguyên khoáng sản. Một vấn đề quan trọng đối với ngành công nghiệp khoáng sản là xử lý an toàn các chất thải từ mỏ để tránh các vấn đề về môi trường. Do đó, Việt Nam cũng cần thiết để xử lý các chất thải mỏ quặng trên khắp cả

nước như Quảng Ninh, Lào Cai, Dak Nông... Công nghệ xử lý hiện nay ở các nước phát triển trên thế giới là sử dụng chất thải mỏ quặng như một thành phần chính của vật liệu san lấp hồ xi măng (Cement Paste Backfill CPB). Thông thường, CPB bao gồm quặng đuôi khử nước (70-85% chất rắn theo trọng lượng), xi măng (3-7% trọng lượng của CPB) và nước. Ngoài việc xử lý an toàn chất thải

của mỏ, CPB còn có những lợi ích khác, bao gồm giảm sụt lún bề mặt và chi phí phục hồi [1]. Hơn nữa, CPB có thể cung cấp hỗ trợ lớp đất mặt thứ cấp cho các hoạt động khai thác để cải thiện môi trường làm việc dưới lòng đất. Tất cả những lợi ích kỹ thuật, kinh tế và môi trường này đã dẫn đến việc áp dụng rộng rãi CPB trên toàn thế giới [2]–[5].

Ưu điểm của CPB phụ thuộc vào tính ổn định cơ học của nó, hiệu suất kinh tế và độ bền. Sau khi đặt, CPB phải có các đặc tính cơ học nhất định để duy trì ổn định trong quá trình đào các đoạn liên kế. Do đó, các đặc tính cơ học của CPB có ý nghĩa rất lớn trong quá trình ứng dụng kỹ thuật của nó. Đặc tính cơ học được sử dụng rộng rãi nhất của CPB là cường độ chịu nén (UCS) vì các thử nghiệm của UCS tương đối đơn giản và kinh tế [6]. Nhiều thí nghiệm UCS đã được thực hiện và đã đạt được nhiều kết quả quan trọng trong việc tìm hiểu mối quan hệ giữa UCS và các biến ảnh hưởng của nó [7], [8]. Việc thực nghiệm xác định các đặc tính cơ học của CPB rất phức tạp và tốn thời gian vì nó bao gồm rất nhiều thử nghiệm cơ học cho từng loại chất thải và các phương pháp thử nghiệm khác nhau, chẳng hạn như UCS. Do đó, một phương pháp có thể dự đoán chính xác các đặc tính cơ học của CPB là hết sức cần thiết cho các kỹ sư thiết kế thành phần chất thải mỏ quặng gia cố xi măng.

Các tính chất cơ học của CPB thường được dự đoán bằng cách sử dụng phương pháp hồi quy tuyến tính hoặc phi tuyến tính thông thường trên dữ liệu thực nghiệm [9], [10]. Tuy nhiên, rất khó để có được một phương trình hồi quy chính xác bằng cách sử dụng các mô hình dựa trên thực nghiệm này. Hơn nữa, các phương trình này thường đặc trưng cho từng mô đối với một lượng quặng nhất định, có nghĩa là khả năng tổng quát hóa của chúng không đạt yêu cầu [11]. Các kỹ thuật tiên tiến hơn là cần thiết để thay thế phân tích hồi quy thông thường trong dự đoán cường độ chịu nén. Hiện nay có nhiều nghiên cứu triển khai ứng dụng các mô hình học máy là các kỹ thuật nền tảng của các mô hình trí tuệ nhân tạo vào nghiên cứu các vấn đề trong lĩnh vực xây dựng công trình. Trần et al. [12] thiết lập một số mô hình học máy để nghiên cứu về cường độ chịu nén của đất gia cố bằng xi măng. Trần [13] sử dụng 8 mô hình học máy để nghiên cứu về hệ số khuếch tán của bê tông xi măng. Cường độ chịu nén của bê tông geopolymer cũng được đánh giá bằng các mô hình học máy trong nghiên cứu của Trần [14]. Do vậy có thể thấy việc ứng dụng mô hình học máy vào đánh giá cường độ chịu nén của chất thải mỏ quặng gia cố bằng xi măng là điều khả thi. Thực tế, một vài nghiên cứu trên thế giới trong việc ứng dụng mô hình học máy vào việc dự đoán cường độ chịu nén của chất thải mỏ quặng gia cố bằng xi măng đã được thực hiện như nghiên cứu của Qi et al. [15], Qi et al. [16], Orejarena and Fall [17]. Tuy nhiên, các nghiên cứu trên có số mẫu để xây dựng mô hình học máy tương đối lớn, tối thiểu là 150 mẫu, do đó nghiên cứu này tập trung vào việc thiết lập mô hình học máy trên số lượng mẫu nhỏ nhưng vẫn có thể dự đoán tương đối chính xác cường độ chịu nén của chất thải mỏ quặng gia cố bằng xi măng. Để đạt được điều này, hai thuật toán học máy phổ biến và đơn giản với các tham số sẵn có trong thư viện Sklearn của ngôn ngữ lập trình Python bao gồm thuật toán tăng cường độ dốc (Gradient Boosting-GB) và thuật toán vector hỗ trợ (Support Vector Regression SVR) đã được lựa chọn để phát triển mô hình học máy dự đoán cường độ chịu nén dựa trên 92 mẫu thí nghiệm và 6 biến đầu vào. Hiệu suất các mô hình học máy sẽ được so sánh bằng các tiêu chí như hệ số xác định R<sup>2</sup>, sai số căn quân phương trung bình RMSE và sai số trung bình tuyệt đối MAE. Độ tin cậy của các mô hình học máy được xác thực bằng mô phỏng ngẫu nhiên Monte Carlo. Từ đó, mô hình học máy trong việc xác định cường độ chịu

nén có hiệu suất tốt hơn sẽ được xác định. Phân tích Shapley Additive Explanation và Feature importance analysis cho phép xác định các yếu tố đầu vào quan trọng trong việc dự đoán chính xác kết quả của mô hình học máy.

## 2. THIẾT LẬP MÔ HÌNH DỰ BÁO

### 2.1. Thuật toán tăng cường độ dốc (Gradient Boosting-GB)

Gradient Boosting là một trong số các thuật toán rất mạnh trong lĩnh vực máy học, được sử dụng rộng rãi cho các vấn đề hồi quy và phân loại [18]. Nó tạo ra một mô hình dự đoán dưới dạng mô hình kết hợp các mô hình dự đoán yếu, điển hình là các cây quyết định (Decision Tree). Nó xây dựng mô hình theo kiểu từng bước giống như các phương pháp boosting khác, và khái quát hóa chúng bằng cách cho phép tối ưu hóa một hàm mất mát (loss function). Mô hình Gradient Boosting có thể được xây dựng gồm 5 bước:

**Bước 1:** Xây dựng mô hình đầu tiên (Mô hình<sub>1</sub>) với hàm mục tiêu mong muốn và tiến hành dự đoán cho bộ dữ liệu.

**Bước 2:** Tính toán sai số dự đoán của mô hình trước (Mô hình<sub>1</sub>) và sử dụng sai số này làm hàm mục tiêu dự đoán.

**Bước 3:** Xây dựng mô hình dự đoán về các sai số này và tiến hành dự đoán.

**Bước 4:** Cập nhật lại kết quả dự đoán của mô hình đầu tiên (Mô hình<sub>1</sub>) dựa trên kết quả dự đoán ở bước 3, ta được mô hình thứ hai (Mô hình<sub>2</sub>).

Mô hình <sub>2</sub> (Original target)	=	Mô hình <sub>1</sub> (Original target)	+	Sai số dự đoán
---	---	---	---	-------------------

**Bước 5:** Lặp lại các bước từ 2 đến 4: tính toán các sai số dự đoán mới và đặt chúng làm hàm mục tiêu dự đoán. Xây dựng mô hình dự đoán về các sai số mới này và tiến hành dự đoán. Kết quả dự đoán lỗi nhận được sẽ được cập nhật vào mô hình thứ hai (Mô hình<sub>2</sub>) ta thu được mô hình ba (Mô hình<sub>3</sub>). Quá trình này (từ bước 2 đến bước 4) sẽ tiếp tục lặp lại nhiều lần cho đến khi sai số dự đoán tiến về 0 hoặc đạt đến điều kiện dừng là số lượng mô hình muốn xây dựng.

### 2.2. Thuật toán vector hỗ trợ (Support Vector Regression-SVR)

Thuật toán vector hỗ trợ (SVR) cho phép xác định mức độ lỗi có thể chấp nhận được trong mô hình và sẽ tìm một đường thích hợp (hoặc siêu phẳng ở các chiều cao hơn) để khớp với dữ liệu [19]. SVR cho phép linh hoạt xác định mức độ lỗi có thể chấp nhận được trong mô hình và sẽ tìm một đường thích hợp (hoặc siêu phẳng ở các chiều cao hơn) để khớp với dữ liệu. Thuật toán vectơ hỗ trợ (SVR) là một loại thuật toán học máy được sử dụng để phân tích hồi quy. Mục tiêu của SVR là tìm một hàm xấp xỉ mối quan hệ giữa các biến đầu vào và biến mục tiêu liên tục, đồng thời giảm thiểu lỗi dự đoán. SVR tìm kiếm một siêu phẳng phù hợp nhất với các điểm dữ liệu trong một không gian liên tục. Điều này đạt được bằng cách ánh xạ các biến đầu vào sang không gian đặc trưng nhiều chiều và tìm siêu phẳng giúp tối đa hóa lề (khoảng cách) giữa siêu phẳng và các điểm dữ liệu gần nhất, đồng thời giảm thiểu lỗi dự đoán. SVR có thể xử lý các mối quan hệ phi tuyến tính giữa các biến đầu vào và biến đích bằng cách sử dụng hàm nhân để ánh xạ dữ liệu sang không gian có chiều cao hơn. Đồng thời, làm cho nó trở thành một công cụ mạnh mẽ cho các nhiệm vụ hồi quy, nơi có thể có các mối quan hệ phức tạp giữa các biến đầu vào và biến mục tiêu. SVR là một kỹ thuật học máy trong đó một mô hình tìm hiểu tầm quan trọng của một biến để mô tả mối quan hệ giữa đầu vào và đầu ra của mô hình.

### 2.3. Đánh giá khả năng dự báo của mô hình học máy

Để đánh giá các giá trị dự đoán của các phương pháp hồi quy,

các giá trị thực tế và dự đoán được so sánh. Trong nghiên cứu này, các tiêu chí đánh giá hiệu suất của mô hình học máy bao gồm hệ số xác định  $R^2$ , sai số căn quân phương trung bình RMSE và sai số trung bình tuyệt đối MAE được sử dụng để đánh giá độ chính xác của dự đoán. Giá trị của  $R^2$  nằm trong khoảng 0 đến 1, giá trị càng gần 1 mô hình có hiệu suất dự đoán càng tốt, ngược lại RMSE và MAE có miền giá trị từ 0 đến vô cùng, giá trị của 2 chỉ số này càng gần 0, các sai số dự đoán càng nhỏ như vậy mô hình có độ chính xác càng lớn.

$$R^2 = \frac{\sum_{k=1}^N (val_k^{ex} - val_{avg}^{ex})^2 - \sum_{k=1}^N (val_k^{ex} - val_k^{pre})^2}{\sum_{k=1}^N (val_k^{ex} - val_{avg}^{ex})^2} \quad (1)$$

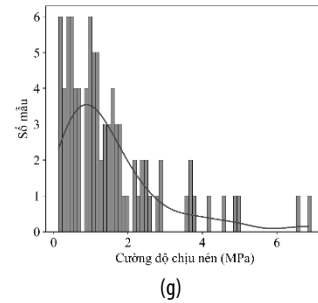
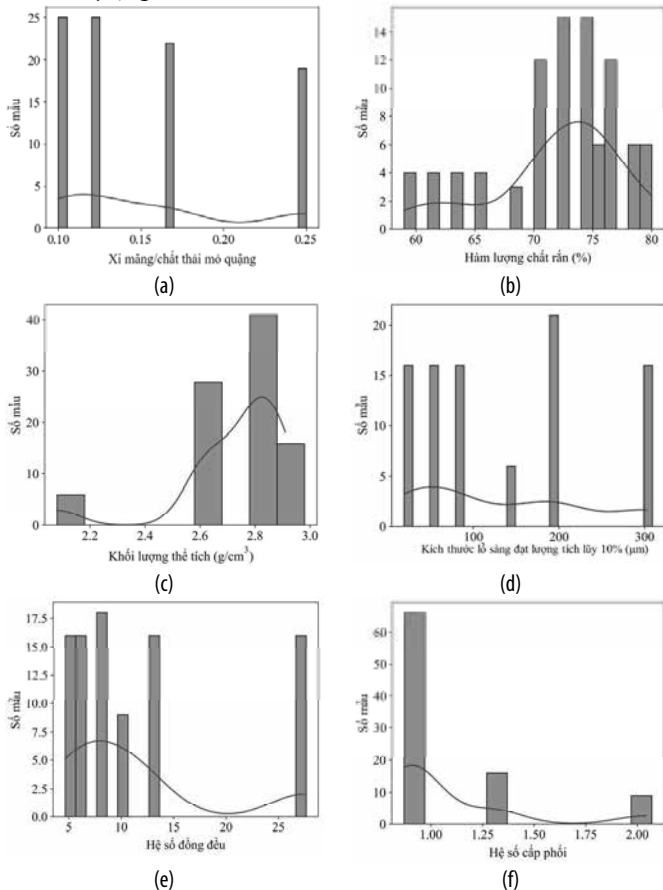
$$RMSE = \sqrt{\frac{1}{N} \sum_{k=1}^N (val_k^{ex} - val_k^{pre})^2} \quad (2)$$

$$MAE = \frac{1}{N} \sum_{k=1}^N |val_k^{pre} - val_k^{ex}| \quad (3)$$

Trong đó: N là số lượng dữ liệu của tập sử dụng,  $val_k^{ex}$  giá trị thực tế thứ k,  $val_k^{pre}$  giá trị dự đoán thứ k tương ứng,  $val_{avg}^{ex}$  giá trị trung bình của tập dữ liệu thực tế.

### 3. MÔ TẢ CƠ SỞ DỮ LIỆU

Cơ sở dữ liệu gồm 92 mẫu được thu thập từ nghiên cứu của Qi et al. [15]. Tập dữ liệu này có 6 biến đầu vào bao gồm tỷ lệ xi măng/chất thải mỏ quặng C/T, hàm lượng chất rắn (Solid content) (%), trọng lượng riêng Gs, kích thước lỗ sàng đạt tích lũy 10%  $D_{10}$  ( $\mu\text{m}$ ), hệ số đồng đều  $C_u$ , hệ số cấp phối  $C_c$ . Tham số đầu ra của của việc xây dựng mô hình học máy là: cường độ chịu nén xi măng chất thải mỏ quặng.



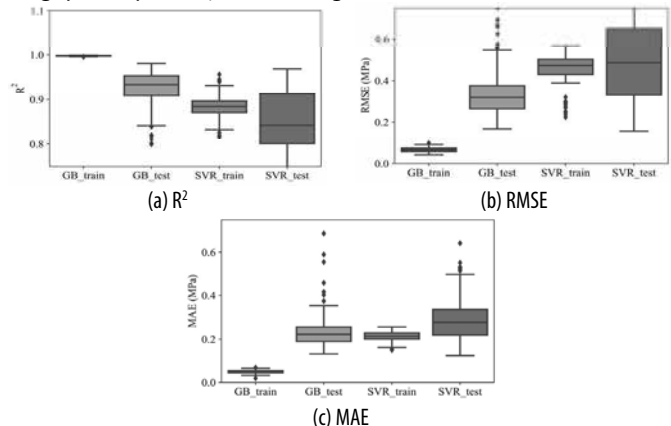
**Hình 1.** Cường độ chịu nén cho phần huấn luyện mô hình học máy GB và SVR với 92 mẫu, cường độ chịu nén 28 ngày

Hình 1a cho thấy tỷ lệ xi măng/chất thải mỏ quặng đạt từ 0.11 đến 0.24. Hình 1b mô tả hàm lượng chất rắn sử dụng trong cấp phối gia cố chất thải mỏ quặng thay đổi từ 58 đến 79 (%), Hình 1c mô tả trọng lượng riêng từ 2.1 đến 2.9 ( $\text{g}/\text{cm}^3$ ), Hình 1d cho thấy Kích thước lỗ sàng đạt lượng tích lũy 10% thay đổi trong khoảng từ 10  $\mu\text{m}$  đến 300  $\mu\text{m}$ , Hình 1e mô tả hệ số đồng đều  $C_u$  từ 4 đến 2, Hình 1f mô tả hệ số cấp phối  $C_c$  từ 0.9 đến 2.1, Hình 1g cho thấy miền giá trị cường độ chịu nén của 92 mẫu thay đổi tương đối lớn từ 1 đến 6 (MPa).

### 4. KẾT QUẢ

#### 4.1. Đánh giá hiệu suất của hai mô hình học máy GB và SVR bằng mô phỏng Monte Carlo

Bộ dữ liệu được chia 70%/30% tương ứng là 64 mẫu/28 mẫu cho tập dữ liệu huấn luyện và kiểm chứng. Trong đó, do có tính ngẫu nhiên trong việc phân chia tập dữ liệu do đó, kỹ thuật mô phỏng ngẫu nhiên Monte Carlo được sử dụng để đánh giá hiệu suất mô hình học máy SVR và GB một cách toàn diện và chặt chẽ. Sự hội tụ của giá trị  $R^2$  cho tập dữ liệu kiểm chứng sau 200 lần lặp ngẫu nhiên Monte Carlo cho thấy giá trị  $R^2$  cho tập dữ liệu kiểm chứng trở nên ổn định sau khoảng 200 lần chạy lặp cho cả hai mô hình học máy SVR và GB trong việc dự đoán cường độ chịu nén. Do đó, việc lựa chọn số lần chạy lặp ngẫu nhiên là 200 đảm bảo tính xác thực chặt chẽ của các giá trị hiệu suất của mô hình học máy được đưa ra. Kết quả so sánh cho thấy mô hình GB cho hiệu suất cao hơn mô hình SVR trong việc dự đoán cường độ chịu nén xi măng chất thải mỏ quặng. Hiệu suất dự đoán cao hơn của mô hình GB được thể hiện qua giá trị trung bình sau 200 lần mô phỏng ngẫu nhiên Monte Carlo của các tiêu chí  $R^2$ , RMSE và MAE. Các giá trị trung bình tương ứng đối với mô hình GB lần lượt là  $R^2=0.9245$ ,  $RMSE=0.3430$  MPa và  $MAE=0.2286$  MPa cho tập dữ liệu kiểm chứng trong khi mô hình SVR có giá trị trung bình của  $R^2=0.8454$ ,  $RMSE=0.5132$  MPa và  $MAE=0.2873$  MPa. Kết quả này được thể hiện rõ ràng qua boxplot được mô tả trong Hình 2.

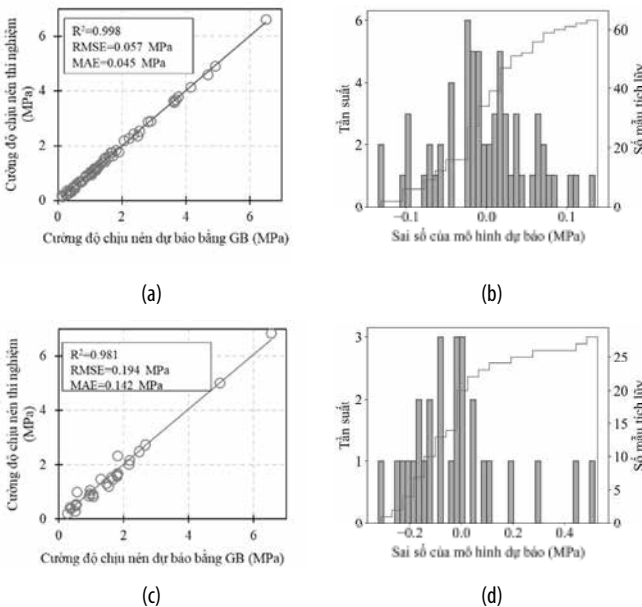


**Hình 2.** So sánh hiệu suất của hai mô hình SVR và GB sau 200 lần mô phỏng Monte Carlo

Do vậy, có thể thấy mô hình GB có hiệu suất tốt hơn mô hình SVR trong việc xác định cường độ chịu nén chất thải mỏ quặng gia cố bằng xi măng.

**4.2. Dự báo cường độ chịu nén chất thải mỏ quặng gia cố xi măng bằng mô hình GB**

Trong mục trước, mô hình học máy GB đã được chứng minh có hiệu suất tốt hơn mô hình học máy SVR trong việc dự đoán cường độ chịu nén của hỗn hợp chất thải mỏ quặng gia cố bằng xi măng. Khả năng dự đoán cường độ chịu nén của hỗn hợp này bằng mô hình học máy GB được mô tả rõ hơn thông qua việc so sánh cường độ chịu nén dự đoán với cường độ chịu nén thực tế cho đồng thời tập dữ liệu huấn luyện mô hình (Hình 3a và Hình 3b) và tập dữ liệu kiểm chứng (Hình 3c và Hình 3d). Kết quả so sánh cho thấy mô hình GB có khả năng dự báo rất tốt cường độ chịu nén trên tập dữ liệu thí nghiệm, khi mà kết biểu đồ tương quan giữa cường độ thí nghiệm và dự đoán hầu như nằm trọn vẹn trên đường tuyến tính  $y=x$  (Hình 3a), với sai số dự đoán rất nhỏ từ -0.15 MPa đến 0.15 MPa so với miền giá trị cường độ chịu nén thực tế từ 1 MPa đến 6 MPa, nhiều sai số dự đoán nằm quanh vùng  $\pm 0.01$  MPa (Hình 3b). Các chỉ số dự đoán của tập huấn luyện đều rất tốt với  $R^2= 0.998$ ,  $RMSE=0.057$  MPa và  $MAE=0.045$  MPa.

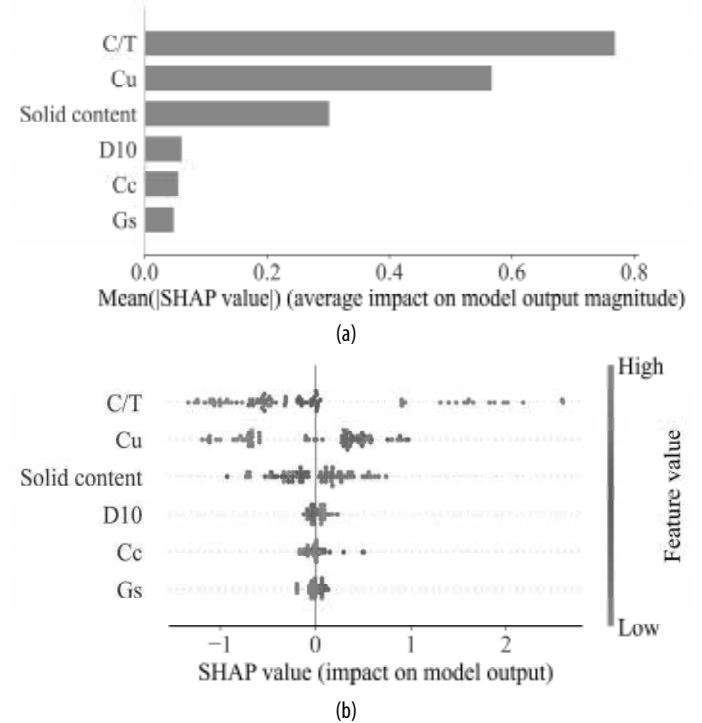


**Hình 3.** So sánh kết quả cường độ chịu nén thí nghiệm và cường độ chịu nén dự đoán bằng mô hình GB

Đối với tập dữ liệu kiểm chứng, tuy khả năng dự đoán cường độ chịu nén kém hơn tập dữ liệu huấn luyện, khi mà Hình 3c cho thấy có một vài điểm nằm ngoài đường thẳng  $y=x$  trên biểu đồ so sánh tương quan giữa cường độ chịu nén thực tế và cường độ chịu nén dự đoán bằng mô hình GB, cũng như miền sai số thay đổi từ -0.3 MPa đến 0.5 MPa, tuy nhiên có thể thấy số lượng sai số chủ yếu vẫn tập trung trong khoảng  $\pm 0.1$  MPa (Hình 3d).

Có thể thấy rằng với số lượng tương đối ít là 92 mẫu, mô hình học máy GB vẫn có thể dự đoán cường độ chịu nén của hỗn hợp chất thải mỏ quặng với độ chính xác tương đối cao với các chỉ số dự đoán  $R^2=0.981$ ,  $RMSE=0.194$  MPa, và  $MAE=0.142$  MPa cho tập dữ liệu kiểm chứng. Độ chính xác của việc dự đoán phụ thuộc rất lớn vào biến đầu vào xây dựng mô hình sẽ được phân tích kỹ hơn trong mục tiếp theo của bài báo.

**4.3. Ảnh hưởng của các giá trị đầu vào đến cường độ chịu nén của hỗn hợp chất thải mỏ quặng gia cố bằng xi măng**



**Hình 4.** Phân tích các yếu tố đầu vào ảnh hưởng đến khả năng dự báo chính xác của mô hình học máy GB

Hiệu suất của mô hình học máy phụ thuộc rất lớn vào các biến đầu vào sử dụng cho việc xây dựng mô hình. Trong nghiên cứu này, mô hình học máy GB có thể dự đoán cường độ chịu nén của hỗn hợp chất thải mỏ quặng gia cố bằng xi măng là dựa vào 6 biến đầu vào bao gồm tỷ lệ xi măng/chất thải mỏ C/T, hàm lượng chất rắn (Solid content) (%), trọng lượng riêng G<sub>s</sub>, kích thước hạt tích lũy đạt 10% D<sub>10</sub> (μm), hệ số đồng đều C<sub>u</sub>, hệ số cấp phối C<sub>c</sub>. Tuy nhiên ảnh hưởng của các biến này đến hiệu suất dự đoán của mô hình GB là không giống nhau. Phân tích yếu tố quan trọng (Hình 4a) cho thấy sự cần thiết của các yếu tố đầu vào đến việc tăng cường hiệu suất của mô hình GB có thể sắp xếp theo thứ tự giảm dần như sau tỷ lệ xi măng/chất thải mỏ C/T > hệ số đồng đều C<sub>u</sub> > hàm lượng chất rắn (Solid content) > Kích thước lỗ sàng đạt lượng tích lũy 10% D<sub>10</sub> > hệ số cấp phối C<sub>c</sub> > trọng lượng riêng G<sub>s</sub>. Do vậy, để có thể xây dựng mô hình học máy GB có hiệu suất cao trong việc dự đoán cường độ chịu nén của hỗn hợp chất thải mỏ quặng gia cố bằng xi măng, ba biến đầu vào cần thiết là tỷ lệ xi măng/chất thải mỏ C/T, hệ số đồng đều C<sub>u</sub>, hàm lượng chất rắn (Solid content).

Ngoài ra, Shapley Additive Explanations (SHAP) được mô tả trong Hình 4b còn cho phân tích định lượng ảnh hưởng của 6 biến đầu vào này đến giá trị cường độ chịu nén của hỗn hợp chất thải mỏ quặng gia cố bằng xi măng. Trong đó, tỷ lệ C/T càng lớn, cường độ chịu nén càng tăng, với miền giá trị ảnh hưởng vào khoảng 4.0 MPa. Điều này đúng khi tỷ lệ C/T tăng tương ứng hàm lượng xi măng sử dụng tăng. Hệ số đồng đều C<sub>u</sub> tăng, cường độ chịu nén càng lớn, với miền ảnh hưởng vào khoảng 2.2 MPa. Hàm lượng chất rắn càng lớn, cường độ chịu nén càng tăng với khoảng tăng là 2.0 MPa. Kích thước lỗ sàng đạt lượng tích lũy 10% càng lớn thì cường độ chịu nén của mẫu càng giảm, tuy nhiên sự suy giảm không lớn. Hệ số đồng đều và trọng lượng riêng ảnh hưởng không rõ ràng tới cường độ chịu nén.

## 5. KẾT LUẬN VÀ KIẾN NGHỊ

Dựa trên bộ dữ liệu nhỏ gồm 92 mẫu và 6 biến đầu vào, nghiên cứu này đã xây dựng thành công hai mô hình học máy thông dụng là mô hình tăng cường độ dốc GB và mô hình vector hỗ trợ để dự đoán cường độ chịu nén của hỗn hợp chất thải mỏ quặng gia cố bằng xi măng. Sử dụng mô phỏng ngẫu nhiên Monte Carlo, mô hình học máy GB được khẳng định có hiệu suất cao hơn mô hình SVR trong việc dự đoán cường độ chịu nén của hỗn hợp mỏ quặng gia cố bằng xi măng với các tiêu chí đánh giá  $R^2$ , RMSE và MAE. Trong đó, độ chính xác của các giá trị dự đoán bằng mô hình tăng cường độ dốc GB cho tập dữ liệu kiểm chứng đạt kết quả cao với các chỉ số đánh giá hiệu  $R^2=0.981$ , RMSE=0.194 MPa, và MAE=0.142 MPa cho tập dữ liệu kiểm chứng.

Phân tích yếu tố quan trọng cho thấy sự cần thiết của các yếu tố đầu vào đến việc tăng cường hiệu suất của mô hình GB có thể sắp xếp theo thứ tự giảm dần như sau tỷ lệ xi măng/chất thải mỏ C/T > hệ số đồng đều  $C_u$  > hàm lượng chất rắn (Solid content) > Kích thước lỗ sàng đạt lượng tích lũy 10%  $D_{10}$  > hệ số cấp phối  $C_c$  > trọng lượng riêng  $G_s$ . Do vậy, để có thể xây dựng mô hình học máy GB có hiệu suất cao trong việc dự đoán cường độ chịu nén của hỗn hợp chất thải mỏ quặng gia cố bằng xi măng, ba biến đầu vào cần thiết là tỷ lệ xi măng/chất thải mỏ C/T, hệ số đồng đều  $C_u$ , hàm lượng chất rắn. Phân tích giá trị SHAP cho thấy cường độ chịu nén tỷ lệ thuận với các giá trị của tỷ lệ xi măng/chất thải mỏ, hệ số đồng đều và hàm lượng chất rắn. Đồng thời, giá trị cường độ chịu nén tỷ lệ nghịch với kích thước lỗ sàng đạt lượng tích lũy 10%. Trọng lượng riêng và hệ số cấp phối ít có ảnh hưởng đến giá trị cường độ chịu nén.

Việc xây dựng thành công mô hình học máy GB trong việc dự đoán cường độ chịu nén của hỗn hợp chất thải mỏ quặng được gia cố bằng xi măng có thể giúp tiết kiệm thời gian và chi phí trong quá trình thiết kế thành phần hỗn hợp chất thải mỏ quặng được gia cố bằng xi măng. Nghiên cứu cần được mở rộng với số lượng mẫu lớn hơn và ít biến đầu vào hơn để nâng cao khả năng ứng dụng mô hình học máy trong thực tế khi có ít dữ liệu đầu vào, cũng như nâng cao độ chính xác của việc thiết kế.

### Lời cảm ơn

Nghiên cứu này được tài trợ bởi Trường Đại học Công nghệ Giao thông vận tải (ĐHCNGTVT) trong đề tài mã số ĐTTĐ2022-17

### TÀI LIỆU THAM KHẢO

[1] E. Yilmaz, 'Investigating the Hydrogeotechnical and Microstructural Properties of Cemented Paste Backfill Using the CUAPS Apparatus', phd, Université du Québec à en Abitibi-Témiscamingue, [Rouyn-Noranda], 2010. Accessed: Feb. 24, 2022. [Online]. Available: <https://depositum.uqat.ca/id/eprint/34/>

[2] Q. Chen, Q. Zhang, A. Fourie, X. Chen, and C. Qi, 'Experimental investigation on the strength characteristics of cement paste backfill in a similar slope model and its mechanism', *Construction and Building Materials*, vol. 154, pp. 34–43, Nov. 2017, doi: 10.1016/j.conbuildmat.2017.07.142.

[3] Q. Chen, Q. Zhang, C. Qi, A. Fourie, and C. Xiao, 'Recycling phosphogypsum and construction demolition waste for cemented paste backfill and its environmental impact', *Journal of Cleaner Production*, vol. 186, pp. 418–429, juin 2018, doi: 10.1016/j.jclepro.2018.03.131.

[4] F. Cihangir, B. Ercikdi, A. Kesimal, H. Deveci, and F. Erdemir, 'Paste backfill of high-sulphide mill tailings using alkali-activated blast furnace slag: Effect of activator nature, concentration and slag properties', *Minerals Engineering*, vol. 83, pp. 117–127, Nov. 2015, doi: 10.1016/j.mineng.2015.08.022.

[5] E. Yilmaz, B. Tikou, M. Benzaazoua, A. Kesimal, and B. Ercikdi, 'Evaluation of the strength properties of deslimed tailings paste backfill', *Mineral Resources Engineering*, vol. 12, pp. 129–144, Nov. 2013.

[6] M. Fall, M. Benzaazoua, and E. G. Saa, 'Mix proportioning of underground cemented tailings backfill', *Tunnelling and Underground Space Technology*, vol. 23, no. 1, pp. 80–90, Jan. 2008, doi: 10.1016/j.tust.2006.08.005.

[7] M. Benzaazoua, J. Ouellet, S. Servant, P. Newman, and R. Verburg, 'Cementitious backfill with high sulfur content Physical, chemical, and mineralogical characterization', *Cement and Concrete Research*, vol. 29, no. 5, pp. 719–725, mai 1999, doi: 10.1016/S0008-8846(99)00023-X.

[8] A. Kesimal, B. Ercikdi, and E. Yilmaz, 'The effect of desliming by sedimentation on paste backfill performance', *Minerals Engineering*, vol. 16, no. 10, pp. 1009–1011, Oct. 2003, doi: 10.1016/S0892-6875(03)00267-X.

[9] M. Fall, J. C. Célestin, M. Pokharel, and M. Touré, 'A contribution to understanding the effects of curing temperature on the mechanical properties of mine cemented tailings backfill', *Engineering Geology*, vol. 114, no. 3, pp. 397–413, août 2010, doi: 10.1016/j.enggeo.2010.05.016.

[10] S. Ouellet, B. Bussière, M. Aubertin, and M. Benzaazoua, 'Microstructural evolution of cemented paste backfill: Mercury intrusion porosimetry test results', *Cement and Concrete Research*, vol. 37, no. 12, pp. 1654–1665, décembre 2007, doi: 10.1016/j.cemconres.2007.08.016.

[11] M. Benzaazoua, M. Fall, and T. Belem, 'A contribution to understanding the hardening process of cemented pastefill', *Minerals Engineering*, vol. 17, no. 2, pp. 141–152, février 2004, doi: 10.1016/j.mineng.2003.10.022.

[12] V. Q. Tran, 'Hybrid gradient boosting with meta-heuristic algorithms prediction of unconfined compressive strength of stabilized soil based on initial soil properties, mix design and effective compaction', *Journal of Cleaner Production*, p. 131683, Apr. 2022, doi: 10.1016/j.jclepro.2022.131683.

[13] V. Quan Tran, 'Machine learning approach for investigating chloride diffusion coefficient of concrete containing supplementary cementitious materials', *Construction and Building Materials*, vol. 328, p. 127103, Apr. 2022, doi: 10.1016/j.conbuildmat.2022.127103.

[14] V. Q. Tran, 'Data-driven approach for investigating and predicting of compressive strength of fly ash–slag geopolymer concrete', *Structural Concrete*, vol. n/a, no. n/a, doi: 10.1002/suco.202300298.

[15] C. Qi, A. Fourie, Q. Chen, and Q. Zhang, 'A strength prediction model using artificial intelligence for recycling waste tailings as cemented paste backfill', *Journal of Cleaner Production*, vol. 183, pp. 566–578, mai 2018, doi: 10.1016/j.jclepro.2018.02.154.

[16] C. Qi, Q. Chen, A. Fourie, and Q. Zhang, 'An intelligent modelling framework for mechanical properties of cemented paste backfill', *Minerals Engineering*, vol. 123, pp. 16–27, Jul. 2018, doi: 10.1016/j.mineng.2018.04.010.

[17] L. Orejarena and M. Fall, 'Artificial neural network based modeling of the coupled effect of sulphate and temperature on the strength of cemented paste backfill', *Can. J. Civ. Eng.*, vol. 38, no. 1, pp. 100–109, Jan. 2011, doi: 10.1139/L10-109.

[18] J. Friedman, 'Stochastic gradient boosting', 2002, doi: 10.1016/S0167-9473(01)00065-2.

[19] J. Cortes and V. Vapnik, 'Support-vector networks', *Mach Learn*, vol. 20, no. 3, pp. 273–297, Sep. 1995, doi: 10.1007/BF00994018.