

MÔ HÌNH DỰ ĐOÁN CƯỜNG ĐỘ ĐẤT GIA CỐ XI MĂNG TRONG MÔI TRƯỜNG SUNFAT CAO BẰNG HỌC MÁY

Phạm Văn Ngọc^{a,*}

^a*Khoa Xây dựng Cầu đường, Trường Đại học Bách khoa, Đại học Đà Nẵng, số 54 Nguyễn Lương Bằng, Phường Liên Chiểu, TP. Đà Nẵng, Việt Nam*

Nhận ngày 30/12/2025, Sửa xong 20/04/2026, Chấp nhận đăng 21/05/2026

Tóm tắt

Nghiên cứu này nhằm phát triển công thức dự đoán để xác định cường độ nén (UCS) của đất gia cố xi măng có xét đến ảnh hưởng của nồng độ sunfat cao (100%, 200%, 500% và 1000% so với nước biển) dựa trên thuật toán lập trình di truyền (MGGP). Mô hình đánh giá tác động của hai biến độc lập, bao gồm nồng độ sunfat và thời gian dưỡng hộ mẫu, đối với cường độ của đất gia cố xi măng. Phương pháp trộn khô được áp dụng để gia cố đất yếu với hàm lượng xi măng ở mức 120 kg/m³. Kết quả cho thấy mô hình MGGP được chọn có thể dự đoán các giá trị mục tiêu với hệ số tương quan cao (giá trị R xấp xỉ 0,987) và sai số thấp (ví dụ: RMSE và MAE). Hiệu quả của mô hình MGGP được đánh giá bằng cách sử dụng một số tiêu chí bên ngoài và thuộc tính phân phối thống kê. Nghiên cứu cho thấy mô hình MGGP có khả năng dự đoán mạnh mẽ và chính xác. Các phương trình dự báo được đề xuất từ mô hình MGGP được lựa chọn có thể giúp các kỹ sư và chuyên gia tư vấn hiểu được ảnh hưởng của nồng độ sunfat đến cường độ ổn định của đất khu vực ven biển trong giai đoạn tiền quy hoạch và tiền thiết kế.

Từ khóa: cường độ chịu nén (UCS); đất gia cố xi măng; lập trình di truyền (GP); lập trình di truyền đa gen (MGGP); đất sét yếu.

MACHINE LEARNING-BASED PREDICTIVE MODEL FOR THE STRENGTH OF CEMENT-STABILIZED SOIL IN HIGH-SULFATE ENVIRONMENTS

Abstract

This study aims to develop a predictive formula for determining the unconfined compressive strength (UCS) of cement-stabilized soil, considering the effects of high sulfate concentrations (100%, 200%, 500%, and 1000% of seawater levels) using the multi-gene genetic programming (MGGP) algorithm. The model evaluates the impact of two independent variables, namely sulfate concentration and curing time, on the strength of cement-stabilized soil. The dry mixing method is applied to stabilize weak soil with a cement content of 120 kg/m³. The results indicate that the selected MGGP model can predict target values with a high correlation coefficient ($R \approx 0,987$) and low error (e.g., RMSE and MAE). The effectiveness of the MGGP model is assessed using several external criteria and statistical distribution properties. The study demonstrates that the MGGP model provides robust and accurate predictive capabilities. The proposed predictive equations derived from the selected MGGP model can assist engineers and consultants in understanding the influence of sulfate concentration on the strength stability of coastal soils during the pre-planning and preliminary design stages.

Keywords: unconfined compressive strength (UCS); cement-stabilized soil; genetic programming (GP); multi-gene genetic programming (MGGP); soft clay.

[https://doi.org/10.31814/stce.huce2026-20\(2V\)-06](https://doi.org/10.31814/stce.huce2026-20(2V)-06) © 2026 Trường Đại học Xây dựng Hà Nội (ĐHXDHN)

1. Giới thiệu

Gia cố đất là một biện pháp xử lý đất tại chỗ, trong đó đất được trộn với xi măng hoặc các chất kết dính khác. Các chất kết dính vô cơ phổ biến là xi măng và vôi [1]. Có một số yếu tố ảnh hưởng đến đặc tính ổn định của đất gia cố, chẳng hạn như loại đất, loại chất kết dính, hàm lượng chất kết dính, cách trộn và thời gian bảo dưỡng. Ở các vùng ven biển, các trụ đất xi măng có thể bị hư hỏng do tác dụng của sunfat trong nước biển [2]. Khi xi măng được trộn với đất tiếp xúc với môi trường sunfat, các phản ứng giữa các ion sunfat và các khoáng chất trong xi măng diễn ra làm giảm lượng canxi (CaO^{2+}) và canxi silicat hydrat (C–S–H), những thành phần đóng vai trò chính trong việc duy trì độ bền của đất gia cố [3]. Hơn nữa, những phản ứng này tạo ra ettringite do phản ứng của canxi

*Tác giả đại diện. Địa chỉ e-mail: pvngoc@dut.udn.vn (Ngọc, P. V.)

aluminat với canxi sunfat, làm cho cấu trúc xi măng trương nở và có khả năng bị nứt [2]. Bên cạnh đó, sự hình thành của magie hydroxit ($Mg(OH)_2$) và magie silicat hydrat (M-S-H) không góp phần vào độ bền của vật liệu đất trộn xi măng [2]. Do đó, các trụ đất-xi măng có thể bị nứt làm giảm cường độ và tăng tính thấm của trụ. Sau thời gian dài tiếp xúc với môi trường sunfat, khả năng chịu lực của các trụ này sẽ giảm đi và cuối cùng là mất khả năng chịu lực trong trường hợp bất lợi nhất.

Các nghiên cứu thực nghiệm cho thấy môi trường ven biển giàu ion sunfat gây suy giảm cường độ trụ đất xi măng theo thời gian. Rajasekaran [2] chỉ ra khi tiếp xúc lâu dài với nước biển, trụ đất xi măng giảm sức chịu rõ rệt; ví dụ trụ đường kính 0,5 m trong môi trường sunfat gấp đôi nước biển sẽ không đạt yêu cầu thiết kế sau khoảng 75 năm. Quan trắc thực tế tại công trình cho thấy ăn mòn có thể xảy ra ở lớp bề mặt và phát triển vào trong trụ gây suy giảm cường độ vật liệu. Một số báo cáo dự án ven biển đã ghi nhận trụ đất xi măng bị ăn mòn sunfat, gây sụt lún và giảm tuổi thọ công trình [4].

Trong lĩnh vực địa kỹ thuật, các phương pháp học máy (machine learning – ML) ngày càng được ứng dụng rộng rãi để mô hình hóa các bài toán phi tuyến và phức tạp, đặc biệt trong nghiên cứu vật liệu đất-xi măng. Nhiều thuật toán như Artificial Neural Network (ANN), Support Vector Machine (SVM), Random Forest (RF) và Extreme Gradient Boosting đã được ứng dụng để dự đoán các đặc trưng cơ học quan trọng như cường độ nén không nở hông (UCS), mô đun biến dạng và sức chịu tải của cọc [5]. Các mô hình này cho thấy độ chính xác dự báo cao nhờ khả năng khai thác mối quan hệ phi tuyến giữa các tham số đầu vào như hàm lượng xi măng, thời gian bảo dưỡng, độ ẩm và loại đất [6]. Tuy nhiên, phần lớn các phương pháp trên thuộc nhóm “hộp đen”, thiếu tính minh bạch và khó diễn giải về mặt cơ học, từ đó hạn chế khả năng ứng dụng trực tiếp trong thiết kế kỹ thuật [7].

Lập trình di truyền (GP) là một loại kỹ thuật học máy có giám sát áp dụng nguyên lý thuyết tiến hóa của Darwin [8]. Đó là một cách tiếp cận khác của mô hình hóa truyền thống. Thuật toán GP có thể tự học từ các dữ liệu nhập vào và tính gần đúng mối quan hệ giữa dữ liệu đầu vào và đầu ra. Lập trình di truyền đa gen (MGGP) [9] là một nhánh của GP truyền thống. Giải pháp của MGGP bao gồm một số gen được kết hợp tuyến tính với trọng số đi kèm. Các nghiên cứu trước đây chỉ ra rằng kỹ thuật MGGP giải quyết các vấn đề chính xác và hiệu quả hơn so với GP truyền thống [10]. Bên cạnh đó, MGGP là một công cụ mạnh mẽ để ước tính hoặc dự đoán các thông số địa kỹ thuật dựa trên dữ liệu đo được cung cấp [10, 11]. Điển hình, Rajabi, và cs. [12] đã nghiên cứu việc sử dụng thuật toán di truyền trong mô hình hóa quá trình ăn mòn của các cọc bê tông trong môi trường nước biển, đặc biệt là môi trường có nồng độ ion sunfat cao. Các gen trong lập trình di truyền được sử dụng để tối ưu hóa các thông số liên quan đến tốc độ ăn mòn và tuổi thọ của cọc. Ngoài ra, Hoang, và cs. [13] sử dụng phương pháp MGGP để dự đoán sự khuếch tán ion clorua trong vữa xi măng. Kết quả nghiên cứu cho thấy MGGP vượt trội so với các mô hình truyền thống nhờ khả năng xử lý mối quan hệ phi tuyến tính phức tạp giữa các yếu tố ảnh hưởng đến sự khuếch tán clorua. Do đó, thuật toán MGGP được chọn để sử dụng trong nghiên cứu này vì nó có khả năng dự đoán mạnh mẽ và chính xác. Hiệu suất của mô hình MGGP sẽ được kiểm tra dựa trên độ chính xác, sai số, tham số thống kê, một số tiêu chí bên ngoài và thuộc tính phân phối.

Nghiên cứu này tập trung vào cường độ nén không nở hông của đất yếu được gia cố bằng xi măng ở nồng độ sunfat cao (100%, 200%, 500% và 1000% so với nước biển). Ảnh hưởng của nồng độ sunfat và thời gian bảo dưỡng đến cường độ UCS của vật liệu được đánh giá nhằm làm rõ sự biến đổi về độ bền của vật liệu trong các điều kiện môi trường khắc nghiệt. Mô hình MGGP được áp dụng để xây dựng phương trình tương quan giữa các tham số. Phân tích thống kê được xem xét để đánh giá độ tin cậy của các mô hình được đề xuất. Nghiên cứu này không chỉ giúp nâng cao hiểu biết về sự tương tác giữa vật liệu xi măng và môi trường, mà còn đóng góp vào việc phát triển các giải pháp xây dựng bền vững hơn, giảm thiểu rủi ro ăn mòn và kéo dài tuổi thọ công trình trong điều kiện môi trường biển.

2. Chuẩn bị dữ liệu

Nghiên cứu này lấy dữ liệu thí nghiệm từ nghiên cứu của Phạm [14]. Loại đất được sử dụng trong nghiên cứu được thu thập từ dự án cải tạo cầu vượt đường sắt tại Newcastle, bang New South Wales, Úc. Đất á cát có độ ẩm tự nhiên 37,7%, giới hạn chảy (LL) là 25%, giới hạn dẻo (PL) là 19%, và dung trọng là 1,68 T/m³. Quá trình thí nghiệm và kết quả thí nghiệm UCS được tóm tắt như sau. Xi măng và đất được trộn theo phương pháp trộn khô với hàm lượng xi măng 120 kg/m³. Các mẫu được chuẩn bị theo tiêu chuẩn Nhật Bản [15]. Các khuôn PVC có đường kính 53,2 mm và chiều cao 106,4 mm đã được chế tạo. Các mẫu được bảo dưỡng trong phòng với độ ẩm không đổi 95% và 22 °C trong 1 đến 388 ngày bảo dưỡng. Sau 28 ngày bảo dưỡng, các mẫu đất-xi măng được ngâm trong các nồng độ sunfat khác nhau (cao gấp 1 ÷ 10 lần nồng độ nước biển thông thường), cụ thể là 100%, 200%, 500% và 1000% nước biển. Thí nghiệm sử dụng Pooclang xi măng loại I (CEM I 42,5N). Thí nghiệm cường độ nén không nở hông (UCS) được sử dụng để đo cường độ của đất gia cố xi măng theo tiêu chuẩn ASTM-D1632-17 [16]. Tất cả các thử nghiệm UCS được thực hiện với tốc độ biến dạng không đổi, xấp xỉ 1% chiều cao mẫu mỗi phút (hoặc 1 mm/phút). Bảng 1 tóm tắt các kết quả UCS về ổn định đất ở các nồng độ sunfat khác nhau.

Bảng 1. Thông số thống kê của dữ liệu đầu vào

Tham số thống kê	Thời gian bảo dưỡng (Ngày)	Nồng độ sunfat (%)	Cường độ UCS (kPa)
Giá trị lớn nhất	388	1000	3002
Giá trị nhỏ nhất	1	0	83
Giá trị trung bình	140	316	1541
Độ lệch chuẩn	118	360	753
Biên độ	387	1000	2919

3. Lập trình Di Truyền Đa Gen (MGGP)

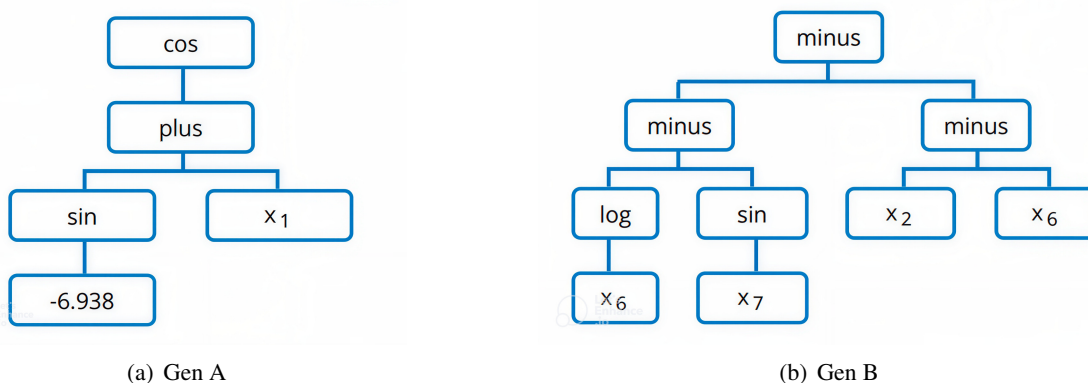
3.1. Lập trình di truyền (GP)

Kỹ thuật GP được giới thiệu bởi và được phát triển bởi Koza [8]. Trong GP, kết quả là một chương trình máy tính có thể được thực thi mà không cần xử lý hậu kỳ [8]. Chúng được biểu diễn giải pháp dưới dạng cấu trúc dạng cây, được gọi là GP dạng cấu trúc cây (GP truyền thống) [8]. Chương trình GP có thể tạo các phương trình dự đoán mà không cần bất kỳ hình thức giả định nào trước đó của mối quan hệ hiện có. Nghiên cứu đã phát hiện ra rằng GP có thể mô hình hóa hành vi phức tạp của các vấn đề địa kỹ thuật với độ chính xác cao và tạo ra kết quả rõ ràng và minh bạch [17]. Các phương pháp GP đã được áp dụng trong nhiều nghiên cứu về địa kỹ thuật và xây dựng nói chung [10].

3.2. Lập trình Di Truyền Đa Gen (MGGP)

Lập trình di truyền đa gen [18] là một nhánh của GP truyền thống. Giải pháp của MGGP bao gồm một số gen được kết hợp tuyến tính với trọng số đi kèm [18]. Hệ số tuyến tính của từng gen của nhiễm sắc thể (cá thể) được xác định bằng phương pháp bình phương nhỏ nhất thông thường [19]. Hình 1 minh họa cấu trúc MGGP điển hình, trong đó kết quả chương trình chứa hai gen tổ hợp có trọng số tuyến tính (Gen A và Gen B). Mỗi gen chứa mối quan hệ phi tuyến tính giữa các biến đầu vào (x_1, x_2, x_6, x_7). Kết quả này được minh họa dưới dạng phương trình toán học: $y = \cos(\sin(-6,938) + x_1) + \ln(x_6) - \sin(x_7) - (x_2 - x_6)$ (trong đó “plus” là toán tử “+”, “minus” là toán tử “-”, “sin” và “cos” là các toán tử lượng giác).

Quần thể ban đầu trong mô hình MGGP bao gồm các cá thể chứa các gen khác nhau được tạo ngẫu nhiên dựa trên các hàm chức năng và bộ đầu cuối cũng như các biến [20]. Các toán tử GP truyền



Hình 1. Ví dụ về cấu trúc MGGP

thông được áp dụng trong MGGP. Số lượng gen tối đa trong mỗi cá thể và độ sâu tối đa của gen là các thông số quan trọng trong mô hình MGGP. Gần đây, mô hình MGGP cũng được áp dụng để giải quyết nhiều bài toán địa kỹ thuật [21, 22]. Tóm lại, các nhà nghiên cứu đã chỉ ra rằng MGGP giải quyết các vấn đề chính xác và hiệu quả hơn so với GP truyền thống và các mô hình dựa trên hồi quy tuyến tính [10, 11]. Hơn nữa, các phương trình dựa trên MGGP rất đơn giản, thiết thực và có thể được sử dụng dễ dàng [10]. Mặc dù mô hình MGGP cho thấy nhiều ưu điểm trong việc phát triển các phương trình dự đoán chính xác, nhưng có ít nghiên cứu áp dụng mô hình MGGP để đánh giá các đặc tính của đất gia cố chất vô cơ.

3.3. Phát triển mô hình MGGP

GPTIPS là một công cụ trong MATLAB được dùng để phân tích MGGP [9]. GPTIPS có thể tự động tạo các phương trình chính xác để dự đoán hành vi của các hệ thống vật lý [23]. Trong GPTIPS, hầu hết các tính năng hoạt động tự động với một vài thông số cài đặt. Ngoài ra, GPTIPS có thể cung cấp báo cáo HTML chứa tất cả thông tin về thông số chạy, biểu đồ và kết quả [24]. GPTIPS đã được áp dụng để giải quyết các vấn đề về vật liệu, địa kỹ thuật và kết cấu [19]. Trong GPTIPS, sai số bình phương trung bình gốc (RMSE) được chọn làm hàm chức năng mặc định để giảm thiểu sai số giữa số liệu đầu ra đo được và kết quả dự đoán [19]. Giá trị phù hợp của cả quá trình huấn luyện và kiểm tra, cũng như tính đơn giản của thuật toán, là các tiêu chí để lựa chọn các mô hình tốt nhất [19]. Bên cạnh đó, các đại lượng MAE, R và RMSE được xem xét để đánh giá tính chính xác của mô hình được đề xuất.

Trong nghiên cứu này, dữ liệu đầu vào bao gồm hai biến độc lập, bao gồm thời gian bảo dưỡng - t ($1 \div 388$ ngày) và nồng độ sunfat - S (100%, 200%, 500% và 1000%), trong khi dữ liệu đầu ra là cường độ UCS của đất gia cố. Phương trình dự đoán của UCS dựa trên các biến đầu vào được phát triển dưới dạng $UCS = f(t, S)$. Trong mô hình MGGP, các biến độc lập (t và S) được gán lần lượt là x_1 và x_2 . Hình 2 minh họa MATLAB code dùng để phân tích MGGP. Dữ liệu được chia ngẫu nhiên thành các tập huấn luyện và tập kiểm tra. Tập dữ liệu huấn luyện được sử dụng để phát triển mô hình, trong khi tập dữ liệu kiểm tra được sử dụng để đánh giá hiệu suất của mô hình đã chọn [19]. Trong nghiên cứu này, việc kết hợp dữ liệu huấn luyện và kiểm tra đã được thử nhiều lần để chọn ra tập dữ liệu phù hợp nhất. Kết quả là, trong số 114 dữ liệu từ đất gia cố xi măng, 90 dữ liệu được sử dụng để huấn luyện mô hình và 24 dữ liệu được sử dụng để kiểm tra mô hình đề xuất. Giá trị trung bình, giá trị lớn nhất, giá trị nhỏ nhất, độ lệch chuẩn và biên độ tham số được xem xét trong tập dữ liệu huấn luyện và kiểm tra. Bảng 2 cho thấy số liệu thống kê mô tả cho các biến đầu vào và đầu ra.

```
my_configNgoc.m

function gp=my_configNgoc(gp);
gp.runcontrol.pop_size=1000;
gp.runcontrol.num_gen=1000;
gp.fitness.fitfun=@regressmulti_fitfun;
gp.fitness.minimisation=true;
load mydata
gp.userdata.xtrain=xtrain;
gp.userdata.ytrain=ytrain;
gp.userdata.xtest=xtest;
gp.userdata.ytest=ytest;
gp.nodes.inputs.num_inp=size(gp.userdata.xtrain,2);
gp.genes.multigene=true;
gp.genes.max_genes=4;
gp.genes.max_depth=4;
gp.nodes.functions.name={'times','minus','plus','sin','cos',
'exp','sqrt','power','log'};
```

Hình 2. MATLAB code dùng để phân tích MGGP

Bảng 2. Thống kê mô tả dữ liệu đầu vào và đầu ra

Tham số	Tập giá trị	Thời gian bảo dưỡng (Ngày)	Nồng độ sunfat (%)	Cường độ UCS (kPa)
Giá trị lớn nhất	Tập huấn luyện	388	1000	3002
	Tập kiểm tra	388	1000	2725
Giá trị nhỏ nhất	Tập huấn luyện	1	0	90
	Tập kiểm tra	58	0	83
Giá trị trung bình	Tập huấn luyện	137	310	1534
	Tập kiểm tra	151	338	1567
Độ lệch chuẩn	Tập huấn luyện	123	367	765
	Tập kiểm tra	96	339	720
Biên độ	Tập huấn luyện	387	1000	2913
	Tập kiểm tra	330	1000	2641

Các tham số trong chương trình MGGP đã được chọn dựa trên các giá trị được đề xuất của Searson, và cs. [25], xem xét nghiên cứu trước đó, và phương pháp tiếp cận sau thử nghiệm và sai sót [19]. Trong mỗi lần thử, giá trị của một tham số được thay đổi, trong khi các tham số khác được đặt không đổi. Số lượng gen và kích thước quần thể là những thông số quan trọng ảnh hưởng đến hiệu suất của mô hình. Một số lượng lớn gen có thể làm tăng độ phức tạp của mô hình và một lượng lớn cá thể yêu cầu thêm thời gian để phân tích mô hình. Do đó, mô hình đã được đào tạo với bốn gen; kích thước quần thể là 1000. Khi sai số của hai tập hợp con nhỏ và càng gần càng tốt, giá trị của tham số được chọn. Các bộ hàm bao gồm nhân, trừ, cộng, sin, cos, exp, sqrt, lũy thừa và log đã được chọn để phát triển mô hình. Bảng 3 trình bày tóm tắt các tham số cài đặt cho mô hình MGGP.

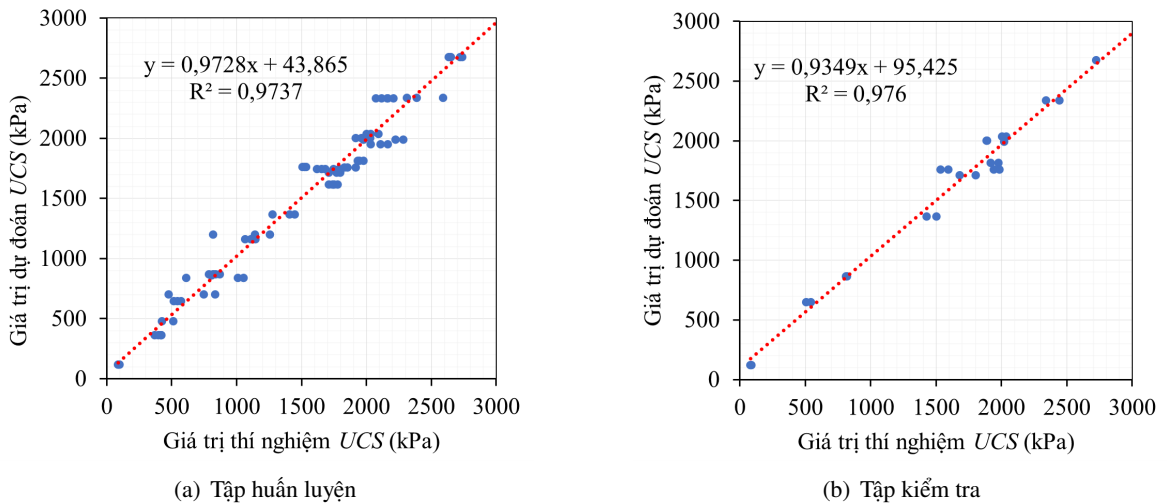
Bảng 3. Cài đặt các tham số cho mô hình MGGP

Các tham số	Giá trị
Kích thước quần thể	1000
Số lượng biến đầu vào	2
Kích thước vòng lặp	10
Số lượng gen	4
Chiều dài tối đa của gen	4
Các hàm chức năng	times, minus, plus, sin, cos, exp, sqrt, power, log

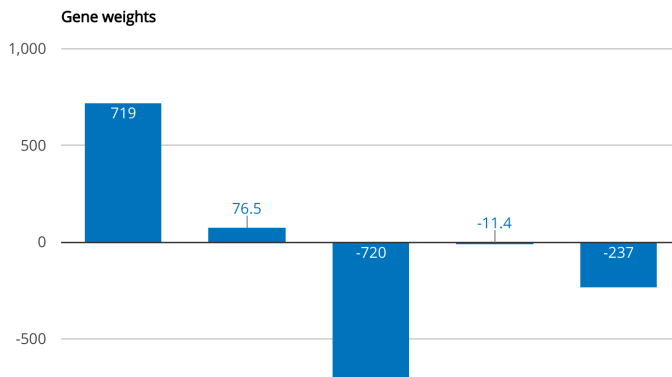
Kết quả MGGP được trình bày trong Hình 3 đến Hình 5. Trong Hình 3, các giá trị cường độ UCS dự đoán từ mô hình MGGP đã chọn được so sánh với dữ liệu UCS đo được (thực tế). Rõ ràng là mô hình MGGP được đề xuất có khả năng dự đoán cao với giá trị R xấp xỉ 1,0 cho tập huấn luyện và 0,965 cho tập dữ liệu kiểm tra. Ngoài ra, Hình 4 cho thấy các giá trị sai lệch và trọng số của bốn gen trong mô hình MGGP đã chọn. Cấu trúc cây của các gen này được thể hiện trong Hình 5. Công thức toán học thu được từ mô hình MGGP đã chọn trong sự kết hợp của độ lệch, trọng số và gen được hiển thị trong phương trình (1). Công thức này đơn giản và dễ sử dụng trong thực tế. Hiệu suất thống kê của mô hình MGGP cho các tập huấn luyện và tập kiểm tra và toàn bộ tập dữ liệu được trình bày trong Bảng 4.

$$\begin{aligned}
 UCS = & 76,5 \sin(t^{1/2}) - 237 \cos(\cos S + t^2) + 720 \ln(S + t) \\
 & - 76,5 \cos S - t^{1/2}(11,4S^{1/4} + 76,5) - 720(\ln t)^{1/2} + 719
 \end{aligned} \tag{1}$$

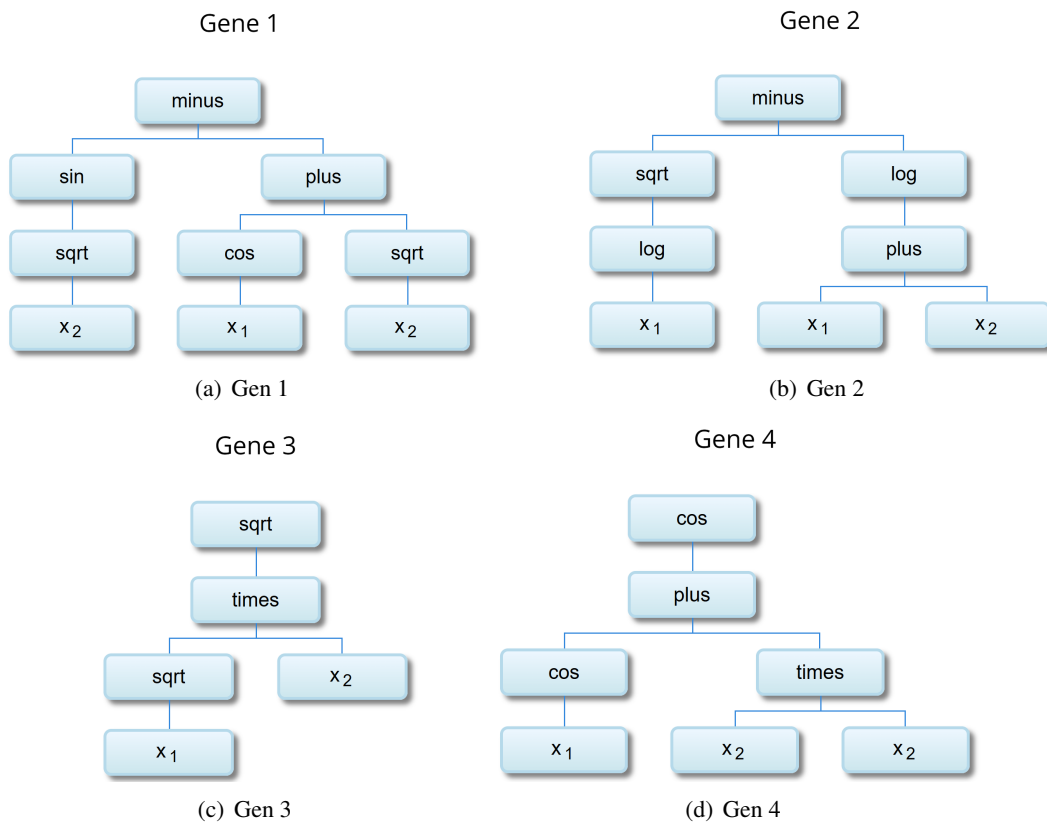
trong đó UCS là cường độ nén không giới hạn của đất gia cố xi măng; S là nồng độ sunfat (%); và t là thời gian bảo dưỡng (ngày).



Hình 3. Các giá trị UCS được dự đoán và đo thực tế trong các giai đoạn huấn luyện và kiểm tra trong mô hình MGGP đã chọn



Hình 4. Giá trị sai lệch và trọng số của các gen trong mô hình MGGP đã chọn



Hình 5. Bốn gen trong mô hình MGGP đã chọn

Bảng 4. Hiệu suất của mô hình MGGP được chọn

Tham số	Tập huấn luyện	Tập kiểm tra	Toàn bộ dữ liệu
R-value	0,987	0,988	0,987
R ² -value	0,974	0,976	0,974
RMSE (kPa)	123,43	113,33	121,37
MAE (kPa)	95,33	93,01	94,84

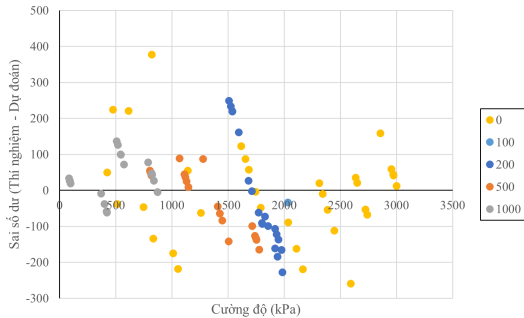
Bảng 5 và 6 minh họa sai số của mô hình tương ứng các miền giá trị theo nồng độ sunfat và theo thời gian. Cụ thể, RMSE và MAE có xu hướng tăng từ khi nồng độ sunfat tăng 0 đến 200%, đạt giá trị lớn nhất tại mức 200% (RMSE ≈ 156,11; MAE ≈ 140,02), cho thấy mô hình gặp nhiều khó khăn trong việc dự đoán ở miền này. Tuy nhiên, khi nồng độ sunfat tiếp tục tăng (500% và 1000%), sai số lại giảm đáng kể (RMSE giảm xuống 87,97 và 65,71), cho thấy mô hình hoạt động ổn định hơn trong các điều kiện sunfat rất cao. Trong khi đó, sai số của mô hình thể hiện sự phụ thuộc rõ rệt vào thời gian dưỡng hệ. Cụ thể, ở giai đoạn sớm (0 ÷ 58 ngày), sai số tương đối lớn (RMSE ≈ 138,04; MAE ≈ 115,49), cho thấy mô hình gặp khó khăn trong việc dự đoán cường độ khi quá trình thủy hóa và phát triển cấu trúc đất-xi măng còn chưa ổn định. Trong khoảng thời gian trung gian (58 ÷ 118 ngày), sai số giảm đáng kể (RMSE ≈ 73,25; MAE ≈ 51,33), phản ánh khả năng dự đoán tốt hơn của mô hình khi vật liệu đạt trạng thái phát triển cường độ tương đối ổn định. Tuy nhiên, ở giai đoạn dài hạn (208 ÷ 388 ngày), sai số lại tăng lên (RMSE ≈ 123,65; MAE ≈ 96,75), cho thấy sự suy giảm độ chính xác của mô hình.

Bảng 5. Sai số của mô hình tương ứng các miền giá trị của nồng độ sunfat

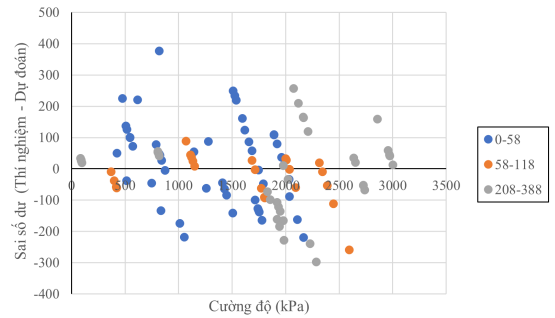
Nồng độ SO ₄ (%)	0	100	200	500	1000
RMSE (kPa)	131,57	134,59	156,11	87,97	65,71
MAE (kPa)	99,720	99,72	140,02	76,71	54,64

Bảng 6. Sai số của mô hình tương ứng các miền giá trị của thời gian

Thời gian	0–58	58–118	208–388
RMSE (kPa)	138,04	73,25	123,65
MAE (kPa)	115,49	51,33	96,75



Hình 6. Đồ thị sai số dư theo các mức nồng độ sunfat



Hình 7. Đồ thị sai số dư theo các mức thời gian

Hình 6 minh họa biểu đồ sai số dư theo các mức nồng độ sunfat. Ở các mức sunfat thấp (0 ÷ 100%), sai số dư phân tán tương đối rộng nhưng vẫn xoay quanh đường 0. Trong khi đó, tại các mức sunfat trung bình đến cao (200 ÷ 500%), xuất hiện xu hướng sai số dư âm chiếm ưu thế trong vùng cường độ trung bình (khoảng 1500 ÷ 2000 kPa), cho thấy mô hình có xu hướng dự đoán cao hơn giá trị thực. Đáng chú ý, ở mức sunfat rất cao (500 ÷ 1000%), các điểm dữ liệu tập trung hơn, ít phân tán đáng kể phản ánh mức độ ổn định gia tăng dưới điều kiện xâm thực mạnh. Bên cạnh đó, Hình 7 cho thấy các giá trị sai số dư nhìn chung phân bố quanh đường 0, chứng tỏ mô hình không có sai lệch hệ thống rõ rệt trên toàn bộ tập dữ liệu. Khi xét theo từng miền, sự khác biệt về đặc trưng sai số trở nên rõ ràng hơn. Ở miền thấp (0 ÷ 58 ngày), các điểm sai số dư phân tán khá rộng, bao gồm cả các giá trị dương lớn (lên tới khoảng 350 ÷ 400) và âm sâu (dưới -200), cho thấy mô hình có độ biến động sai số cao và chưa ổn định trong vùng này. Điều này có thể liên quan đến sự thiếu dữ liệu hoặc độ nhạy cao của cường độ đối với các yếu tố đầu vào ở mức thấp. Trong miền trung bình (58 ÷ 118 ngày), các sai số dư có xu hướng tập trung hơn và phần lớn nằm gần vùng từ -100 đến 100. Đối với miền cao (208 ÷ 388 ngày), mặc dù số lượng điểm không lớn, nhưng có thể nhận thấy độ phân tán của sai số dư tăng lên, đặc biệt xuất hiện các giá trị âm lớn (dưới -250), cho thấy mô hình có xu hướng giảm độ chính xác trong điều kiện giá trị biến đầu vào cao.

3.4. Nghiên cứu thống kê về mô hình MGGP

Golbraikh và cs. [26] đã cung cấp một số tiêu chí để đánh giá các tập dữ liệu thí nghiệm. Đầu tiên, một trong các hệ số góc của đường hồi quy (k hoặc k') qua gốc phải xấp xỉ bằng 1. Ngoài ra, chỉ số m và n (chỉ số hiệu suất) phải thấp hơn 0,1. Roy và cs. [27] đề xuất rằng chỉ số xác nhận (R_m) về khả năng dự đoán bên ngoài của các mô hình nên cao hơn 0,5. Ngoài ra, các hệ số tương quan bình phương giữa giá trị dự đoán và thí nghiệm (R_0^2 hoặc $R_0'^2$) được khuyến nghị gần bằng 1 [10, 28].

Bảng 7 trình bày kết quả của các tiêu chí này. Mô hình được lựa chọn đáp ứng các điều kiện cần thiết và cho thấy hiệu suất rất tốt với khả năng dự đoán chính xác. Ngoài ra, các tiêu chí khác để đánh giá hiệu suất của các mô hình MGGP là dựa trên các thuộc tính thống kê của chúng. Bảng 8 cho thấy giá trị tối đa, tối thiểu, giá trị trung bình, độ lệch chuẩn và phạm vi của dữ liệu tính toán từ mô hình MGGP được đề xuất và dữ liệu đo được thực tế. Các kết quả thống kê của mô hình MGGP khá gần với dữ liệu thử nghiệm.

Bảng 7. Các tham số thống kê để xác thực độ chính xác của mô hình MGGP

Tham số thống kê	Yêu cầu	Mô hình MGGP
$k = \sum_{i=1}^n (O_i \cdot P_i) / \sum_{i=1}^n O_i^2$	$0,85 < k < 1,15$	0,994
$k' = \sum_{i=1}^n (O_i \cdot P_i) / \sum_{i=1}^n P_i^2$	$0,85 < k' < 1,15$	1,001
$R_m = R_0^2 \cdot (1 - \sqrt{ R^2 - R_0^2 })$	$R_m > 0,5$	0,938
$R_0^2 = 1 - \sum_{i=1}^n (P_i - O_i^0)^2 / \sum_{i=1}^n (P_i - \bar{P}_i)^2$ trong đó $O_i^0 = k' \cdot O_i$	Gần bằng 1	0,973
$R_0'^2 = 1 - \sum_{i=1}^n (O_i - P_i^0)^2 / \sum_{i=1}^n (O_i - \bar{O}_i)^2$ trong đó $P_i^0 = k \cdot P_i$	Gần bằng 1	0,974
$m = (R^2 - R_0^2) / R^2$	$m < 0,1$	0,001
$n = (R^2 - R_0'^2) / R^2$	$n < 0,1$	0,001

Bảng 8. Thuộc tính thống kê của cường độ đất gia cố xi măng tính từ mô hình MGGP và dữ liệu đo được

Tham số	Dữ liệu thực tế	Kết quả tính từ mô hình MGGP
Giá trị lớn nhất	3002	3015
Giá trị nhỏ nhất	83	117
Giá trị trung bình	1541	1541
Độ lệch chuẩn	753	737
Biên độ	2919	2898

4. Kết luận

Nghiên cứu này đã phát triển mô hình dự đoán cường độ chịu nén không nở hông (UCS) của đất gia cố xi măng trong môi trường có nồng độ sunfat cao, sử dụng phương pháp Lập trình di truyền đa gen (MGGP). Mô hình được xây dựng dựa trên dữ liệu thực nghiệm, trong đó hai biến độc lập quan trọng là nồng độ sunfat và thời gian dưỡng hộ được xem xét nhằm đánh giá ảnh hưởng của chúng đến cường độ UCS của đất gia cố xi măng. Kết quả nghiên cứu cho thấy mô hình MGGP có khả năng dự đoán chính xác giá trị UCS với hệ số tương quan cao ($R \approx 0,987$), đồng thời sai số dự đoán thấp ($RMSE < 123$ kPa và $MAE < 95$ kPa). Điều này chứng tỏ MGGP là một công cụ mạnh mẽ để mô hình hóa mối quan hệ phi tuyến giữa các yếu tố ảnh hưởng và cường độ chịu nén của đất ổn định bằng xi măng trong điều kiện môi trường khắc nghiệt. Hiệu suất của mô hình MGGP đã được đánh giá dựa trên một số tiêu chí bên ngoài và các thuộc tính phân phối thống kê. Kết quả phân tích cho thấy mô hình không chỉ có độ chính xác cao mà còn có tính ổn định tốt, phù hợp để áp dụng trong các điều kiện thực tế.

Bên cạnh đó, nghiên cứu cũng nhấn mạnh tầm quan trọng của việc xem xét tác động của môi trường có nồng độ sunfat cao đối với trụ đất gia cố xi măng, đặc biệt trong các khu vực ven biển hoặc khu vực có điều kiện địa chất phức tạp. Việc sử dụng mô hình MGGP không chỉ giúp dự đoán chính xác mức độ ảnh hưởng của sunfat đến cường độ UCS mà còn hỗ trợ các kỹ sư trong quá trình thiết kế nền móng, tối ưu hóa việc sử dụng xi măng và lựa chọn phương pháp xử lý nền đất phù hợp.

Nghiên cứu này đề xuất các phương trình dự báo có thể được sử dụng như một công cụ hỗ trợ kỹ sư trong việc ước tính nhanh cường độ nén của đất gia cố xi măng trong môi trường sunfat cao, đặc biệt trong giai đoạn tiền quy hoạch và thiết kế sơ bộ. Các kỹ sư có thể áp dụng trực tiếp mô hình bằng cách nhập các giá trị của nồng độ sunfat và thời gian dưỡng hộ trong phạm vi nghiên cứu để dự đoán cường độ tương ứng. Kết quả dự báo có thể giúp giảm số lượng thí nghiệm cần thiết và hỗ trợ ra quyết định trong lựa chọn giải pháp xử lý nền phù hợp. Tuy nhiên, cần lưu ý rằng mô hình MGGP được phát triển dựa trên một loại đất và các điều kiện thí nghiệm cụ thể trong nghiên cứu này. Vì vậy, mô hình chủ yếu áp dụng cho các loại đất có đặc tính tương tự và trong phạm vi các biến đã khảo sát. Khi áp dụng cho các loại đất hoặc điều kiện khác, mô hình cần được hiệu chỉnh hoặc cập nhật bằng dữ liệu thực nghiệm bổ sung để đảm bảo độ tin cậy của kết quả dự báo.

Tài liệu tham khảo

- [1] Donald, B. (2000). *An introduction to the deep soil mixing methods as used in geotechnical applications*.
- [2] Rajasekaran, G. (2005). [Sulphate attack and ettringite formation in the lime and cement stabilized marine clays](#). *Ocean Engineering*, 32(8):1133–1159.
- [3] Rajasekaran, G., Rao, S. (2005). [Sulphate attack in lime-treated marine clay](#). *Marine Georesources and Geotechnolgy*, 23(1-2):93–116.
- [4] Li, W., Fu, Z., Xiao, H., Xi, L., Yang, K. (2025). [Effect of sulphate type on the strength and swelling properties of lime-GGBS-stabilised sulphate-bearing soil](#). *International Journal of Pavement Engineering*, 26(1):2464203.
- [5] Shao, W., Yue, W., Zhang, Y., Zhou, T., Zhang, Y., Dang, Y., Wang, H., Feng, X., Chao, Z. (2023). [The application of machine learning techniques in geotechnical engineering: A review and comparison](#). *Mathematics*, 11(18):3976.
- [6] Chian, S. C., Bi, J. (2025). [Data-driven prediction of strength in cement-treated clayey soils](#). *Journal of Rock Mechanics and Geotechnical Engineering*.
- [7] Harle, S. M., Wankhade, R. L. (2025). [Machine learning techniques for predictive modelling in geotechnical engineering: a succinct review](#). *Discover Civil Engineering*, 2(1):86.
- [8] Koza, J. R. (1992). *Genetic programming: on the programming of computers by means of natural selection*. MIT press, Cambridge, MA, USA.
- [9] Searson, D. (2009). *GPTIPS: Genetic programming and symbolic regression for MATLAB*.
- [10] Soleimani, S., Rajaei, S., Jiao, P., Sabz, A., Soheilinia, S. (2018). [New prediction models for unconfined compressive strength of geopolymer stabilized soil using multi-gen genetic programming](#). *Measurement*, 113:99–107.
- [11] Kurugodu, H., Bordoloi, S., Hong, Y., Garg, A., Garg, A., Sreedeeep, S., Gandomi, A. H. (2018). [Genetic programming for soil-fiber composite assessment](#). *Advances in Engineering Software*, 122:50–61.
- [12] Rajabi, Z., Eftekhari, M., Ghorbani, M., Ehteshamzadeh, M., Beirami, H. (2022). [Prediction of the degree of steel corrosion damage in reinforced concrete using field-based data by multi-gene genetic programming approach](#). *Soft Computing*, 26(18):9481–9496.
- [13] Hoang, N.-D., Chen, C.-T., Liao, K.-W. (2017). [Prediction of chloride diffusion in cement mortar using multi-gene genetic programming and multivariate adaptive regression splines](#). *Measurement*, 112:141–149.
- [14] Pham, N. V. (2016). *Durability of soil-cement columns in coastal areas*. Master Thesis, School of Engineering, Faculty of Engineering and Built Environment, The University of Newcastle.
- [15] JGS, JGS 0821-2000 (2005). *Practice for Making and Curing Stabilized Soil Specimens Without Compaction*. Japanese Geotechnical Society, Tokyo, Japan.

- [16] ASTM, ASTM D 1632-2017 (2017). *Standard Practice for Making and Curing Soil-Cement Compression and Flexure Test Specimens in the Laboratory*. American Society for Testing and Materials, West Conshohocken, PA, USA.
- [17] Pham, V.-N., Oh, E., Ong, D. E. (2022). [Effects of binder types and other significant variables on the unconfined compressive strength of chemical-stabilized clayey soil using gene-expression programming](#). *Neural Computing and Applications*, 34(11):9103–9121.
- [18] Searson, D., Willis, M., Montague, G. (2007). [Co-evolution of non-linear PLS model components](#). *Journal of Chemometrics: A Journal of the Chemometrics Society*, 21(12):592–603.
- [19] Gandomi, A. H., Alavi, A. H. (2012). [A new multi-gene genetic programming approach to non-linear system modeling. Part II: geotechnical and earthquake engineering problems](#). *Neural Computing and Applications*, 21(1):189–201.
- [20] Muduli, P. K., Das, S. K., Bhattacharya, S. (2014). [CPT-based probabilistic evaluation of seismic soil liquefaction potential using multi-gene genetic programming](#). *Georisk: Assessment and Management of Risk for Engineered Systems and Geohazards*, 8(1):14–28.
- [21] Barati, R., Mehraein, M., Alizadeh, M. J., Atashi, V., Mohajeri, S. H. (2025). [Temporal Evaluation of Scour Hole Dimensions due to Plain Wall Jets in Noncohesive Sediments Using a Soft Computing Approach: White-Box versus Black-Box Modeling](#). *Journal of Hydrologic Engineering*, 30(2):04024060.
- [22] Hasan, M. M., Rahman, M. Z., Fahim, A. K. F. (2025). [Earthquake induced liquefaction hazard analysis for Chittagong City using machine learning](#). *Geomatics, Natural Hazards and Risk*, 16(1):2451126.
- [23] Searson, D. P. (2015). GPTIPS 2: an open-source software platform for symbolic data mining. *Handbook of genetic programming applications*, Springer, Cham, 551–573.
- [24] Gandomi, A. H., Atefi, E. (2019). [Software review: the GPTIPS platform](#). *Genetic Programming and Evolvable Machines*, 20(1):1–8.
- [25] Searson, D. P., Leahy, D. E., Willis, M. J. (2010). GPTIPS: an open source genetic programming toolbox for multigene symbolic regression. *Proceedings of the International multiconference of engineers and computer scientists*, Hong Kong, 1:77–80.
- [26] Golbraikh, A., Tropsha, A. (2002). [Beware of q²!](#) *Journal of Molecular Graphics and Modelling*, 20(4): 269–276.
- [27] Roy, P. P., Roy, K. (2008). [On some aspects of variable selection for partial least squares regression models](#). *QSAR & Combinatorial Science*, 27(3):302–313.
- [28] Alavi, A. H., Gandomi, A. H., Mollahasani, A., Bazaz, J. B., Talatahari, S. (2013). Linear and tree-based genetic programming for solving geotechnical engineering problems. *Metaheuristics in water, geotechnical and transport engineering*, Elsevier, Amsterdam, 289–310.