# USO DO APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DA RESISTÊNCIA A COMPRESSÃO DO CONCRETO COM MATERIAL CIMENTÍCIO SUPLEMENTAR

**Maria Heloisa Rodrigues Siqueira**
Pontifícia Universidade Católica do Rio de Janeiro | Rio de Janeiro, RJ | helloisarodrigues.hr@gmail.com

**Alexandre Miguel Silva Araujo**
Pontifícia Universidade Católica do Rio de Janeiro | Rio de Janeiro, RJ | alexandremiguelsa@gmail.com

**Rafael Otávio Alves Abreu**
Universidade Estadual do Maranhão | São Luís, MA
Instituto Tecgraf, Pontifícia Universidade Católica do Rio de Janeiro | Rio de Janeiro, RJ | rafaeloabreu@tecgraf.puc-rio.br

**Brunno Ferreira dos Santos**
Pontifícia Universidade Católica do Rio de Janeiro | Rio de Janeiro, RJ | bsantos@puc-rio.br

**Lourdes Maria Silva de Souza**
Pontifícia Universidade Católica do Rio de Janeiro | Rio de Janeiro, RJ | lourdesmariasouza@puc-rio.br

## Resumo

*A produção de cimento Portland (CP) é uma das principais fontes de emissões de $CO_2$. Para reduzir seu impacto ambiental, materiais cimentícios suplementares (MCS), como sílica, cinza volante e escória de alto-forno, têm sido utilizados, melhorando as propriedades do concreto e reutilizando subprodutos industriais. No entanto, a dosagem de concretos com MCS ainda enfrenta desafios práticos. Com os avanços tecnológicos, o aprendizado de máquina (ML) tem se mostrado uma ferramenta promissora na engenharia civil, permitindo otimizar a dosagem e prever a resistência à compressão (RC) do concreto. Este estudo aplica sete modelos de ML para prever a RC de concretos com MCS, avaliando as variáveis mais influentes na previsão. O banco de dados conta com 1.568 amostras experimentais de três fontes do Scopus, considerando variáveis como conteúdo de cimento, sílica, cinza volante, escória, agregado miúdo, agregado graúdo, água, aditivos e idade. Os resultados preliminares indicam que modelos baseados em Gradient Boosting, como o LGBM, apresentam o melhor desempenho, com $R^2$ = 0,9773, sugerindo alta acurácia na previsão da RC. A análise de importância das variáveis destacou o conteúdo de cimento, a idade e o teor de água como os fatores mais influentes na resistência do concreto.*

**Palavras-chave:** Concreto; Material cimentício suplementar; Inteligência artificial; Aprendizado de máquina; Resistência à compressão.

## ABSTRACT

*The production of Portland cement (PC) is one of the main sources of $CO_2$ emissions. To mitigate its environmental impact, supplementary cementitious materials (SCMs), such as silica fume, fly ash, and ground granulated blast furnace slag, have been widely used, enhancing concrete properties while repurposing industrial by-products. However, the mix proportioning of concrete incorporating SCMs still presents practical challenges. With technological advancements, machine learning (ML) has emerged as a promising tool in civil engineering, enabling the optimization of mix designs and the prediction of concrete compressive strength (CS). This study applies seven ML models to predict the CS of concrete containing SCMs, assessing the most influential variables in the prediction. The database comprises 1,568 experimental samples collected from three sources in Scopus, considering variables such as cement content, silica fume, fly ash, slag, fine aggregate, coarse aggregate, water, admixtures, and curing age. Preliminary results indicate that Gradient Boosting-based models, such as LGBM, achieve the best performance, with $R^2$ = 0.9773, suggesting high accuracy in CS prediction. The variable importance analysis identified cement content, curing age, and water content as the most influential factors affecting concrete strength.*

**Keywords:** Concrete; Supplementary cementitious material; Artificial intelligence; Machine learning; Compressive strength.

# 1 INTRODUÇÃO

O concreto, amplamente empregado na construção civil, é composto principalmente por cimento Portland (CP), cujo principal componente é o clínquer, resultante da fusão de calcário e argila a altas temperaturas. Materiais cimentícios suplementares (MCS), como pozolanas, são uma alternativa para substituir parcialmente o CP. Eles reagem com o hidróxido de cálcio formado na hidratação do CP, melhorando a resistência e durabilidade do concreto (MEHTA E MONTEIRO, 2014) e reaproveitando subprodutos industriais (RADWAN et al., 2021). No entanto, sua aplicação prática é dificultada pela complexidade na dosagem, como o método de empacotamento de partículas (TUTIKIAN E HELENE, 2011).

O aprendizado de máquina (ML) tem sido utilizado na engenharia civil para prever o desempenho do concreto, com destaque para a previsão da resistência à compressão (RC), reduzindo custos e tempo (CUI et al., 2021). Contudo, a acurácia depende do modelo, dados e parâmetros utilizados, sendo necessária a regularização para garantir resultados confiáveis (AL-SHAMIRI et al., 2020). A generalização dos modelos é dificultada pela variação nas propriedades dos materiais e pela insuficiência de dados representativos (CHEN et al., 2022), além da desconsideração de fatores ambientais e outras propriedades mecânicas, como a resistência à tração na flexão (AHMAD et al., 2021). Diante desse cenário, este estudo tem como objetivo compilar dados experimentais para criar uma base de dados abrangente, avaliar o desempenho de sete modelos de ML para previsão da RC do concreto com MCS e identificar as variáveis mais influentes, tendo em vista aplicações práticas de tais modelos.

# 2 METODOLOGIA

A metodologia deste trabalho seguiu cinco etapas principais: compilação do banco de dados, pré-processamento, treinamento dos modelos, avaliação de desempenho e comparação com outros modelos. O banco de dados foi composto por 1.599 registros experimentais, provenientes de três estudos: Ibrahim et al. (2023), Yeh (1998) e Wongkeo et al. (2014), com 425, 1.030 e 144 registros, respectivamente. Os dados consideram concreto variados, informando teor de CP, sílica ativa, cinza volante, escória de alto-forno (EAF), agregados fino e graúdo, água, superplastificante, idade e RC. Os estudos não trazem informações como a classe da cinza volante e tipo de sílica, nem informações acerca da reatividade dos MCS. Yeh (1998) e Wongkeo et al. (2014) usaram areia de rio e calcário britado ($D_{max}$ 20 mm) como agregados.

Os dados foram divididos em 85% para treinamento e 15% para teste, conforme a proposta de Joseph (2022). Para treinar e avaliar, foram aplicados sete modelos de regressão: *Adaptive Boosting* (ADA), *Artificial Neural Network* (ANN), *Histogram-based Gradient Boosting Machine* (HGBM), *Light Gradient Boosting Machine* (LGBM), *Random Forest* (RF), *Support Vector Regression* (SVR) e *eXtreme Gradient Boosting Machine* (XGBM). O desempenho dos modelos foi avaliado por métricas como coeficiente de determinação ($R^2$), erro absoluto médio (MAE) e erro percentual absoluto médio (MAPE). Foi analisada também a importância das variáveis por permutação para avaliar as variáveis que mais influenciam os modelos.

# 3 RESULTADOS E ANÁLISES

## 3.1 BASE DE DADOS

A base de dados inicial consiste em 1.599 amostras. Após o pré-processamento, 31 duplicatas foram removidas. A base de dados final contém 1.568 amostras, 9 variáveis de entrada e 1 variável de saída (RC). O resumo estatístico das variáveis da base de dados apresentados na Tabela 1 mostra que o CP tem uma média de 274,88 kg/m³ e desvio padrão de 107,74 kg/m³, indicando grande variação. As variáveis sílica ativa, cinza volante e EAF têm médias de 2,40, 72,37 e 69,07 kg/m³, respectivamente, com desvios padrão elevados, refletindo variações significativas no consumo de materiais. A idade do concreto varia de 1 a 365 dias, com média de 40 dias, e a resistência à compressão média é de 37,40 MPa, com desvio padrão de 17,49 MPa, mostrando também grande variabilidade.

## 3.2 AVALIAÇÃO DA PERFORMANCE DOS MODELOS

Os modelos de ML para previsão da RC do concreto apresentaram diferenças significativas em seus desempenhos (Tabela 2). Um dos modelos que se destacou foi o ADA, alcançando $R^2$ de 0,9986 no conjunto de treinamento e 0,9547 no conjunto de teste, evidenciando uma excelente capacidade preditiva. Em contraste, modelos como ANN e SVR tiveram desempenhos inferiores, com $R^2$ abaixo de 0,7 no conjunto de teste, que podem ser melhorados com mais ajustes nos hiperparâmetros.

**Tabela 1:** Resumo estatístico das variáveis.

| Variável | Unidade | Média | Desvio padrão | Mínimo | Primeiro quartil | Segundo quartil | Terceiro quartil | Máximo |
|---|---|---|---|---|---|---|---|---|
| CP | (kg/m³) | 274.88 | 107.74 | 102.00 | 180.00 | 255.40 | 337.90 | 600.00 |
| Silica ativa | (kg/m³) | 2.40 | 9.99 | 0.00 | 0.00 | 0.00 | 0.00 | 60.00 |
| Cinza Volante | (kg/m³) | 72.37 | 86.85 | 0.00 | 0.00 | 24.48 | 124.16 | 420.00 |
| EAF | (kg/m³) | 69.07 | 85.65 | 0.00 | 0.00 | 0.02 | 141.30 | 359.40 |
| Ag. miúdo | (kg/m³) | 793.57 | 107.27 | 594.00 | 733.75 | 781.50 | 845.00 | 1166.00 |
| Ag. graúdo | (kg/m³) | 936.87 | 126.75 | 595.00 | 882.00 | 959.00 | 1022.80 | 1145.00 |
| Água | (kg/m³) | 182.22 | 19.80 | 121.75 | 170.17 | 181.95 | 192.00 | 247.20 |
| Superplastificante | (kg/m³) | 5.78 | 5.74 | 0.00 | 0.00 | 6.00 | 9.88 | 32.20 |
| Idade | (dias) | 40 | 53 | 1 | 28 | 28 | 28 | 365 |
| RC | (MPa) | 37.40 | 17.49 | 2.33 | 24.89 | 35.34 | 47.15 | 106.60 |

**Fonte**: Autores (2025)

O MAE e MAPE reforçam tais observações, com o modelo ADA apresentando um MAE de 12,6719 MPa no conjunto de teste, capturando a tendência dos dados. Modelos baseados em árvores, como HGBM, LGBM e RF, também apresentam ótimos desempenhos, destacando-se o LGBM com $R^2$ de 0,9773 no conjunto de teste. A capacidade desses modelos de lidar com dados heterogêneos contribui para tais desempenhos.

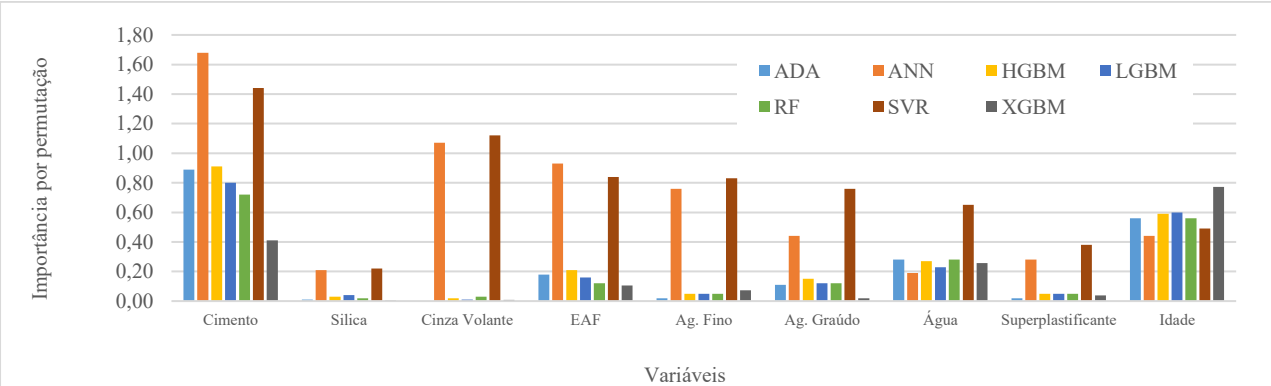
**Tabela 2:** Métricas dos modelos nos conjuntos de treinamento e teste

| Modelos | Conjunto de treinamento | | | Conjunto de teste | | |
|---|---|---|---|---|---|---|
| | $R^2$ | MAE (MPa) | MAPE (%) | $R^2$ | MAE (MPa) | MAPE (%) |
| ADA | 0.9986 | 0.4399 | 0.0405 | 0.9547 | 12.6719 | 1.9519 |
| ANN | 0.7150 | 8.0684 | 0.2015 | 0.6963 | 7.7012 | 0.2069 |
| HGBM | 0.9959 | 0.6542 | 0.0230 | 0.9724 | 1.8322 | 0.0671 |
| LGBM | 0.9973 | 0.4368 | 0.0144 | 0.9773 | 1.5541 | 0.0550 |
| RF | 0.9928 | 0.9414 | 0.0321 | 0.9624 | 2.2551 | 0.0850 |
| SVR | 0.6803 | 8.1658 | 0.2067 | 0.6819 | 7.5037 | 0.1957 |
| XGBM | 0.9966 | 0.5790 | 0.0192 | 0.9758 | 1.6407 | 0.0576 |

**Fonte**: Autores (2025)

A análise da importância das variáveis (Figura 1) revelou que a quantidade de cimento é a variável mais influente na maioria dos modelos, especialmente no ADA, ANN, HGBM e LGBM. Modelos como RF e XGBM atribuem menor importância ao cimento, conferindo mais importância às outras variáveis. A variável idade também tem grande impacto em modelos como HGBM, LGBM, RF e XGBM. Estes resultados já eram esperados, uma vez que é a influência do teor de cimento e a idade na resistência à compressão é notoriamente conhecida. Por outro lado, as quantidades de sílica, cinza volante e escória tiveram baixa importância, especialmente no ADA, HGBM, LGBM e RF. Provavelmente as dosagens presentes no banco de dados não possuíram variabilidade o suficiente para que os modelos acessassem a conhecida influência dos MCS na resistência à compressão.

**Figura 1:** Importância das variáveis por permutação




**Fonte**: Autores (2025)

## 4 CONCLUSÃO

Este estudo avaliou os desempenhos de modelos de ML em uma base de dados experimental, com foco em sua capacidade preditiva frente a variações nos materiais e na identificação das variáveis mais relevantes. Os modelos ADA, LGBM, XGBM e RF apresentaram ótimos desempenhos. Por outro lado, os modelos ANN e SVR revelaram desempenhos insatisfatórios, com $R^2$ abaixo de 0,7. Aumentar a complexidade ou ajustar

outros hiperparâmetros de tais modelos podem melhorar estes desempenhos. As variáveis mais impactantes na modelagem foram a quantidade de cimento, a idade e o teor de água. Para estudos futuros, recomenda-se a criação de novos conjuntos de dados e modelos que considerem variações na composição dos materiais, reatividade dos SCMs e controle granulométrico, além de outros fatores externos.

## 5 AGRADECIMENTOS

Os autores agradecem à CAPES (processo nº 8887.827958/2023-00), ao CNPq (processo nº 132090/2023-6 e 306536/2022-6), à PUC-Rio e à FAPERJ (processo nº E-26/201.266/2022) pelo apoio financeiro.

## REFERÊNCIAS

AHMAD, Waqas; AHMAD, Ayaz; OSTROWSKI, Krzysztof Adam; ASLAM, Fahid; JOYKLAD, Panuwat; ZAJDEL, Paulina. Application of advanced machine learning approaches to predict the compressive strength of concrete containing supplementary cementitious materials. **Materials**, v. 14, n. 19, p. 5762, 2021. doi:10.3390/ma14195762.

AL-SHAMIRI, Abobakr Khalil; YUAN, Tian-Feng; KIM, Joong Hoon. Non-tuned machine learning approach for predicting the compressive strength of high-performance concrete. **Materials**, v. 13, n. 5, p. 1023, 2020. doi:10.3390/ma13051023.

CHEN, Ning; ZHAO, Shibo; GAO, Zhiwei; WANG, Dawei; LIU, Pengfei; OESER, Markus; HOU, Yue; WANG, Linbing. Virtual Mix Design: Prediction of Compressive Strength of Concrete with Industrial Wastes Using Deep Data Augmentation. **Construction and Building Materials**, v. 323, p. 126580, 2022. doi: 10.1016/j.conbuildmat.2022.126580.

CUI, Liyun; CHEN, Peiyuan; WANG, Liang; LI, Jin; LING, Hao. Application of extreme gradient boosting based on grey relation analysis for prediction of compressive strength of concrete. **Advances in Civil Engineering**, v. 2021, p. 1-14, 2021. doi:10.1155/2021/8878396.

IBRAHIM, Syed Muhammad; ANSARI, Saad Shamim; HASAN, Syed Danish. Towards white box modeling of compressive strength of sustainable ternary cement concrete using explainable artificial intelligence (XAI). **Applied Soft Computing**, v. 149, p. 110997, 2023. doi:10.1016/j.asoc.2023.110997.

JOSEPH, V. Roshan. Optimal ratio for data splitting. Statistical Analysis and Data Mining: **The ASA Data Science Journal,** v. 15, n. 4, p. 531-538, 2022. doi:10.1002/sam.11583.

MEHTA, P.; MONTEIRO, Paulo. **Concreto: Microestrutura, propriedades e materiais**. 2. ed. São Paulo: IBRACON, 2014. 504 p.

RADWAN, Mohammed K. H.; ONN, Chiu Chuen; MO, Kim Hung; YAP, Soon Poh; NG, Chee Guan; YUSOFF, Sumiani. Eco-mechanical performance of binary and ternary cement blends containing fly ash and slag. **Proceedings of the Institution of Civil Engineers – Engineering Sustainability**, v. 174, n. 1, p. 23-36, 2021. doi:10.1680/jensu.20.00009.

TUTIKIAN, B. F.; HELENE, P. R. L. **Dosagem dos concretos de cimento Portland. Concreto: ciência e tecnologia**. Tradução. São Paulo: IBRACON, 2022. Disponível em: https://repositorio.usp.br/directbitstream/1a05cae2-8454-49ce-aebe-c334e88b8c20/Helene-2022-dosagem.pdf. Acesso em: 20 set. 2024.

YEH, I. C. Modeling of strength of high-performance concrete using artificial neural networks. **Cement and Concrete**, v. 28, n. 12, p. 1797-1808, 1998. doi:10.1016/S0008-8846(98)00165-3.

WONGKEO, Watcharapong; THONGSANITGARN, Pailyn; NGAMJARUROJANA, Athipong; CHAIPANICH, Arnon. Compressive strength and chloride resistance of self-compacting concrete containing high level fly ash and silica fume. **Materials & Design**, v. 64, p. 261-269, 2014. doi:10.1016/j.matdes.2014.07.042