

Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.

P. Kumar Yadav^{1*} , A. Kumar² 

* Autor de Contato: pin2kryadav@gmail.com, ajaykumar8mail@gmail.com

DOI: <https://doi.org/10.21041/ra.v16i3.1024>

Recebido: 20/02/2026 | Correções recebidas: 03/05/2026 | Aceito: 08/05/2026 | Publicado: 01/09/2026

RESUMO

Este estudo tem como objetivo desenvolver uma estrutura de aprendizado de máquina explicável para prever a resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados (RA-SCC), utilizando indicadores de empacotamento de partículas. Um conjunto de dados composto por 365 misturas provenientes de 102 estudos foi analisado utilizando seis modelos de regressão, com aplicação do método SHapley Additive exPlanations (SHAP) para interpretação dos resultados. O modelo Extra Trees apresentou a maior precisão ($R^2 = 0,9724$; Raiz do Erro Quadrático Médio (RMSE) = 1,19 MPa). O conjunto de dados está limitado à resistência aos 28 dias e não considera os efeitos dos aditivos. O estudo apresenta uma abordagem combinada que integra conceitos de empacotamento de partículas, avaliação de multi-colinearidade e inteligência artificial explicável. Os resultados demonstram que o aprendizado de máquina pode prever a resistência de forma confiável, ao mesmo tempo em que captura implicitamente os mecanismos relacionados ao empacotamento de partículas.

Palavras-chave: concreto autoadensável com agregados reciclados; aprendizado de máquina; predição da resistência à compressão; indicadores de empacotamento de partículas; inteligência artificial explicável; análise SHAP; multicolinearidade; concreto sustentável.

Citar como: Kumar Yadav, P., Kumar, A. (2026), “*Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.*”, Revista ALCONPAT, 16 (3), pp. 411 – 440, DOI: <https://doi.org/10.21041/ra.v16i3.1024>

¹ Research Scholar, Department of Civil Engineering, Sandip University, Sijoul, Madhubani-847235, Bihar, India

² Assistant Professor, Department of Civil Engineering, Sandip University, Sijoul, Madhubani-847235, Bihar, India.

Contribuição de cada autor

Neste trabalho, a contribuição dos autores foi dividida igualmente entre Pintu Kumar Yadav (50%) e Dr. Ajay Kumar (50%).

Licença Creative Commons

Copyright (2026) é propriedade dos autores. Este trabalho é um artigo de acesso aberto publicado sob os termos e condições de uma Licença Internacional Creative Commons Atribuição 4.0 ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Discussões e correções pós-publicação

Qualquer discussão, incluindo as respostas dos autores, será publicada na segunda edição de 2027, desde que as informações sejam recebidas antes do encerramento da primeira edição de 2027.

Explainable machine learning based prediction of 28-day compressive strength of recycled aggregate self-compacting concrete using particle packing indicators.

ABSTRACT

This study aims to develop an explainable machine learning framework for predicting the 28-day compressive strength of recycled aggregate self-compacting concrete (RA-SCC) using particle packing indicators. A dataset comprising 365 mixtures from 102 studies was analyzed using six regression models, with SHapley Additive exPlanations (SHAP) applied for interpretability. The Extra Trees model achieved the highest accuracy ($R^2 = 0.9724$, Root Mean Square Error (RMSE) = 1.19 MPa). The dataset is limited to 28-day strength and excludes admixture effects. The study introduces a combined approach integrating particle packing concepts, multicollinearity assessment, and explainable AI. The results demonstrate that machine learning can reliably predict strength while implicitly capturing packing-related mechanisms.

Keywords: recycled aggregate self-compacting concrete; machine learning; compressive strength prediction; particle packing indicators; explainable artificial intelligence; shap analysis; multicollinearity; sustainable concrete.

Predicción explicable basada en aprendizaje automático de la resistencia a la compresión a 28 días de hormigón autocompactante de áridos reciclados utilizando indicadores de empaquetado de partículas.

RESUMEN

Este estudio tiene como objetivo desarrollar un marco entendible de aprendizaje automático para predecir la resistencia a compresión a 28 días del hormigón autocompactante de agregados reciclados (RA-SCC) utilizando indicadores de empaquetado de partículas. Se analizó un conjunto de datos que comprendía 365 mezclas de 102 estudios utilizando seis modelos de regresión, aplicando SHapley Additives ExPlanations (SHAP) para su interpretabilidad. El modelo Extra Trees alcanzó la mayor precisión ($R^2 = 0,9724$, Root Mean Square Error (RMSE) = 1,19 MPa). El conjunto de datos está limitado a una resistencia a la compresión de 28 días y excluye efectos de mezclado. El estudio introduce un enfoque combinado que integra conceptos de empaquetado de partículas, evaluación de multicolinealidad e IA explicable. Los resultados demuestran que el aprendizaje automático puede predecir de forma fiable resistencia a la compresión mientras captura implícitamente los mecanismos relacionados con el empaquetado.

Palabras clave: concreto autocompactante de áridos reciclados; aprendizaje automático; predicción de resistencia a la compresión; indicadores de empaquetamiento de partículas; inteligencia artificial explicable; análisis de shap; multicolinealidad; hormigón sostenible.

Informações legais

Revista ALCONPAT é uma publicação trimestral da Associação Latino-Americana de Controle de Qualidade, Patologia e Recuperação de Construção, Internacional, A.C., Km. 6, antiga estrada para Progreso, Merida, Yucatán, C.P. 97310, Tel. +52 1 983 419 8241, alconpat.int@gmail.com, Website: www.alconpat.org

Reserva de direitos de uso exclusivo No.04-2013-011717330300-203, eISSN 2007-6835, ambos concedidos pelo Instituto Nacional de Direitos Autorais. Editor responsável: Dr. Pedro Castro Borges. Responsável pela última atualização deste número, Unidade de Informática ALCONPAT, Eng. Elizabeth Sabido Maldonado.

As opiniões expressas pelos autores não refletem necessariamente a posição do editor.

A reprodução total ou parcial do conteúdo e das imagens da publicação é realizada de acordo com o código COPE e a licença CC BY 4.0 da Revista ALCONPAT.

1. INTRODUÇÃO

A pesquisa sobre concreto com agregados reciclados (RAC) como potencial substituto do concreto convencional tem aumentado em razão da crescente demanda por materiais de construção ambientalmente sustentáveis. A utilização de agregados reciclados reduz o consumo de recursos naturais e a geração de resíduos provenientes da construção e demolição, que representam uma importante carga ambiental em escala global (Behera et al., 2014; Evangelista & de Brito, 2007; Silva et al., 2015). Devido à sua excelente fluidez, capacidade de passagem e resistência à segregação, o concreto autoadensável (SCC) tornou-se um material de construção avançado, capaz de ser aplicado de forma eficiente em estruturas densamente armadas, sem a necessidade de vibração mecânica (Khayat, 2016; Okamura & Ouchi, 2003). A incorporação de agregados reciclados ao SCC oferece benefícios significativos em termos de sustentabilidade; entretanto, também introduz uma variabilidade acentuada nas propriedades no estado fresco e endurecido, particularmente na resistência à compressão (Kou & Poon, 2009; Poon et al., 2004).

A resistência à compressão continua sendo um dos principais parâmetros de projeto e controle de qualidade na engenharia do concreto, influenciando diretamente a segurança estrutural e o desempenho em serviço. Uma interação complexa entre as propriedades dos materiais, como a taxa de substituição por agregados reciclados, a composição do aglomerante, a relação água/aglomerante, a distribuição granulométrica dos agregados e a dosagem de aditivos, controla o desenvolvimento da resistência do concreto autoadensável com agregados reciclados (RA-SCC) (Guo et al., 2018; Tam et al., 2007; Xiao et al., 2012). A presença de argamassa aderida aos agregados reciclados altera a morfologia das partículas, a absorção de água e o comportamento da zona de transição interfacial, afetando, conseqüentemente, a eficiência do empacotamento das partículas e a estrutura interna dos poros (Butler et al., 2013; Etxeberria et al., 2007). Esses efeitos combinados tornam particularmente desafiadora a predição confiável da resistência por meio de formulações empíricas convencionais.

A teoria do empacotamento de partículas fornece uma base mecanicista para a compreensão da evolução da resistência em materiais cimentícios. O empacotamento denso dos constituintes sólidos minimiza os vazios, aumenta o contato entre as partículas e melhora a transferência de tensões no interior da matriz endurecida (de Larrard, 1999; Funk & Dinger, 1994). De acordo com pesquisas experimentais, o empacotamento otimizado dos agregados pode aumentar significativamente a durabilidade, a resistência à compressão e o módulo de elasticidade, ao mesmo tempo em que reduz a demanda de aglomerante (Cheng et al., 2021; He et al., 2022; Su et al., 2001). Nos sistemas de concreto autoadensável (SCC), a densidade de empacotamento é especialmente importante, pois a trabalhabilidade e a resistência à segregação estão intimamente relacionadas ao equilíbrio volumétrico entre agregados, materiais finos e pasta (Felekoğlu et al., 2007; Khayat, 2016). Apesar desses avanços, a maioria das abordagens de empacotamento de partículas depende de procedimentos experimentais trabalhosos e apresenta dificuldades de generalização para diferentes fontes de materiais e diferentes dosagens de concreto.

Simultaneamente, o uso de técnicas de aprendizado de máquina (ML) para modelar interações complexas e não lineares em materiais à base de concreto tem se tornado cada vez mais difundido (Tang et al., 2023). Redes neurais artificiais, regressão por vetores de suporte, Random Forest e algoritmos de Gradient Boosting têm sido utilizados com sucesso em diversos estudos para prever a resistência à compressão com maior precisão do que os modelos convencionais de regressão (Asteris et al., 2020; Liu et al., 2021; Yeh, 1998). A eficácia das técnicas de aprendizado por conjunto (ensemble learning) no tratamento da variabilidade dos materiais e das interações não lineares foi validada por estudos recentes voltados ao concreto com agregados reciclados (RAC) (Li et al., 2022; Park et al., 2022). Entretanto, muitos modelos

de aprendizado de máquina são considerados “caixas-pretas”, o que limita sua interpretabilidade e reduz a confiança em sua utilização nos processos de tomada de decisão em engenharia (Bardhan et al., 2021; Zhou et al., 2024).

Métodos de inteligência artificial explicável, incluindo o SHapley Additive exPlanations (SHAP), foram desenvolvidos para quantificar o impacto das variáveis individuais de entrada nas previsões dos modelos de aprendizado de máquina, com o objetivo de superar essa limitação (Lundberg & Lee, 2017). A predição da resistência do concreto tem se beneficiado do uso eficaz da interpretação baseada em SHAP, que fornece informações fisicamente significativas sobre o papel dos constituintes da mistura, como cimento, teor de água e materiais cimentícios suplementares (Almutairi, 2025; Fanijo et al., 2023). Embora a teoria do empacotamento de partículas tenha sido amplamente investigada no contexto da dosagem de concretos, sua integração com abordagens de aprendizado de máquina explicável aplicadas ao concreto autoadensável com agregados reciclados ainda permanece relativamente limitada.

Além disso, as variáveis de dosagem do concreto associadas ao empacotamento de partículas são frequentemente fortemente interdependentes. Parâmetros como as frações de agregados, o teor de finos e a relação água/aglomerante apresentam multicolinearidade significativa, o que compromete a confiabilidade dos modelos estatísticos lineares e dificulta a interpretação física dos resultados (Kim & Lee, 2023; Li et al., 2022; J. Zhang et al., 2019).

Embora a multi-colinearidade tenha implicações significativas para a seleção dos modelos e sua interpretabilidade, ela raramente é quantificada de forma explícita em estudos sobre concreto baseados em aprendizado de máquina (ML). Entretanto, as pesquisas atualmente publicadas ainda não apresentam uma abordagem abrangente para a predição da resistência do concreto autoadensável com agregados reciclados (RA-SCC) que considere, simultaneamente, a precisão da predição, a multicolinearidade e a interpretabilidade física.

Diante disso, o presente estudo incorpora indicadores de empacotamento de partículas derivados das características de composição da mistura em uma estrutura baseada em aprendizado de máquina, destinada à predição da resistência à compressão do concreto autoadensável com agregados reciclados (RA-SCC). Diferentemente de estudos que calculam explicitamente a densidade de empacotamento ou o índice de vazios, este trabalho captura implicitamente os mecanismos relacionados ao empacotamento por meio de modelagem orientada por dados e interpreta sua influência utilizando a análise SHAP. Além disso, é empregada a análise do fator de inflação da variância (VIF) para quantificar a multicolinearidade entre as variáveis relacionadas ao empacotamento, fornecendo uma base estatisticamente robusta para a seleção dos modelos. Ao integrar precisão preditiva, validação estatística e interpretabilidade física, este estudo busca contribuir com uma ferramenta confiável e explicável para a dosagem sustentável de RA-SCC. Reconhece-se que o comportamento do empacotamento de partículas no concreto é regido por múltiplas variáveis de dosagem que interagem entre si e pode não ser plenamente representado apenas por índices individuais de empacotamento. Diversos estudos anteriores investigaram o efeito da taxa de substituição por agregados reciclados sobre a resistência à compressão, estabelecendo-a como um parâmetro fundamental para o projeto de concretos sustentáveis (Kou & Poon, 2009; Pedro et al., 2017; Xiao et al., 2012). Esses estudos destacam, de forma consistente, a influência do nível de substituição na redução e na variabilidade da resistência, fornecendo uma base para sua inclusão como uma variável primária de entrada em modelos preditivos.

2. OBJETIVOS DA PESQUISA

O principal objetivo deste estudo é desenvolver uma estrutura baseada em aprendizado de máquina capaz de prever, com precisão e interpretabilidade, a resistência à compressão do concreto autoadensável com agregados reciclados (RA-SCC), por meio da incorporação de

indicadores de empacotamento de partículas derivados de características relacionadas à composição da mistura. Para atingir esse objetivo, são definidos os seguintes objetivos específicos:

1. Compilar e analisar um conjunto de dados de RA-SCC baseado na literatura, incorporando variáveis de dosagem associadas ao comportamento do empacotamento de partículas, incluindo as proporções dos agregados, o teor de aglomerante, a relação água/aglomerante, os índices volumétricos e o nível de substituição por agregados reciclados.
2. Criar e avaliar diversos modelos de regressão baseados em aprendizado de máquina para a predição da resistência à compressão do RA-SCC, incluindo algoritmos de conjunto (*ensemble*), baseados em árvores e lineares.
3. Avaliar o desempenho dos modelos utilizando indicadores estatísticos abrangentes, incluindo o coeficiente de determinação, métricas baseadas em erro, erro médio de viés e testes de hipóteses, a fim de avaliar a precisão preditiva e a confiabilidade estatística.
4. Utilizar a análise do fator de inflação da variância (VIF) para quantificar o grau de multicolinearidade entre as variáveis de entrada relacionadas ao empacotamento de partículas e investigar suas implicações para a estabilidade das predições e a seleção dos modelos.
5. Relacionar os resultados do aprendizado de máquina ao comportamento físico dos materiais, interpretando o impacto dos indicadores de empacotamento de partículas na predição da resistência à compressão por meio do método SHapley Additive exPlanations.
6. Demonstrar a capacidade do aprendizado de máquina explicável de capturar implicitamente os mecanismos de empacotamento de partículas, sem o cálculo explícito da densidade de empacotamento ou do índice de vazios, contribuindo para uma dosagem sustentável e orientada por dados do RA-SCC.

3. MATERIAIS E DESCRIÇÃO DO CONJUNTO DE DADOS

3.1 Fonte dos dados e estratégia de compilação.

Os resultados experimentais publicados na literatura científica revisada por pares sobre concreto com agregados reciclados e concreto autoadensável com agregados reciclados foram sistematicamente compilados para constituir o conjunto de dados utilizado nesta pesquisa. Foram considerados apenas estudos que apresentavam proporções de mistura completas e claramente definidas, características dos materiais e valores de resistência à compressão. Essa abordagem garantiu a confiabilidade dos dados, ao mesmo tempo em que permitiu abranger uma ampla variedade de combinações de materiais e níveis de desempenho. O conjunto de dados final é composto por $N = 365$ misturas de concreto, sistematicamente compiladas a partir de 102 estudos científicos independentes revisados por pares. Cada estudo selecionado forneceu informações completas e consistentes sobre as proporções das misturas e a resistência à compressão, garantindo a qualidade e a comparabilidade dos dados entre diferentes programas experimentais. O conjunto de dados abrange uma ampla faixa de níveis de substituição por agregados reciclados, composições de aglomerantes e configurações de mistura, representando, assim, uma variabilidade realista dos materiais. Essa diversidade aumenta a representatividade estatística e a robustez da estrutura de aprendizado de máquina desenvolvida neste estudo. O conjunto de dados é composto exclusivamente por valores de resistência à compressão obtidos experimentalmente e relatados nos estudos de origem. Não foram utilizados dados sintéticos ou simulados. Além da divisão dos dados em conjuntos de treinamento e teste adotada neste estudo, reconhece-se que a utilização de conjuntos de dados experimentais externos para validação independente constitui uma etapa importante para pesquisas futuras, permitindo

avaliar de forma mais abrangente a capacidade de generalização do modelo para sistemas de materiais ainda não analisados.

Reconhece-se que os modelos de aprendizado de máquina desenvolvidos com dados provenientes de um único programa experimental bem controlado podem apresentar maior consistência interna. Entretanto, a utilização, neste estudo, de um conjunto de dados obtido a partir da literatura permite incorporar uma variedade mais ampla de propriedades dos materiais, proporções das misturas e condições experimentais. Essa diversidade aumenta a capacidade de generalização do modelo desenvolvido e permite que ele capture a variabilidade real encontrada em aplicações práticas. Considerando a complexidade inerente e a interdependência das variáveis do concreto autoadensável com agregados reciclados, a utilização de múltiplas fontes independentes de dados proporciona uma representação mais abrangente do comportamento do material. A abordagem adotada, portanto, busca estabelecer uma estrutura preditiva generalizada, em vez de um modelo específico para um determinado projeto, calibrado a partir de um único conjunto de dados experimentais. Os estudos selecionados abrangem diferentes fontes de agregados reciclados, taxas de substituição, sistemas de aglomerantes e condições de cura, proporcionando variabilidade suficiente para uma análise robusta por meio de aprendizado de máquina (Ajdukiewicz & Kliszczewicz, 2002; Limbachiya et al., 2000; Xiao et al., 2012).

As principais fontes incluem estudos experimentais sobre o comportamento mecânico do concreto com agregados reciclados, a influência do nível de substituição dos agregados e o papel de materiais cimentícios suplementares, como as cinzas volantes. Em conjunto, esses estudos representam tanto sistemas de concreto convencional quanto de concreto de alto desempenho e refletem a variabilidade real dos materiais encontrada na prática (Corinaldesi, 2010; Kou & Poon, 2013; Pedro et al., 2017). Também foram utilizadas revisões abrangentes e monografias sobre concreto com agregados reciclados, a fim de garantir a consistência das definições dos materiais e do escopo experimental (de Brito & Saikia, 2013; Oikonomou, 2005). Uma lista detalhada das referências utilizadas para a compilação do conjunto de dados é apresentada na seção Referências. Em conjunto, essas fontes abrangem uma ampla variedade de teores de agregados reciclados, sistemas de aglomerantes e condições de ensaio, garantindo que o conjunto de dados represente de forma adequada a variabilidade experimental real relatada na literatura.

3.2 Materiais constituintes.

3.2.1 Agregados reciclados.

Os agregados reciclados utilizados nos estudos compilados foram obtidos principalmente a partir de resíduos de concreto britado provenientes de elementos estruturais demolidos. Esses agregados são geralmente constituídos por partículas de agregados naturais recobertas por argamassa de cimento aderida, o que altera a textura superficial, a porosidade e as características de absorção de água em comparação com os agregados naturais (Etxeberria et al., 2007; Poon et al., 2004). Diversos estudos relataram a utilização de métodos de pré-tratamento ou beneficiamento, como processamento mecânico ou técnicas assistidas por micro-ondas, com o objetivo de melhorar a qualidade dos agregados reciclados. Entretanto, os agregados não tratados foram utilizados com maior frequência, a fim de representar condições práticas de aplicação na construção civil (Akbarnezhad et al., 2011; Tam et al., 2005). A taxa de substituição por agregados reciclados variou entre os estudos, abrangendo desde a substituição parcial até a substituição total dos agregados graúdos naturais. Essa variação é essencial para avaliar a influência do teor de agregados reciclados sobre a eficiência do empacotamento de partículas e o desenvolvimento da resistência à compressão (Evangelista & de Brito, 2007; Pedro et al., 2017). Esses efeitos estão amplamente documentados em estudos anteriores que analisaram a influência dos agregados reciclados sobre o desempenho mecânico

e as características microestruturais do concreto (Butler et al., 2013; Etxeberria et al., 2007; Poon et al., 2004).

3.2.2 Materiais cimentícios.

O cimento Portland comum (OPC) foi o principal material aglomerante utilizado nos estudos experimentais empregados para a compilação do conjunto de dados. No conjunto de dados compilado, o termo OPC refere-se aos aglomerantes cimentícios relatados nos estudos originais, que podem incluir tanto cimento Portland puro (CEM I) quanto cimentos compostos (como CEM II), dependendo da fonte. Devido às variações nas práticas de apresentação das informações encontradas na literatura, nem sempre foi possível distinguir explicitamente esses tipos de cimento em todas as entradas do conjunto de dados. O OPC desempenha um papel fundamental no processo de hidratação, na formação da pasta e no desenvolvimento da resistência nas primeiras idades do RA-SCC. As variações no teor de cimento influenciam diretamente o volume de pasta, o grau de recobrimento das partículas e a capacidade de transferência de esforços na matriz endurecida, fatores que estão diretamente relacionados ao desempenho da resistência à compressão.

Diversos estudos incorporaram materiais cimentícios suplementares, especialmente cinzas volantes, como substituição parcial do cimento. A incorporação de cinzas volantes tem sido amplamente relatada como capaz de melhorar o desempenho mecânico em longo prazo e a durabilidade do concreto com agregados reciclados, por meio do refinamento da estrutura de poros e da melhoria da zona de transição interfacial entre as partículas de agregado e a pasta de cimento (Behera et al., 2014; Kou & Poon, 2013). A morfologia esférica e a granulometria fina das partículas de cinza volante contribuem para melhorar o empacotamento de partículas, preenchendo os microvazios existentes entre as partículas mais grossas, aumentando, assim, a densidade de empacotamento e reduzindo a porosidade capilar.

Sob a perspectiva do empacotamento de partículas, o cimento e a cinza volante controlam, conjuntamente, a disponibilidade de pasta e a distribuição dos finos no sistema de concreto. Um volume adequado de pasta é essencial no concreto autoadensável para garantir fluidez adequada, resistência à segregação e completo envolvimento das partículas de agregado.

No concreto autoadensável com agregados reciclados (RA-SCC), essa exigência torna-se ainda mais crítica devido à maior rugosidade superficial e à maior absorção de água dos agregados reciclados. A cinza volante contribui para a melhoria da trabalhabilidade e para a redução da demanda de água, influenciando indiretamente a eficiência do empacotamento e a estrutura interna de vazios (de Brito & Saikia, 2013). Esses resultados são consistentes com a literatura consolidada sobre o papel dos materiais cimentícios suplementares na melhoria da durabilidade e do refinamento microestrutural dos sistemas de concreto com agregados reciclados (Behera et al., 2014; Kou & Poon, 2013; Silva et al., 2015).

Deve-se observar que, em diversos estudos utilizados como fonte, os sistemas de cimentos compostos (por exemplo, CEM II) podem já conter materiais cimentícios suplementares, como cinza volante. Quando esses cimentos são relatados juntamente com teores adicionais de cinza volante, a contribuição efetiva desse material pode ser superestimada. Isso introduz certo grau de viés inerente ao conjunto de dados, resultando em valores aparentes elevados de cinza volante e em uma forte interdependência estatística entre as variáveis relacionadas aos aglomerantes. Esse efeito contribui diretamente para o VIF extremamente elevado observado para o teor de cinza volante, refletindo uma multi-colinearidade severa entre os componentes cimentícios. Entretanto, essa interdependência é característica de conjuntos de dados de concreto obtidos a partir da literatura, nos quais a composição dos aglomerantes é determinada por restrições práticas de dosagem, e não por variáveis controladas de forma independente.

A utilização, neste estudo, de modelos de aprendizado de máquina por conjunto (*ensemble*) baseados em árvores reduz essa limitação, uma vez que esses algoritmos são intrinsecamente

robustos à multicolinearidade e não dependem da hipótese de independência entre as variáveis preditoras. No conjunto de dados compilado, os teores de cimento e cinza volante apresentaram forte interdependência com outros parâmetros de dosagem, como a relação água/aglomerante e as frações de agregados. Essa interação evidencia o papel predominante desses componentes na definição do arranjo microestrutural das partículas sólidas e dos produtos de hidratação. Consequentemente, os teores de cimento e cinza volante foram considerados indicadores-chave do empacotamento de partículas na estrutura de aprendizado de máquina, permitindo que os modelos capturassem implicitamente os mecanismos relacionados ao empacotamento que influenciam o desenvolvimento da resistência à compressão.

3.2.3 Agregados e água de amassamento.

Tanto os agregados graúdos quanto os agregados miúdos foram incluídos no conjunto de dados, sendo o agregado miúdo constituído, geralmente, por areia natural de rio. A distribuição granulométrica e as proporções volumétricas dos agregados foram explicitamente relatadas na maioria dos estudos, permitindo a determinação das frações de agregados e dos parâmetros de fração volumétrica dos agregados (Domingo-Cabo et al., 2009; Limbachiya et al., 2000). O teor de água de amassamento e a relação água/aglomerante foram extraídos diretamente dos estudos, uma vez que esses parâmetros exercem forte influência sobre a trabalhabilidade, a estrutura de vazios e a resistência à compressão (Poon et al., 2004; Tam et al., 2005). No concreto autoadensável, os aditivos químicos, como superplastificantes e agentes modificadores de viscosidade, desempenham um papel fundamental na obtenção da fluidez, capacidade de passagem e resistência à segregação requeridas. Entretanto, esses parâmetros não foram relatados de maneira consistente nas fontes bibliográficas compiladas e, portanto, não foram incluídos como variáveis de entrada na estrutura de aprendizado de máquina utilizada no presente estudo. Essa omissão representa uma limitação, uma vez que a dosagem de aditivos influencia significativamente tanto as propriedades no estado fresco quanto no estado endurecido do SCC.

Nos sistemas de concreto autoadensável, as propriedades no estado fresco, como fluidez, capacidade de passagem e resistência à segregação, constituem critérios essenciais de desempenho. Essas características são normalmente avaliadas por meio de ensaios padronizados, incluindo espalhamento (*slump flow*), funil V (*V-funnel*) e caixa L (*L-box*). Variações na composição da mistura, especialmente nas proporções dos agregados, no teor de aglomerante e na relação água/aglomerante, influenciam significativamente essas propriedades. Entretanto, devido à ausência de dados sobre as propriedades no estado fresco relatados de maneira consistente nas fontes bibliográficas compiladas, o presente estudo concentra-se exclusivamente na predição da resistência à compressão. Além disso, parâmetros fundamentais que diferenciam o concreto autoadensável do concreto convencional, como a dosagem de aditivos redutores de água de alta eficiência, o teor de agentes modificadores de viscosidade e as características reológicas relacionadas ao teor de vazios, não estavam disponíveis de maneira consistente no conjunto de dados compilado. Consequentemente, o modelo não diferencia explicitamente o comportamento do SCC com base no desempenho no estado fresco, mas captura seus efeitos de forma indireta por meio das variáveis relacionadas à composição da mistura.

3.3 Variáveis de entrada e saída.

Um conjunto de variáveis de entrada representativas da composição dos materiais e das propriedades de empacotamento de partículas foi elaborado com base nos resultados experimentais publicados. O conjunto de dados utilizado para o desenvolvimento do modelo é composto por 365 amostras, fornecendo uma base suficientemente ampla e diversificada para o treinamento e a validação confiáveis dos modelos de aprendizado de máquina. Essas variáveis

incluem teor de cimento, teor de cinza volante, teor de agregado graúdo, proporção de agregado miúdo, percentual de substituição por agregado reciclado, teor de água, relação água/aglomerante, fração volumétrica de agregados, índice de empacotamento e idade de cura. Embora a densidade de empacotamento de partículas ou o índice de vazios não tenham sido calculados explicitamente nos estudos utilizados como fonte, essas variáveis atuam, em conjunto, como indicadores de empacotamento de partículas, uma vez que governam a distribuição espacial das partículas sólidas e da pasta na matriz de concreto (Cheng et al., 2021; de Brito & Saikia, 2013)

A resistência à compressão do concreto, determinada em idades de cura preestabelecidas e sob condições laboratoriais controladas, constitui a variável de saída. Os valores de resistência à compressão foram relatados de forma consistente nos estudos selecionados, permitindo uma comparação e agregação significativas dos dados (Ajdukiewicz & Kliszczewicz, 2002; Corinaldesi, 2010). A resistência à compressão foi considerada como parâmetro de saída, enquanto um conjunto de fatores relacionados à composição dos materiais e ao empacotamento de partículas foi selecionado como variáveis de entrada do modelo, com base no conjunto de dados compilado da literatura. A lista completa das variáveis de entrada e saída, incluindo seus símbolos, unidades e funções na estrutura de aprendizado de máquina, está resumida na Tabela 1. Foram calculadas estatísticas descritivas básicas para o conjunto de dados compilado, com o objetivo de fornecer informações sobre a distribuição e a variabilidade das variáveis de entrada e saída utilizadas na modelagem por aprendizado de máquina. Os valores mínimo, máximo, médio, desvio padrão e variância de cada variável estão resumidos na Tabela 2.

Tabela 1. Descrição das variáveis de entrada e saída utilizadas na modelagem por aprendizado de máquina.

Variável	Símbolo	Unidade	Descrição	Função no modelo
Idade	Idade	Dias	Período de cura dos corpos de prova de concreto	Entrada
Cimento	Cimento	kg/m ³	Teor de cimento Portland comum na mistura	Entrada
Cinza volante	FA	kg/m ³	Teor de cinza volante utilizada como material cimentício suplementar	Entrada
Agregado graúdo	CA	kg/m ³	Teor de agregado graúdo na mistura de concreto	Entrada
Água	Água	kg/m ³	Teor de água de amassamento	Entrada
Relação água/aglomerante	W/B	–	Relação entre a quantidade total de água de amassamento e o teor total de aglomerante	Entrada
Teor de agregado reciclado	RA (%)	%	Percentual de substituição do agregado natural por agregado	Entrada
Proporção de agregado miúdo	FAR	–	Proporção de agregado miúdo em relação ao teor total de agregados	Entrada
Fração volumétrica de agregados	AVF	–	Fração volumétrica dos agregados na mistura de concreto	Entrada
Índice de empacotamento	PI	–	Índice que representa a eficiência do empacotamento de partículas	Entrada
Resistência à compressão	CS	MPa	Resistência à compressão medida do concreto	Saída
Água	Água	kg/m ³	Teor de água de amassamento	Entrada
Relação água/aglomerante	W/B	–	Relação entre a quantidade total de água de amassamento e o teor total de aglomerante	Entrada

Nota: FAR indica a proporção de agregado miúdo, AVF indica a fração volumétrica de

Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.

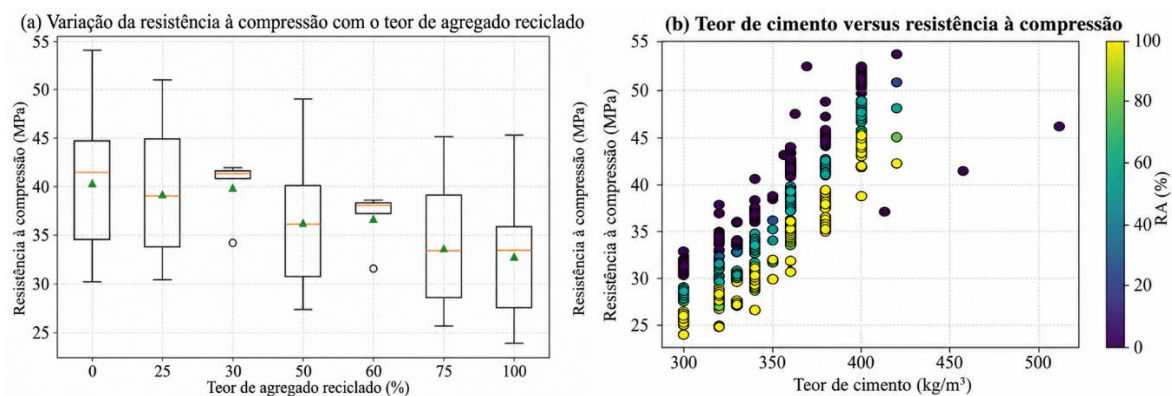
agregados, PI indica o índice de empacotamento e W/B indica a relação água/aglomerante.

Tabela 2. Resumo estatístico das variáveis do conjunto de dados utilizadas no estudo.

Parâmetro de entrada	Mínimo	Máximo	Média	Desvio padrão	Variável
Idade (dias)	28	28	28	0	0
Cimento (kg/m ³)	300	511.9	352.249	33.001	1089.069
Cinza volante (kg/m ³)	511	780	716.027	41.548	1726.209
Cinza volante (kg/m ³)	980	1330	1069.904	46.106	2125.783
Água (kg/m ³)	155.1	215	171.202	6.618	43.796
Relação água/aglomerante (W/B)	0.38	0.58	0.491	0.052	0.003
Teor de agregado reciclado (%)	0	100	48.932	39.757	1580.586
Proporção de agregado miúdo (FAR)	0.31	0.407	0.401	0.011	0
Fração volumétrica de agregados (AVF)	0.613	0.728	0.663	0.019	0
Índice de empacotamento (PI)	7.67	12.61	10.475	0.515	0.265
Resistência à compressão (MPa)	23.9	54.1	36.591	6.79	46.102

Nota: É importante observar que o conjunto de dados utilizado neste estudo está restrito a uma única idade de cura, de 28 dias, não havendo, portanto, variabilidade no parâmetro idade. Consequentemente, os modelos de aprendizado de máquina desenvolvidos são especificamente destinados à predição da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados, e não à representação da evolução da resistência ao longo do tempo. FAR indica a proporção de agregado miúdo, AVF indica a fração volumétrica de agregados, PI indica o índice de empacotamento e W/B indica a relação água/aglomerante.

Embora a resistência à compressão aos 28 dias seja amplamente aceita como referência padrão para o projeto estrutural e o controle de qualidade na engenharia do concreto, essa limitação implica que o modelo proposto não considera explicitamente os processos de hidratação dependentes do tempo. Sabe-se que materiais cimentícios suplementares, como a cinza volante, contribuem para o desenvolvimento da resistência após os 28 dias, frequentemente proporcionando ganhos significativos de resistência em idades de cura mais avançadas, como 56 ou 90 dias. Portanto, o presente estudo concentra-se na modelagem da resistência à compressão no período padrão de cura de 28 dias, que continua sendo o parâmetro mais frequentemente relatado e de maior relevância prática nos conjuntos de dados obtidos a partir da literatura. Pesquisas futuras deverão incorporar conjuntos de dados com múltiplas idades de cura, permitindo uma modelagem dependente do tempo e uma representação mais adequada da cinética de hidratação em longo prazo dos sistemas cimentícios compostos.



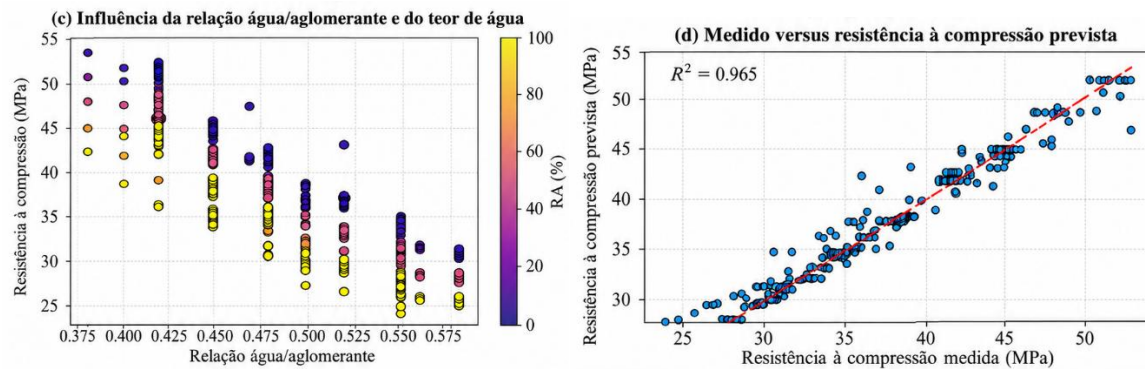


Figura 1. Caracterização do conjunto de dados, ilustrando (a) a variabilidade da resistência à compressão em função do teor de agregado reciclado, (b) a dependência da resistência à compressão em relação ao teor de cimento sob diferentes níveis de agregado reciclado, (c) os efeitos combinados da relação água/aglomerante e do teor de água sobre a resistência à compressão e (d) a resistência à compressão medida versus prevista com base nos parâmetros de entrada combinados ($R^2 = 0,965$).

A Figura 1 apresenta uma avaliação visual concisa do conjunto de dados de concreto autoadensável com agregados reciclados antes da modelagem por aprendizado de máquina. Os efeitos do teor de agregado reciclado, da dosagem de cimento e dos parâmetros relacionados à água sobre a resistência à compressão são apresentados nas Figuras 1(a-c), evidenciando a variabilidade dos materiais e os efeitos combinados das variáveis de dosagem. Com um elevado coeficiente de determinação, a Figura 1(d) demonstra uma forte correlação entre os valores medidos e previstos da resistência à compressão, utilizando os parâmetros de entrada combinados ($R^2 = 0,965$), confirmando que o conjunto de dados captura de forma eficaz os principais mecanismos que governam o desenvolvimento da resistência.

3.4 Pré-processamento e consistência dos dados.

Para garantir a compatibilidade entre os diferentes programas experimentais, todas as variáveis de entrada foram convertidas para unidades consistentes e normalizadas antes da análise por aprendizado de máquina. Os registros de dados que apresentavam informações incompletas sobre as misturas ou condições de ensaio ambíguas foram excluídos. O conjunto de dados final apresenta uma distribuição equilibrada dos teores de agregados reciclados, das composições dos aglomerantes e das relações água/aglomerante, permitindo que os modelos de aprendizado de máquina identifiquem tanto os efeitos individuais quanto os comportamentos de interação associados ao empacotamento de partículas. Esse conjunto de dados baseado na literatura mantém uma estreita relação com o comportamento dos sistemas de concreto com agregados reciclados observado experimentalmente, ao mesmo tempo em que fornece uma base fisicamente significativa para a modelagem orientada por dados.

4. METODOLOGIA

O fluxo de trabalho da estrutura de aprendizado de máquina proposta, desenvolvida para prever a resistência à compressão do concreto autoadensável com agregados reciclados (RA-SCC), é apresentado na Figura 2. O processo inicia-se com a compilação de um conjunto de dados baseado na literatura e prossegue com o pré-processamento, a normalização e o particionamento do conjunto de dados. Antes do desenvolvimento dos modelos, a multicolinearidade entre as variáveis relacionadas ao empacotamento de partículas é quantificada por meio da análise do fator de inflação da variância (VIF). Em seguida, diversos modelos de regressão e modelos de aprendizado de máquina por conjunto (*ensemble*) são treinados e avaliados utilizando métricas estatísticas de

desempenho e testes de hipóteses. Por fim, o método SHapley Additive exPlanations (SHAP) é empregado para interpretar as previsões dos modelos e identificar os indicadores de empacotamento de partículas mais influentes, assegurando tanto a precisão preditiva quanto a interpretabilidade física.

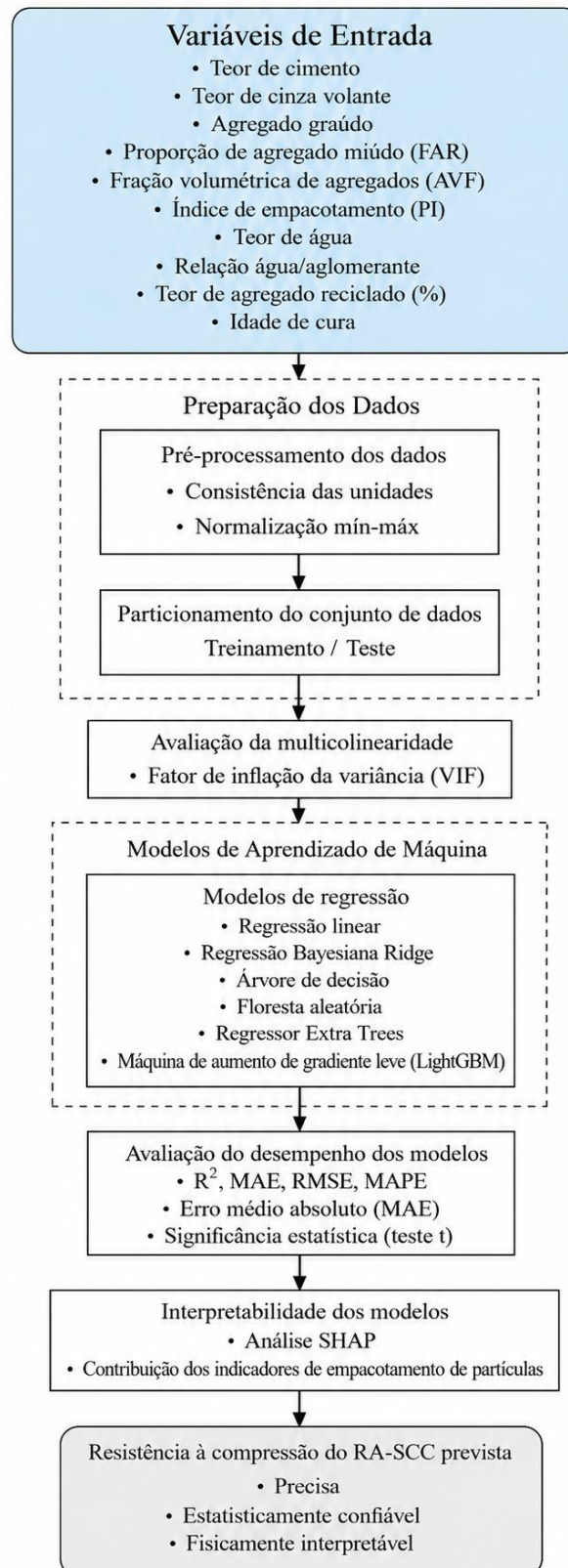


Figura 2. Fluxo de trabalho da estrutura de aprendizado de máquina proposta.

Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.

4.1 Visão geral da metodologia proposta

A forte não linearidade, a multi-colinearidade e a variabilidade dos materiais são características inerentes ao concreto autoadensável com agregados reciclados (RA-SCC), e a abordagem científica utilizada neste estudo foi desenvolvida para tratar essas dificuldades de forma abrangente. A metodologia foi dividida em quatro etapas consecutivas e interligadas: pré-processamento e normalização dos dados, avaliação da multicolinearidade, desenvolvimento dos modelos de aprendizado de máquina e avaliação de desempenho com análise explicável. Essa abordagem foi adotada para garantir tanto a precisão preditiva quanto a interpretabilidade sob a perspectiva da engenharia.

Na primeira etapa, o conjunto de dados compilado a partir da literatura foi submetido a um pré-processamento, a fim de garantir consistência, estabilidade numérica e compatibilidade entre as diferentes fontes experimentais. Essa etapa é fundamental na modelagem orientada por dados das propriedades do concreto, uma vez que os dados experimentais brutos frequentemente apresentam grandes variações de escala e distribuição, o que pode afetar negativamente o aprendizado e a convergência dos modelos (Liu et al., 2021; Yeh, 1998). A normalização das variáveis de entrada também possibilita uma comparação adequada entre os diferentes algoritmos de aprendizado de máquina.

Utilizando a análise do fator de inflação da variância (VIF), a segunda etapa concentrou-se na quantificação das interdependências entre as variáveis de entrada. Os parâmetros de dosagem associados ao empacotamento de partículas, como as frações de agregados, o teor de aglomerante e a relação água/aglomerante, são frequentemente altamente correlacionados. A avaliação explícita da multicolinearidade fornece uma base estatística para compreender as limitações dos modelos de regressão linear e auxilia na seleção de abordagens não paramétricas e de aprendizado por conjunto (*ensemble learning*), que apresentam maior robustez diante de variáveis de entrada correlacionadas (Kutner et al., 2005; Li et al., 2022).

Para capturar as interações complexas e não lineares entre os indicadores de empacotamento de partículas e a resistência à compressão, diversos modelos de regressão baseados em aprendizado de máquina foram desenvolvidos na terceira etapa. Foram utilizados tanto métodos tradicionais de regressão quanto modelos avançados de aprendizado por conjunto (*ensemble*) baseados em árvores, permitindo a comparação entre as diferentes abordagens e a investigação de como a complexidade dos algoritmos influencia os resultados das previsões (Asteris et al., 2020).

A etapa final integrou uma avaliação abrangente do desempenho com técnicas de inteligência artificial explicável. Além das métricas convencionais de precisão e erro, foram utilizadas a avaliação estatística de viés e testes de hipóteses para analisar a confiabilidade dos modelos. Além disso, o método SHapley Additive exPlanations (SHAP) foi utilizado para interpretar as previsões dos modelos e quantificar a contribuição individual dos indicadores de empacotamento de partículas, relacionando, assim, os resultados orientados por dados ao comportamento fisicamente significativo dos materiais (Almutairi, 2025; Lundberg & Lee, 2017). Essa metodologia integrada garante que a estrutura desenvolvida seja não apenas precisa, mas também transparente e adequada para aplicações em engenharia.

4.2 Normalização dos dados e particionamento do conjunto de dados

Todas as variáveis de entrada foram normalizadas utilizando o escalonamento Min-Max, com o objetivo de aumentar a estabilidade numérica durante o treinamento dos modelos e reduzir o impacto das diferenças de escala entre as variáveis de entrada.

O procedimento de normalização é expresso como:

$$x_{\text{norm}} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (1)$$

em que x representa o valor original de uma variável de entrada, e $x_{\min}$ and $x_{\max}$ indicam, respectivamente, os valores mínimo e máximo dessa variável no conjunto de dados. A normalização das variáveis é particularmente importante para algoritmos de regressão que envolvem cálculos de distância ou otimização iterativa, pois evita que variáveis com maiores amplitudes numéricas dominem o processo de aprendizado (Liu et al., 2021; Yeh, 1998).

Após a normalização, o conjunto de dados foi dividido aleatoriamente em subconjuntos de treinamento e teste, com o objetivo de avaliar a capacidade de generalização dos modelos de aprendizado de máquina desenvolvidos. Neste estudo, 30% do conjunto de dados foi reservado para testes independentes, enquanto os 70% restantes foram utilizados para o treinamento dos modelos e o ajuste dos hiperparâmetros. O subconjunto de treinamento foi utilizado exclusivamente para a calibração dos modelos e o ajuste dos hiperparâmetros, enquanto o subconjunto de teste não foi utilizado em nenhuma etapa do desenvolvimento dos modelos. Essa estratégia minimiza o vazamento de informações (*data leakage*) e permite uma avaliação rigorosa do desempenho preditivo em dados não utilizados durante o treinamento, aspecto particularmente importante para conjuntos de dados provenientes da literatura e caracterizados pela variabilidade dos materiais.

4.3 Avaliação da multicolinearidade utilizando o fator de inflação da variável

Os parâmetros de dosagem do concreto associados ao empacotamento de partículas são frequentemente fortemente correlacionados, o que pode afetar negativamente a estabilidade e a interpretabilidade dos modelos. Para quantificar o grau de multicolinearidade entre as variáveis de entrada, foi calculado o fator de inflação da variância (VIF) para cada variável.

O VIF é definido como:

$$VIF_i = \frac{1}{1-R_i^2} \quad (2)$$

em que R_i^2 é o coeficiente de determinação obtido pela regressão da variável preditora considerada em relação a todas as demais variáveis predictoras. Um valor de VIF superior a 10 geralmente indica multicolinearidade severa (Kutner et al., 2005; Li et al., 2022). Para quantificar o grau de multicolinearidade entre as variáveis de entrada relacionadas ao empacotamento de partículas, foram calculados os valores de VIF. Os valores resultantes para cada variável preditora são apresentados na Tabela 3. Valores de VIF extremamente elevados indicam que a interpretação convencional baseada nos coeficientes dos modelos não é confiável, reforçando, assim, a adequação dos modelos de aprendizado por conjunto (*ensemble*) baseados em árvores para esse problema.

Tabela 3. Valores do fator de inflação da variância (VIF) das variáveis de entrada.

Variável de entrada	Variável de entrada
Cimento	137.495
Cinza volante (FA)	1405.819
Agregado graúdo (CA)	580.775
Relação água/aglomerante (W/B)	240.825
Água	19.978
Teor de agregado reciclado (RA %)	1.025
FProporção de agregado miúdo (FAR)	777.145
Fração volumétrica de agregados (AVF)	35.664
Índice de empacotamento (PI)	12.89

Nota: Valores de VIF superiores a 10 indicam multi-colinearidade severa entre as variáveis predictoras. FAR indica a proporção de agregado miúdo, AVF indica a fração volumétrica de agregados, PI indica o índice de empacotamento e W/B indica a relação água/aglomerante.

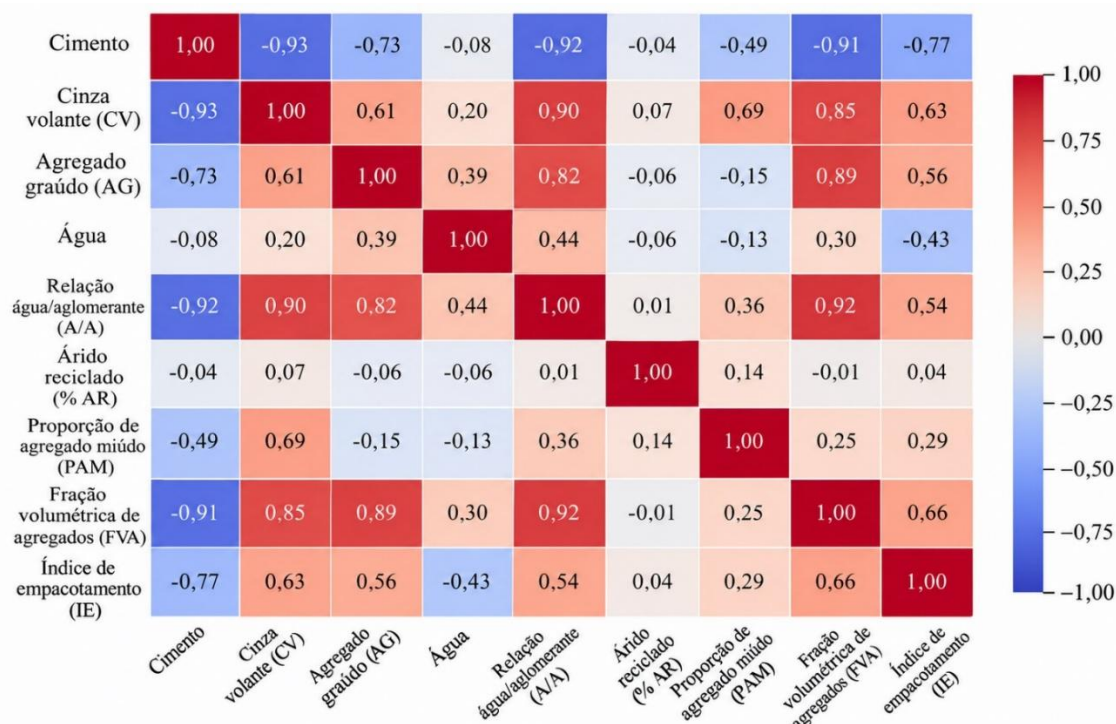


Figura 3. Matriz de correlação das variáveis de entrada relacionadas ao empacotamento de partículas utilizadas para a predição da resistência à compressão do RA-SCC.

A matriz de correlação de Pearson das variáveis de entrada relacionadas ao empacotamento de partículas consideradas nesta investigação é apresentada na Figura 3. A presença de multicolinearidade é indicada pelas fortes correlações observadas entre diversas variáveis relacionadas aos aglomerantes e aos agregados. Esse aspecto é investigado quantitativamente na seção seguinte por meio da análise do fator de inflação da variância (VIF), a fim de avaliar suas implicações para a estabilidade e a seleção dos modelos. As fortes correlações negativas observadas entre o teor de cimento e o teor de cinza volante ($r = -0,93$), a relação água/aglomerante ($r = -0,92$) e a fração volumétrica de agregados ($r = -0,91$) indicam uma interdependência significativa entre a composição do aglomerante, a qualidade da pasta e as características volumétricas de empacotamento. Por outro lado, o teor de cinza volante apresenta fortes correlações positivas com a fração volumétrica de agregados ($r = 0,85$) e com a relação água/aglomerante ($r = 0,90$), indicando a atuação combinada da dispersão da pasta e do teor de finos nas misturas de concreto autoadensável.

As restrições volumétricas que controlam o empacotamento de partículas são evidenciadas pelas correlações moderadas a significativas observadas entre o teor de agregado graúdo e a relação água/aglomerante ($r = 0,82$) e a fração volumétrica de agregados ($r = 0,89$). A relativa independência estatística do teor de agregado reciclado no conjunto de dados é confirmada por suas baixas correlações com todas as demais variáveis. As fortes correlações entre as variáveis relacionadas aos aglomerantes e aos agregados indicam uma multicolinearidade significativa. Esse resultado sustenta a utilização de modelos de aprendizado de máquina baseados em árvores e métodos de conjunto (*ensemble*) para uma predição robusta da resistência e justifica a subsequente análise do fator de inflação da variância (VIF).

4.4 Modelos de regressão de aprendizado de máquina.

Para prever a resistência à compressão do concreto autoadensável com agregados reciclados, foram desenvolvidos seis modelos de regressão.

4.4.1 Regressão linear e regressão Ridge Bayesiana.

De acordo com o modelo de regressão linear, as variáveis de entrada e a resistência à compressão apresentam uma relação linear, representada da seguinte forma:

$$\hat{y} = \beta_0 + \sum_{i=1}^n \beta_i x_i \quad (3)$$

em que $\hat{y}$ representa a resistência à compressão prevista, x_i representam as variáveis de entrada, e β_i são os coeficientes de regressão.

A regressão Ridge Bayesiana amplia essa formulação por meio da introdução de distribuições probabilísticas a priori nos coeficientes de regressão, com o objetivo de melhorar a estabilidade numérica na presença de multicolinearidade (Montgomery et al., 2012).

4.4.2 Modelos baseados em árvores de decisão.

A regressão por árvore de decisão, um método de aprendizado supervisionado não paramétrico, modela relações não lineares por meio da divisão recursiva do espaço das variáveis de entrada em um conjunto de regiões disjuntas. Em cada nó interno da árvore, o conjunto de dados é dividido com base em uma variável de entrada selecionada e em um valor limite correspondente, com o objetivo de minimizar o erro de predição nos subconjuntos resultantes. Esse processo recursivo de divisão binária continua até que sejam atendidos critérios de parada predefinidos, como o número mínimo de amostras por nó ou a profundidade máxima da árvore. A média dos valores observados correspondentes às amostras contidas em um determinado nó terminal (folha) é utilizada para calcular a resistência à compressão prevista para esse nó. A função de predição de um modelo de árvore de decisão pode ser expressa da seguinte forma:

$$\hat{y}(x) = \sum_{m=1}^M c_m \mathbb{I}(x \in R_m) \quad (4)$$

em que M representa o número total de regiões terminais, R_m indica a região do espaço de variáveis, c_m representa a resistência à compressão média das amostras contidas na região, e $\mathbb{I}(\cdot)$ é uma função indicadora, que assume valor 1 quando a condição é satisfeita e 0 caso contrário (Breiman et al., 1984).

Em problemas de regressão, a divisão ótima em cada nó é geralmente determinada pela minimização da variância ou do erro quadrático médio da variável-alvo nos nós filhos. Esse critério permite que as árvores de decisão capturem relações complexas entre as variáveis de entrada, sem a necessidade de estabelecer hipóteses sobre sua estrutura funcional. Como os indicadores de empacotamento de partículas e os parâmetros de dosagem interagem de forma altamente não linear em sistemas de materiais heterogêneos, como o concreto autoadensável com agregados reciclados, os modelos de árvore de decisão são, portanto, particularmente adequados para a modelagem desses sistemas (Asteris et al., 2020; Yeh, 1998).

4.4.3 Regressores Random Forest e Extra Trees

Na regressão Random Forest, é construído um conjunto (*ensemble*) de árvores de decisão, treinadas a partir de amostras obtidas por reamostragem *bootstrap*, sendo a predição final obtida pela média das predições das árvores individuais (Breiman, 2001):

$$\hat{y} = \frac{1}{M} \sum_{m=1}^M \hat{y}_m \quad (5)$$

em que M representa o número de árvores, e $\hat{y}_m$ representa a predição realizada pela m^{th} árvore.

O regressor Extra Trees aumenta ainda mais o grau de aleatoriedade ao selecionar aleatoriamente os limiares de divisão, o que melhora a capacidade de generalização e reduz a

Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.

variância do modelo, particularmente em conjuntos de dados com variáveis preditoras correlacionadas.

4.4.4 *light gradient boosting machine*.

O Light Gradient Boosting Machine (LightGBM) é uma estrutura de gradient boosting que constrói, de forma sequencial, árvores de regressão aditivas. O resultado previsto é dado por:

$$\hat{y}_k = \sum_{t=1}^T f_t(x_k) \quad (6)$$

em que f_t representa a t^{th} árvore de decisão, e T corresponde ao número total de iterações de *boosting* (Ke et al., 2017).

4.5 Métricas de avaliação do desempenho dos modelos.

Para garantir uma avaliação abrangente e estatisticamente confiável do desempenho dos modelos, foram utilizados diversos indicadores quantitativos para avaliar a precisão das predições, a magnitude dos erros, o viés sistemático e a consistência estatística entre os valores previstos e experimentais da resistência à compressão. A utilização de múltiplas métricas de avaliação é essencial na predição da resistência do concreto, uma vez que a adoção de um único indicador pode levar a conclusões equivocadas quanto à robustez do modelo (Montgomery et al., 2012; Willmott & Matsuura, 2005).

O coeficiente de determinação (R^2) foi utilizado para quantificar a proporção da variância da resistência à compressão medida que é explicada pelas predições do modelo. Ele é definido como:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (7)$$

em que y_i e $\hat{y}_i$ representam, respectivamente, os valores de resistência à compressão medidos experimentalmente e previstos para a i^{th} amostra, $\bar{y}$ representa a média dos valores medidos, e N corresponde ao número total de observações. Um valor mais elevado de R^2 indica uma maior concordância entre as predições e os resultados experimentais.

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (8)$$

O erro absoluto médio (MAE) fornece uma medida direta da magnitude média dos erros de predição e é menos sensível a valores extremos quando comparado às métricas baseadas em erros quadráticos. Por outro lado, a raiz do erro quadrático médio (RMSE) penaliza de forma mais acentuada os erros de maior magnitude e é expressa como:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (9)$$

Em aplicações de engenharia, nas quais desvios mais elevados em relação aos valores medidos podem apresentar implicações estruturais importantes, o RMSE é particularmente útil para a avaliação do desempenho dos modelos (Willmott & Matsuura, 2005).

Para quantificar a superestimação ou subestimação sistemática das predições, foi calculado o erro médio de viés (MBE) da seguinte forma:

$$\text{MBE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i) \quad (10)$$

Uma tendência à subestimação é indicada por um MBE positivo, enquanto um valor negativo indica superestimação. Valores próximos de zero indicam previsões não enviesadas, condição necessária para uma tomada de decisão confiável em engenharia.

Para determinar se as diferenças entre os valores experimentais e previstos da resistência à compressão eram estatisticamente significativas, foi realizado um teste de hipótese estatística, além da utilização das métricas baseadas em erro. Foi aplicado um teste t pareado para avaliar a hipótese nula de que a diferença média entre os valores medidos e previstos é igual a zero. A concordância estatística foi avaliada por meio da estatística t resultante e do valor-p associado; para um nível de confiança de 95%, valores de p superiores a 0,05 indicaram ausência de diferença estatisticamente significativa (Montgomery et al., 2012). Essa validação estatística aumenta a confiabilidade das previsões realizadas pelos modelos de aprendizado de máquina desenvolvidos.

4.6 Interpretabilidade do modelo baseada em SHAP.

O método SHapley Additive exPlanations (SHAP) foi utilizado para interpretar o papel dos indicadores de empacotamento de partículas nas previsões do modelo. Os valores SHAP satisfazem a propriedade de aditividade e são derivados da teoria dos jogos cooperativos:

$$f(x) = \phi_0 + \sum_{i=1}^n \phi_i \quad (11)$$

em que ϕ_i representa a contribuição da i^{th} característica para a predição (Lundberg & Lee, 2017). Essa abordagem permite quantificar tanto a magnitude quanto a direção da influência de cada variável, possibilitando uma relação direta entre os resultados do aprendizado de máquina e os mecanismos físicos de empacotamento de partículas. É importante observar que os valores SHAP quantificam a contribuição das variáveis de entrada para as previsões do modelo, não estabelecendo, contudo, uma relação de causalidade física estrita. Em conjuntos de dados caracterizados por forte multi-colinearidade, como no presente estudo, variáveis correlacionadas podem compartilhar ou redistribuir entre si suas contribuições. Consequentemente, a importância atribuída a cada variável individual deve ser interpretada com cautela, especialmente quando são observados valores elevados do fator de inflação da variância (VIF). No conjunto de dados utilizado neste estudo, o teor de cimento, o teor de cinza volante e a relação água/aglomerante apresentam forte interdependência, refletindo restrições práticas inerentes à dosagem do concreto, em vez de representarem variáveis físicas independentes. Nessas condições, a importância baseada em SHAP deve ser interpretada como uma representação da influência conjunta das variáveis correlacionadas, e não como efeitos causais isolados. Apesar dessa limitação, os resultados de SHAP obtidos neste estudo permanecem fisicamente coerentes, uma vez que as variáveis identificadas como predominantes — teor de agregado reciclado, teor de cimento e relação água/aglomerante — são reconhecidamente fatores determinantes que influenciam a resistência à compressão e o comportamento do empacotamento de partículas. A consistência entre a classificação de importância obtida por SHAP e os princípios consolidados da mecânica do concreto fornece confiança de que o modelo captura um comportamento realista do material, mesmo na presença de multicolinearidade. Portanto, neste estudo, a análise SHAP é interpretada como uma ferramenta de explicabilidade que fornece informações sobre a influência das variáveis dentro de um sistema correlacionado, e não como um método de atribuição causal estrita. Essa distinção é essencial para garantir uma interpretação cientificamente rigorosa dos resultados.

4.7 Implementação computacional

Para garantir reprodutibilidade, transparência e eficiência computacional, foram utilizadas bibliotecas de computação científica baseadas em Python em todas as etapas de processamento de dados, modelagem por aprendizado de máquina e análises estatísticas.

O Python foi selecionado devido ao seu amplo ecossistema de ferramentas de código aberto, amplamente adotadas em pesquisas de materiais orientadas por dados e em aplicações de engenharia civil (McKinney, 2010; Pedregosa et al., 2011).

O fluxo de trabalho computacional compreendeu etapas sequenciais, incluindo a normalização dos dados, o particionamento do conjunto de dados, o cálculo do fator de inflação da variância (VIF), o desenvolvimento dos modelos, a avaliação de desempenho e a análise por aprendizado de máquina explicável. Para proporcionar uma comparação justa do desempenho, os procedimentos de normalização das variáveis e pré-processamento dos dados foram aplicados de maneira uniforme a todos os modelos. Foram utilizadas funções padronizadas de bibliotecas computacionais para o desenvolvimento dos algoritmos de aprendizado de máquina, permitindo processos confiáveis e reprodutíveis de treinamento e teste dos modelos. O ajuste dos hiperparâmetros foi realizado no conjunto de dados de treinamento utilizando validação cruzada, com o objetivo de melhorar a capacidade de generalização dos modelos e evitar o sobreajuste (*overfitting*). Para evitar o vazamento de informações (*data leakage*) e garantir uma avaliação objetiva do desempenho preditivo, todos os modelos treinados foram testados exclusivamente no conjunto de dados de teste independente. As métricas estatísticas de avaliação e os testes de hipóteses foram calculados utilizando bibliotecas numéricas consolidadas, a fim de assegurar precisão e estabilidade numérica (Montgomery et al., 2012). A interpretabilidade do modelo foi obtida por meio do método SHapley Additive exPlanations (SHAP), cujos valores foram calculados utilizando algoritmos eficientes de aproximação compatíveis com modelos baseados em árvores. Essa abordagem permite uma atribuição consistente das contribuições das variáveis, tanto para previsões individuais quanto para o comportamento global do modelo, possibilitando uma interpretação significativa dos indicadores de empacotamento de partículas que influenciam a resistência à compressão (Lundberg & Lee, 2017). As etapas intermediárias do fluxo de trabalho de modelagem, incluindo pré-processamento dos dados, avaliação da multicolinearidade, treinamento dos modelos e avaliação de desempenho, foram sistematicamente validadas por meio da abordagem de divisão dos dados em conjuntos de treinamento e teste. A validação cruzada foi empregada durante o ajuste dos hiperparâmetros para garantir a estabilidade dos modelos e evitar o sobreajuste (*overfitting*). Portanto, os resultados apresentados refletem tanto a precisão dos modelos quanto sua capacidade de generalização, com base em dados de teste não utilizados durante o treinamento.

A utilização de software de código aberto e de práticas padronizadas de implementação garante que a estrutura desenvolvida possa ser facilmente reproduzida e ampliada por outros pesquisadores. Essa abordagem aberta facilita tanto o desenvolvimento futuro de modelos de aprendizado de máquina explicável aplicados a materiais de concreto sustentáveis quanto sua validação e avaliação comparativa (*benchmarking*).

5. RESULTADOS E DISCUSSÃO

5.1 Comparação do desempenho dos modelos de aprendizado de máquina.

Diversas métricas estatísticas, como o coeficiente de determinação (R^2), o erro absoluto médio (MAE), a raiz do erro quadrático médio (RMSE), o erro percentual absoluto médio (MAPE), o erro médio de viés (MBE) e testes de hipóteses, foram utilizadas para avaliar o desempenho preditivo dos seis modelos de aprendizado de máquina. Os resultados demonstram claramente que, na predição da resistência à compressão do concreto autoadensável com agregados

Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.

reciclados (RA-SCC), os modelos baseados em métodos de conjunto (*ensemble*) apresentam desempenho superior aos modelos lineares e aos modelos baseados em uma única árvore de decisão. Com um valor de $R^2 = 0,9724$, o Extra Trees Regressor apresentou a maior precisão preditiva entre todos os modelos, indicando que foi capaz de explicar mais de 97% da variabilidade da resistência à compressão. Esse modelo também apresentou os menores valores de MAE (0,79 MPa) e RMSE (1,19 MPa), demonstrando desempenho superior na minimização dos erros. Os modelos Random Forest e Light Gradient Boosting Machine também apresentaram excelente desempenho, com valores de R^2 de 0,968 e 0,965, respectivamente. Esses resultados corroboram as conclusões de estudos anteriores sobre concreto com agregados reciclados, segundo as quais as técnicas de aprendizado por conjunto (*ensemble*) são adequadas para modelar o comportamento altamente não linear do concreto (Ahmed et al., 2024; Asteris et al., 2020; Liu et al., 2021). A Tabela 4 apresenta um resumo do coeficiente de determinação, das métricas baseadas em erro, do erro médio de viés e dos resultados dos testes de hipóteses para cada modelo.

Tabela 4. Comparação do desempenho dos modelos de aprendizado de máquina para a predição da resistência à compressão.

Modelo	R^2	MAE (MPa)	MSE (MPa ²)	RMSE (MPa)	MAPE (%)	MBE (MPa)	t-Estatística	Valor p
Regressão linear	0.9526	1.336	2.444	1.563	3.8	-0.042	-0.282	0.779
Regressão Ridge Bayesiana	0.9146	1.525	4.404	2.099	4.13	-0.042	-0.211	0.833
Regressor de árvore de decisão	0.9363	1.167	3.28	1.811	3.06	0.203	1.177	0.242
Regressor Random Forest	0.968	0.855	1.648	1.284	2.31	0.095	0.771	0.443
Regressor Extra Trees	0.9724	0.788	1.423	1.193	2.12	0.122	1.074	0.285
Máquina de Reforço de Gradiente Leve	0.965	0.866	1.802	1.343	2.3	0.205	1.61	0.11

Nota: MAE indica o erro absoluto médio, MSE indica o erro quadrático médio, RMSE indica a raiz do erro quadrático médio, MAPE indica o erro percentual absoluto médio e MBE indica o erro médio de viés.

A precisão das predições e os níveis de erro da Regressão Ridge Bayesiana e da Regressão Linear foram significativamente inferiores. A principal causa dessa redução de desempenho é a forte interdependência entre as variáveis de dosagem, o que contraria as premissas da modelagem linear. Estudos anteriores sobre a predição da resistência do concreto baseada em aprendizado de máquina documentaram limitações semelhantes da regressão linear na presença de multicolinearidade (Li et al., 2022).

A Figura 4 apresenta gráficos de dispersão que comparam os valores experimentais e previstos da resistência à compressão para cada um dos seis modelos de aprendizado de máquina considerados nesta investigação. O agrupamento concentrado dos pontos de dados em torno da linha de referência de 45° indica elevada precisão preditiva e baixo viés nos modelos baseados em métodos de conjunto (*ensemble*), especialmente nos regressores Extra Trees e Random Forest. O Light Gradient Boosting Machine também apresenta forte concordância, embora exiba uma dispersão ligeiramente maior para valores mais elevados de resistência. Em contrapartida, os modelos de Regressão Linear, Regressão Ridge Bayesiana e Árvore de Decisão apresentam maior dispersão, refletindo sua menor capacidade de capturar as interações

não lineares entre as variáveis relacionadas ao empacotamento de partículas. Os padrões visuais corroboram a maior confiabilidade preditiva das técnicas de aprendizado por conjunto (*ensemble*) para o concreto autoadensável com agregados reciclados e estão de acordo com os indicadores quantitativos de desempenho apresentados na Tabela 4.

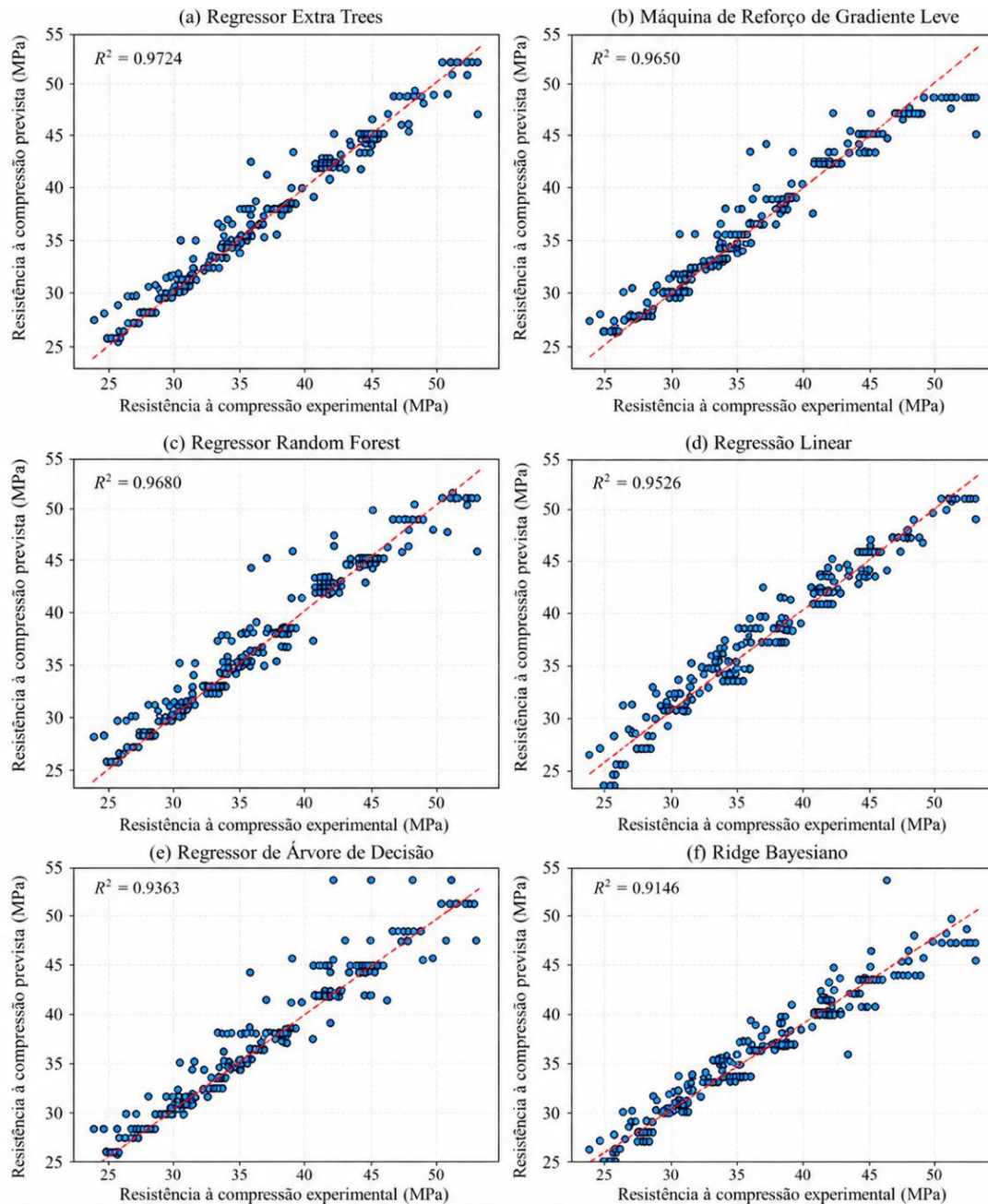


Figura 4. Comparação entre a resistência à compressão prevista e experimental obtida utilizando diferentes modelos de aprendizado de máquina: (a) regressor Extra Trees, (b) Máquina de Reforço de Gradiente Leve, (c) regressor Random Forest, (d) regressão linear, (e) regressor de árvore de decisão e (f) regressão Ridge Bayesiana.

A Figura 5 apresenta as distribuições dos resíduos obtidas para os seis modelos de aprendizado de máquina, fornecendo informações sobre o viés das previsões e a dispersão dos erros, além das métricas convencionais de precisão. O Regressor Extra Trees apresenta um perfil de resíduos estreitamente distribuído, aproximadamente simétrico e centrado em zero, consistente com seu baixo erro médio de viés (MBE = 0,122 MPa), reduzido RMSE (1,19 MPa) e elevado Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.

coeficiente de determinação ($R^2 = 0,9724$). A elevada estabilidade preditiva também é confirmada pelo Regressor Random Forest, que apresenta uma distribuição compacta dos resíduos e baixo viés (MBE = 0,0945 MPa; RMSE = 1,28 MPa). Embora apresente uma dispersão ligeiramente maior, evidenciada por seu RMSE de 1,34 MPa, a Máquina de Reforço de Gradiente Leve (Light Gradient Boosting Machine) também apresenta resíduos concentrados próximos de zero (MBE = 0,2046 MPa). Em contrapartida, os modelos de Regressão Linear e Regressão Ridge Bayesianas apresentam distribuições de resíduos mais amplas, apesar dos valores de viés próximos de zero (MBE $\approx -0,04$ MPa), indicando maior erro aleatório e menor capacidade de representar as interações não lineares entre as variáveis relacionadas ao empacotamento de partículas. O Regressor de Árvore de Decisão apresenta assimetria e dispersão moderadas, em concordância com seu RMSE comparativamente mais elevado (1,81 MPa).

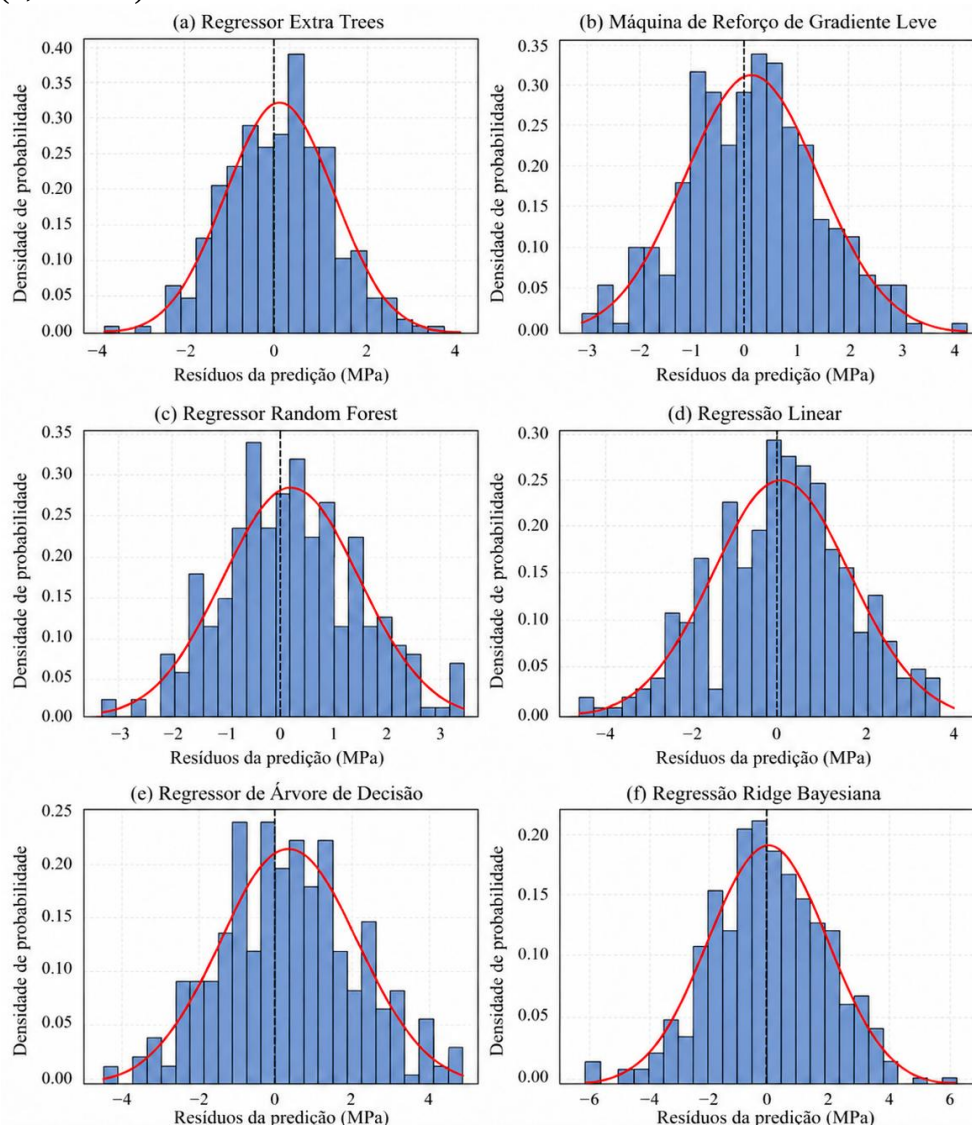


Figura 5. Distribuição dos resíduos de predição obtidos utilizando diferentes modelos de aprendizado de máquina: (a) regressor Extra Trees, (b) Máquina de Reforço de Gradiente Leve, (c) regressor Random Forest, (d) regressão linear, (e) regressor de árvore de decisão e (f) regressão Ridge Bayesianas.

5.2 Confiabilidade estatística e avaliação do viés.

Além das métricas convencionais de precisão, a confiabilidade estatística foi avaliada por meio do erro médio de viés (MBE) e de testes de hipóteses. Os valores de MBE para todos os modelos

ficaram próximos de zero, indicando a ausência de superestimação ou subestimação sistemática da resistência à compressão. Destaca-se que o Regressor Extra Trees apresentou um MBE de 0,12 MPa, confirmando sua capacidade de realizar previsões sem viés significativo. A estatística *t* e os respectivos valores-*p* reforçam essa observação. Para todos os modelos, os valores-*p* foram superiores ao nível de significância de 0,05, indicando que não houve diferença estatisticamente significativa entre os valores previstos e experimentais da resistência à compressão. A consistência estatística observada nos modelos de conjunto (*ensemble*) aumenta a confiança em sua aplicabilidade na tomada de decisões em engenharia, na qual previsões confiáveis e sem viés são essenciais (Montgomery et al., 2012).

5.3 Influência da multi-colinearidade no desempenho dos modelos.

A análise do fator de inflação da variância (VIF) revelou multicolinearidade severa entre diversas variáveis de entrada relacionadas ao empacotamento de partículas. O teor de cinza volante, o teor de agregado graúdo, a proporção de agregado miúdo e a relação água/aglomerante apresentaram valores de VIF extremamente elevados, confirmando a forte interdependência entre esses parâmetros. Essa multicolinearidade explica o desempenho inferior dos modelos de regressão linear e regressão Bayesiana, uma vez que a instabilidade dos coeficientes e o aumento da variância comprometem a confiabilidade das previsões. Conforme apresentado na Tabela 3, diversas variáveis relacionadas ao empacotamento de partículas apresentam multicolinearidade severa, particularmente o teor de cinza volante, as frações de agregados e a relação água/aglomerante, enquanto o teor de agregado reciclado permanece estatisticamente independente. A taxa de substituição por agregado reciclado apresentou um valor de VIF próximo à unidade, indicando independência estatística. Essa característica torna o teor de agregado reciclado uma variável preditora robusta, tanto em estruturas de modelagem linear quanto não linear. O desempenho superior dos modelos de conjunto (*ensemble*) baseados em árvores, observado neste estudo, está de acordo com resultados de pesquisas anteriores que demonstram que esses algoritmos são inerentemente robustos à multicolinearidade e capazes de capturar interações complexas entre as variáveis sem exigir uma descorrelação explícita das variáveis de entrada (Breiman, 2001; Li et al., 2022). Parte da multicolinearidade observada também pode ser atribuída a inconsistências na forma de apresentação dos tipos de cimento e dos materiais cimentícios suplementares entre os diferentes estudos. Em particular, a utilização de cimentos compostos contendo cinza volante pode resultar em uma representação sobreposta dos componentes do aglomerante, ampliando, assim, as correlações estatísticas entre as variáveis de entrada.

5.4 Interpretação dos indicadores de empacotamento de partículas baseada em SHAP.

Para determinar a importância relativa dos indicadores de empacotamento de partículas e interpretar as previsões do modelo de melhor desempenho, foi realizada uma análise por meio do método SHapley Additive exPlanations (SHAP). Os resultados mostram que o fator mais importante é a porcentagem de substituição por agregado reciclado, seguido pelo teor de cimento e pela relação água/aglomerante. Essa classificação evidencia o papel predominante da qualidade dos agregados, da disponibilidade de pasta e da estrutura interna de vazios no desenvolvimento da resistência à compressão do RA-SCC (Tang et al., 2023). A forte influência do teor de agregado reciclado pode ser atribuída ao seu impacto sobre a eficiência do empacotamento de partículas e as características da zona de transição interfacial. Os agregados reciclados introduzem porosidade adicional e maior rugosidade superficial devido à presença de argamassa aderida, o que afeta tanto a densidade de empacotamento quanto os mecanismos de transferência de tensões (Etxeberria et al., 2007; Poon et al., 2004; H. Zhang et al., 2022). O teor de cimento e a relação água/aglomerante controlam diretamente o volume de pasta e os produtos de hidratação, influenciando ainda mais o preenchimento dos vazios e a conectividade

entre as partículas. Para o modelo de aprendizado de máquina de melhor desempenho, o método SHapley Additive exPlanations foi utilizado para analisar a contribuição individual dos indicadores de empacotamento de partículas para a predição da resistência à compressão. O gráfico-resumo SHAP mostra como os valores SHAP individuais estão distribuídos ao longo do conjunto de dados de treinamento, enquanto o gráfico de barras de importância das variáveis foi elaborado utilizando os valores médios absolutos finais de SHAP obtidos a partir do Regressor Extra Trees treinado. A Tabela 5 apresenta um resumo dos valores médios absolutos de SHAP e das respectivas classificações de importância das variáveis. Essa consistência entre os resultados obtidos por meio do aprendizado de máquina explicável e os princípios consolidados da mecânica do concreto aumenta a confiança na aplicabilidade prática da estrutura proposta.

Tabela 5. Classificação da importância das variáveis baseada em SHAP para o modelo de melhor desempenho.

Classificação	Variável	Valor médio absoluto de SHAP
1	Teor de agregado reciclado (RA %)	2.476
2	Cimento	1.483
3	Relação água/aglomerante (W/B)	1.401
4	Cinza volante (FA)	0.728
5	Agregado graúdo (CA)	0.703
6	Fração volumétrica de agregados (AVF)	0.694
7	Índice de empacotamento (PI)	0.157
8	Água	0.088
9	Proporção de agregado miúdo (FAR)	0.08
10	Idade	0

Nota: Os valores médios absolutos de SHAP representam a magnitude média da contribuição de cada variável para as predições do modelo. Valores mais elevados indicam maior influência sobre a resistência à compressão. FAR indica a proporção de agregado miúdo, AVF indica a fração volumétrica de agregados, PI indica o índice de empacotamento e W/B indica a relação água/aglomerante.

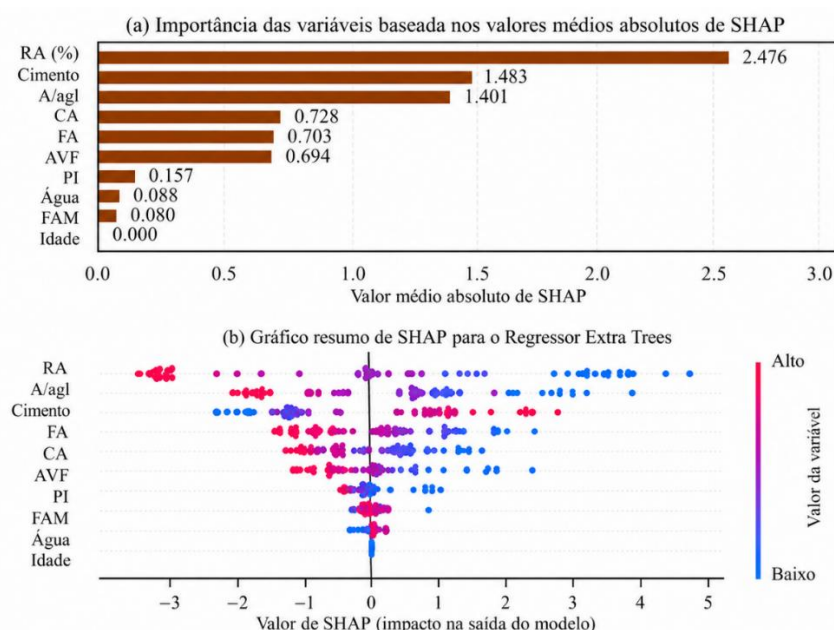


Figura 6. Interpretação baseada em SHAP do Regressor Extra Trees, mostrando (a) a importância das variáveis classificada com base nos valores médios absolutos de SHAP e (b) o gráfico-resumo SHAP, ilustrando a influência das variáveis de entrada relacionadas ao empacotamento de partículas na predição da resistência à compressão.

Predição explicável, baseada em aprendizado de máquina, da resistência à compressão aos 28 dias do concreto autoadensável com agregados reciclados utilizando indicadores de empacotamento de partículas.

A Figura 6 apresenta a análise de explicabilidade baseada em SHAP para o Regressor Extra Trees, fornecendo uma avaliação quantitativa da contribuição relativa das variáveis relacionadas ao empacotamento de partículas para a predição da resistência à compressão. De acordo com o gráfico de importância das variáveis apresentado na Figura 6(a), a taxa de substituição por agregado reciclado é a variável preditora mais significativa, apresentando o maior valor médio absoluto de SHAP (2,48), seguida pelo teor de cimento (1,48) e pela relação água/aglomerante (1,40). Cabe destacar também que a importância relativamente maior do teor de agregado reciclado pode ser parcialmente influenciada por sua ampla faixa de variação (0–100%), quando comparada às demais variáveis de entrada. Essa maior variabilidade aumenta sua contribuição estatística dentro do modelo e deve ser considerada na interpretação da classificação de importância das variáveis. Essas variáveis controlam diretamente a qualidade dos agregados, a disponibilidade de pasta e a estrutura interna de vazios, fatores que, em conjunto, governam a eficiência do empacotamento de partículas no concreto autoadensável. O teor de cinza volante (0,73), o teor de agregado graúdo (0,70) e a fração volumétrica de agregados (0,69) apresentam contribuições moderadas, indicando seu papel secundário no ajuste da densidade de empacotamento e da dispersão da pasta. Em contrapartida, o índice de empacotamento (0,16), o teor de água (0,09) e a proporção de agregado miúdo (0,08) apresentam influência relativamente limitada, enquanto a idade de cura apresenta contribuição desprezível no conjunto de dados analisado.

O gráfico-resumo SHAP apresentado na Figura 6(b) ilustra, adicionalmente, a direção da influência dessas variáveis nas predições individuais. Maiores teores de agregado reciclado e valores mais elevados da relação água/aglomerante exercem, em geral, um efeito negativo sobre a resistência à compressão, enquanto maiores teores de cimento contribuem positivamente.

Os resultados da análise SHAP confirmam que o modelo Extra Trees é capaz de capturar mecanismos de empacotamento de partículas fisicamente coerentes, fornecendo um suporte transparente e orientado por dados para a estrutura preditiva proposta. Deve-se destacar também que, devido à multicolinearidade entre as variáveis relacionadas aos aglomerantes, os valores SHAP refletem contribuições distribuídas entre variáveis de entrada correlacionadas. Consequentemente, a importância relativa das variáveis individuais deve ser interpretada em conjunto com o conhecimento técnico do domínio, e não como representação de efeitos causais independentes.

O teor de materiais cimentícios suplementares e os parâmetros volumétricos dos agregados apresentaram contribuições moderadas, indicando seu papel secundário, porém não desprezível, no aprimoramento do empacotamento de partículas e da microestrutura. Por outro lado, a idade de cura apresentou influência desprezível no conjunto de dados compilado, sugerindo que os parâmetros de composição da mistura predominam na variação da resistência dentro da faixa de dados selecionada. É importante interpretar criticamente a importância relativamente baixa do índice de empacotamento (PI) e da proporção de agregado miúdo (FAR), apesar de sua relevância conceitual para o comportamento do empacotamento de partículas. Embora esses parâmetros tenham como objetivo representar características de empacotamento, eles constituem indicadores derivados ou secundários, dependentes das variáveis subjacentes de composição da mistura. Como resultado, sua contribuição independente para as predições do modelo é comparativamente limitada. Em contraste com a variabilidade da dosagem, resultados semelhantes foram documentados em conjuntos de dados provenientes da literatura com distribuição restrita de idades (Asteris et al., 2020; Fanijo et al., 2023). A análise SHAP indica que variáveis primárias de dosagem, como o teor de agregado reciclado, o teor de cimento e a relação água/aglomerante, exercem maior influência sobre a resistência à compressão. Isso ocorre porque esses parâmetros controlam diretamente o volume de pasta, a interação entre as partículas e a estrutura de vazios, que, em conjunto, determinam a eficiência do empacotamento. Em contrapartida, o PI e o FAR representam esses efeitos de

maneira agregada ou indireta, resultando em menor importância individual no modelo de aprendizado de máquina. Essa observação não contradiz a estrutura baseada no empacotamento de partículas proposta neste estudo. Ao contrário, demonstra que o modelo de aprendizado de máquina captura implicitamente os mecanismos de empacotamento por meio das variáveis predominantes de dosagem, sem depender fortemente de índices de empacotamento explicitamente definidos. Os resultados sugerem que o comportamento do empacotamento está intrinsecamente incorporado às relações entre o teor de materiais cimentícios, as proporções de agregados e a demanda de água, em vez de ser governado por uma única variável indicadora. Portanto, a classificação relativamente inferior do PI e do FR deve ser interpretada como uma indicação de que os efeitos do empacotamento de partículas estão distribuídos entre múltiplas variáveis interativas, em vez de estarem isolados em índices específicos. Isso reforça a capacidade do aprendizado de máquina explicável de capturar implicitamente mecanismos físicos complexos, em consonância com o objetivo central deste estudo.

5.5 Implicações para a dosagem orientada pelo empacotamento de partículas.

Os resultados combinados das análises de VIF e SHAP mostram que o aprendizado de máquina pode capturar implicitamente os mecanismos de empacotamento de partículas por meio das variáveis de composição da mistura, mesmo que a densidade de empacotamento de partículas ou a relação de vazios não tenham sido determinadas explicitamente. A forte concordância entre a importância das variáveis obtida por meio do SHAP e a teoria consolidada de empacotamento confirma a consistência física do modelo desenvolvido. Essa representação implícita oferece uma alternativa prática aos ensaios experimentais de empacotamento, que frequentemente demandam tempo e são específicos para cada material (Cheng et al., 2021; de Larrard, 1999). Sob uma perspectiva prática, os resultados sugerem que o controle do teor de agregado reciclado, da dosagem de aglomerante e da relação água/aglomerante é fundamental para otimizar a resistência à compressão do RA-SCC. A estrutura desenvolvida proporciona tanto capacidade preditiva quanto informações interpretativas, contribuindo para uma dosagem de concreto sustentável e orientada por dados.

6. CONCLUSÕES E DIREÇÕES PARA PESQUISAS FUTURAS

6.1 Conclusões.

Por meio da utilização de indicadores de empacotamento de partículas derivados das características de composição da mistura, este estudo desenvolveu uma estrutura de aprendizado de máquina explicável para a predição da resistência à compressão do concreto autoadensável com agregados reciclados (RA-SCC). Sob a perspectiva da engenharia, o modelo desenvolvido pode auxiliar os profissionais na seleção do teor de agregado reciclado e das proporções de aglomerantes, buscando equilibrar os requisitos de sustentabilidade e resistência, sem recorrer a ensaios laboratoriais baseados em tentativa e erro.

Com base em um conjunto de dados proveniente da literatura e em uma análise estatística abrangente, podem ser estabelecidas as seguintes conclusões:

1. Entre os seis modelos avaliados, o Regressor Extra Trees apresentou a maior precisão preditiva, alcançando R^2 de 0,9724, MAE de 0,79 MPa, RMSE de 1,19 MPa e MAPE de 2,12%. Esses resultados demonstram uma elevada concordância entre os valores experimentais e previstos da resistência à compressão.
2. Os modelos baseados em métodos de conjunto (*ensemble*), incluindo Extra Trees, Random Forest ($R^2 = 0,968$) e Light Gradient Boosting Machine ($R^2 = 0,965$), apresentaram desempenho consistentemente superior aos modelos de regressão linear e regressão Bayesiana. Essa vantagem de desempenho evidencia a capacidade dos modelos *ensemble* de capturar interações não lineares entre as variáveis de dosagem que governam

- o comportamento do empacotamento de partículas.
3. Os modelos linear e Bayesiano apresentaram menor precisão devido à multicolinearidade severa entre os parâmetros relacionados ao empacotamento de partículas. A análise do fator de inflação da variância (VIF) confirmou uma forte interdependência entre o teor de cinza volante, as frações de agregados e a relação água/aglomerante, enquanto a taxa de substituição por agregado reciclado permaneceu estatisticamente independente.
 4. A análise da confiabilidade estatística revelou que todos os modelos apresentaram viés de predição muito reduzido. A robustez da estrutura desenvolvida foi corroborada pelo erro médio de viés de 0,12 MPa do Regressor Extra Trees e pelos testes de hipóteses, que não identificaram diferença estatisticamente significativa entre as resistências previstas e medidas ($p > 0,05$).
 5. A análise por meio do método SHapley Additive exPlanations (SHAP) identificou a taxa de substituição por agregado reciclado como o parâmetro mais influente, seguida pelo teor de cimento e pela relação água/aglomerante. Essa classificação é fisicamente coerente com a teoria do empacotamento de partículas, uma vez que a qualidade dos agregados, a disponibilidade de pasta e a estrutura interna de vazios governam, em conjunto, o desenvolvimento da resistência à compressão.
 6. A significativa concordância entre a importância das variáveis determinada por SHAP e os princípios consolidados de empacotamento demonstra que o aprendizado de máquina pode capturar implicitamente os mecanismos de empacotamento de partículas por meio das variáveis de composição da mistura, mesmo que a densidade de empacotamento de partículas ou a relação de vazios não tenham sido estimadas explicitamente.
 7. A estrutura proposta contribui para o desenvolvimento sustentável do concreto e para a avaliação de seu desempenho, ao oferecer um método preciso, estatisticamente robusto e fisicamente interpretável para a predição da resistência à compressão do concreto autoadensável com agregados reciclados (RA-SCC). Os resultados demonstram que o aprendizado de máquina explicável pode atuar como uma ferramenta confiável de apoio à tomada de decisões para a dosagem sustentável do RA-SCC, tanto no âmbito da pesquisa quanto na prática profissional. A utilização de um conjunto de dados composto por 365 misturas, compiladas a partir de 102 estudos independentes, reforça ainda mais a confiabilidade estatística e a aplicabilidade prática da estrutura preditiva proposta.

6.2 Direções para pesquisas futuras.

Embora o presente trabalho demonstre que o aprendizado de máquina explicável é uma ferramenta útil para a predição da resistência à compressão, existem diversas áreas que podem ser objeto de investigações adicionais:

- Estudos futuros poderão incorporar a densidade de empacotamento de partículas ou a relação de vazios medidas experimentalmente como variáveis de entrada explícitas, a fim de fortalecer a relação mecânica entre as predições do aprendizado de máquina e o comportamento físico do empacotamento.
- Para ampliar a capacidade de generalização dos modelos em diferentes cenários de construção, o conjunto de dados poderá ser expandido para abranger uma faixa mais ampla de idades de cura, classes de agregados reciclados e características reológicas do concreto autoadensável
- Poderão ser investigadas abordagens de modelagem híbrida, combinando a teoria do empacotamento de partículas com estruturas de aprendizado de máquina orientadas por dados, com o objetivo de aprimorar ainda mais a precisão das predições e a interpretabilidade dos modelos.
- A extensão da metodologia proposta para a predição de propriedades relacionadas à durabilidade, como permeabilidade, retração e desenvolvimento da resistência em longo

prazo, ampliaria sua aplicabilidade no projeto de infraestruturas sustentáveis

- Por fim, modelos de aprendizado de máquina explicável podem ser associados a técnicas de quantificação de incertezas, permitindo estabelecer intervalos de confiança para as previsões e contribuindo para uma prática de engenharia fundamentada na avaliação de riscos.
- Pesquisas futuras devem se concentrar na validação da estrutura de aprendizado de máquina proposta utilizando conjuntos de dados experimentais controlados, obtidos em programas de pesquisa individuais. Esses estudos possibilitariam uma avaliação mais precisa do desempenho dos modelos sob condições consistentes de materiais e ensaios, além de contribuir para o desenvolvimento de ferramentas preditivas específicas para cada projeto.
- Além disso, estudos futuros devem incorporar conjuntos de dados abrangendo múltiplas idades de cura, possibilitando a predição da resistência em função do tempo. Esse aspecto é particularmente importante para sistemas de concreto que incorporam materiais cimentícios suplementares, como a cinza volante, nos quais o desenvolvimento da resistência em longo prazo se estende além do período convencional de 28 dias.

7. REFERÊNCIAS

- Ahmed, M., Khan, M. I., Al-Hussein, M. (2024). *Performance Evaluation of Ensemble Machine Learning Models for Compressive Strength Prediction of Green Concrete*. Journal of Building Engineering, 82, 108091. <https://doi.org/10.1016/j.jobbe.2024.108091>
- Ajdukiewicz, A., Kliszczewicz, A. (2002). Influence of recycled aggregates on mechanical properties of HS/HPC. *Cement and Concrete Composites*, 24(2), 269–279. [https://doi.org/10.1016/S0958-9465\(01\)00012-9](https://doi.org/10.1016/S0958-9465(01)00012-9)
- Akbarnezhad, A., Ong, K. C. G., Zhang, M. H., Tam, C. T., Foo, T. W. J. (2011). *Microwave-assisted beneficiation of recycled concrete aggregates*. Construction and Building Materials, 25(8), 3469–3479. <https://doi.org/10.1016/j.conbuildmat.2011.03.038>
- Almutairi, A. (2025). *Explainable machine learning-based prediction of compressive strength in sustainable recycled aggregate self-compacting concrete using SHAP analysis*. Sustainability, 17(24), 11334. <https://doi.org/10.3390/su172411334>
- Asteris, P. G., Skentou, A. D., Bardhan, A., Samui, P., Pilakoutas, K. (2020). *Predicting concrete strength using machine learning*. Construction and Building Materials, 252, 119134. <https://doi.org/10.1016/j.conbuildmat.2020.119134>
- Bardhan, A., Samui, P., Ghosh, K., Asteris, P. G. (2021). *Machine learning approaches for prediction of concrete compressive strength: A state-of-the-art review*. Archives of Computational Methods in Engineering, 28, 2495–2511. <https://doi.org/10.1007/s11831-020-09494-9>
- Behera, M., Bhattacharyya, S. K., Minocha, A. K., Deoliya, R., Maiti, S. (2014). *Recycled aggregate from C&D waste and its use in concrete – A breakthrough towards sustainability in construction sector*. Construction and Building Materials, 68, 501–516. <https://doi.org/10.1016/j.conbuildmat.2014.07.003>
- Breiman, L. (2001). *Random forests*. Machine Learning, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth International Group.
- Butler, L., West, J. S., Tighe, S. L. (2013). *Effect of recycled concrete coarse aggregate from multiple sources on the hardened properties of concrete with equivalent compressive strength*. Construction and Building Materials, 47, 1292–1301. <https://doi.org/10.1016/j.conbuildmat.2013.06.021>

- Cheng, Y.-H., Zhu, B.-L., Yang, S.-H., Tong, B.-Q. (2021). *Design of concrete mix proportion based on particle packing voidage and test research on compressive strength and elastic modulus of concrete*. Materials, 14(3), 623. <https://doi.org/10.3390/ma14030623>
- Corinaldesi, V. (2010). *Mechanical and elastic behaviour of concretes made of recycled-concrete coarse aggregates*. Construction and Building Materials, 24(9), 1616–1620. <https://doi.org/10.1016/j.conbuildmat.2010.02.031>
- de Brito, J., Saikia, N. (2013). *Recycled Aggregate in Concrete: Use of Industrial, Construction and Demolition Waste*. Springer. <https://doi.org/10.1007/978-1-4471-4540-0>
- de Larrard, F. (1999). *Concrete Mixture Proportioning: A Scientific Approach*. E & FN Spon.
- Domingo-Cabo, A., Lázaro, C., López-Gayarre, F., Serrano-López, M. A., Serna, P., Castaño-Tabares, J. O. (2009). *Creep and shrinkage of recycled aggregate concrete*. Construction and Building Materials, 23(7), 2545–2553. <https://doi.org/10.1016/j.conbuildmat.2009.02.018>
- Etcheberria, M., Vázquez, E., Marí, A., Barra, M. (2007). *Influence of amount of recycled coarse aggregates and production process on properties of recycled aggregate concrete*. Cement and Concrete Research, 37(5), 735–742. <https://doi.org/10.1016/j.cemconres.2007.02.002>
- Evangelista, L., de Brito, J. (2007). *Mechanical behaviour of concrete made with fine recycled concrete aggregates*. Cement and Concrete Composites, 29(5), 397–401. <https://doi.org/10.1016/j.cemconcomp.2006.12.004>
- Fanijo, E. O., Aladegboye, O. J., Oyeyemi, K. D. (2023). *Explainable machine learning prediction of compressive strength of sustainable concrete incorporating waste materials*. Journal of Building Engineering, 63, 105458. <https://doi.org/10.1016/j.jobbe.2022.105458>
- Felekoğlu, B., Türkel, S., Baradan, B. (2007). *Effect of water/cement ratio on the fresh and hardened properties of self-compacting concrete*. Building and Environment, 42(4), 1795–1802. <https://doi.org/10.1016/j.buildenv.2006.01.012>
- Funk, J. E., Dinger, D. R. (1994). *Predictive Process Control of Crowded Particulate Suspensions*. Springer.
- Guo, H., Shi, C., Guan, X., Zhu, J., Ding, Y., Ling, T. C., Zhang, H. (2018). *Durability of recycled aggregate concrete: A review*. Cement and Concrete Composites, 89, 251–259. <https://doi.org/10.1016/j.cemconcomp.2018.03.008>
- He, Z., Du, S., Chen, E., Zhang, J. (2022). *Particle Packing–Based Optimization of Sustainable Concrete Incorporating Recycled Aggregates*. Journal of Cleaner Production, 365, 132772. <https://doi.org/10.1016/j.jclepro.2022.132772>
- Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., Liu, T.-Y. (2017). *LightGBM: A highly efficient gradient boosting decision tree*. Advances in Neural Information Processing Systems, 30, 3146–3154.
- Khayat, K. H. (2016). *Workability, testing, and performance of self-consolidating concrete*. ACI Materials Journal, 113(2), 185–194. <https://doi.org/10.14359/51688664>
- Kim, T., Lee, S. (2023). *Impact of Multicollinearity on Machine Learning Models for Predicting Concrete Strength*. Engineering Structures, 287, 116041. <https://doi.org/10.1016/j.engstruct.2023.116041>
- Kou, S. C., Poon, C. S. (2009). *Properties of self-compacting concrete prepared with coarse and fine recycled concrete aggregates*. Cement and Concrete Composites, 31(9), 622–627. <https://doi.org/10.1016/j.cemconcomp.2009.06.005>
- Kou, S. C., Poon, C. S. (2013). *Long-term mechanical and durability properties of recycled aggregate concrete prepared with the incorporation of fly ash*. Cement and Concrete Composites, 37, 12–19. <https://doi.org/10.1016/j.cemconcomp.2012.12.011>
- Kutner, M. H., Nachtsheim, C. J., Neter, J., Li, W. (2005). *Applied Linear Statistical Models* (5th ed.). McGraw-Hill Irwin. <https://doi.org/10.2307/2238088>
- Li, L., Wang, Y., Zhang, J., Li, Z. (2022). *Effects of multicollinearity on machine learning prediction of concrete strength incorporating recycled aggregates*. Engineering Structures,

- 252, 113620. <https://doi.org/10.1016/j.engstruct.2021.113620>
- Limbachiya, M. C., Leelawat, T., Dhir, R. K. (2000). *Use of Recycled Concrete Aggregate in High-Strength Concrete*. *Materials and Structures*, 33(9), 574–580. <https://doi.org/10.1007/BF02480538>
- Liu, H., Wu, J., Chen, B. (2021). *Prediction of compressive strength of recycled aggregate concrete using ensemble machine learning techniques*. *Construction and Building Materials*, 304, 124609. <https://doi.org/10.1016/j.conbuildmat.2021.124609>
- Lundberg, S. M., Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*. *Advances in Neural Information Processing Systems*, 30, 4765–4774.
- McKinney, W. (2010). *Data Structures for Statistical Computing in Python*. *Proceedings of the 9th Python in Science Conference*, 51–56. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Montgomery, D. C., Peck, E. A., Vining, G. G. (2012). *Introduction to Linear Regression Analysis* (5th ed.). Wiley.
- Oikonomou, N. D. (2005). *Recycled Concrete Aggregates*. *Cement and Concrete Composites*, 27(2), 315–318. <https://doi.org/10.1016/j.cemconcomp.2004.02.020>
- Okamura, H., Ouchi, M. (2003). *Self-Compacting Concrete*. *Journal of Advanced Concrete Technology*, 1(1), 5–15. <https://doi.org/10.3151/jact.1.5>
- Park, S., Choi, S., Kim, Y. (2022). *Prediction of Compressive Strength of Recycled Aggregate Concrete Using Ensemble Machine Learning Techniques*. *Construction and Building Materials*, 314, 125634. <https://doi.org/10.1016/j.conbuildmat.2021.125634>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Duchesnay, E. (2011). *Scikit-Learn: Machine Learning in Python*. *Journal of Machine Learning Research*, 12, 2825–2830. <https://doi.org/10.5555/1953048.2078195>
- Pedro, D., de Brito, J., Evangelista, L. (2017). *Influence of the Use of Recycled Concrete Aggregates on the Compressive Strength of Concrete*. *Construction and Building Materials*, 133, 531–542. <https://doi.org/10.1016/j.conbuildmat.2016.12.093>
- Poon, C. S., Shui, Z. H., Lam, L. (2004). *Effect of Microstructure of ITZ on Compressive Strength of Concrete Prepared with Recycled Aggregates*. *Construction and Building Materials*, 18(6), 461–468. <https://doi.org/10.1016/j.conbuildmat.2004.03.005>
- Silva, R. V., de Brito, J., Dhir, R. K. (2015). *Properties and Composition of Recycled Aggregates from Construction and Demolition Waste Suitable for Concrete Production*. *Construction and Building Materials*, 65, 201–217. <https://doi.org/10.1016/j.conbuildmat.2014.04.117>
- Su, N., Hsu, K.-C., Chai, H.-W. (2001). *A Simple Mix Design Method for Self-Compacting Concrete*. *Cement and Concrete Research*, 31(12), 1799–1807. [https://doi.org/10.1016/S0008-8846\(01\)00566-X](https://doi.org/10.1016/S0008-8846(01)00566-X)
- Tam, V. W. Y., Gao, X. F., Tam, C. M. (2005). *Microstructural Analysis of Recycled Aggregate Concrete Produced from Two-Stage Mixing Approach*. *Cement and Concrete Research*, 35(6), 1195–1203. <https://doi.org/10.1016/j.cemconres.2004.10.025>
- Tam, V. W. Y., Tam, C. M., Le, K. N. (2007). *Removal of Cement Mortar Remains from Recycled Aggregate Using Pre-Soaking Approaches*. *Resources, Conservation and Recycling*, 50(1), 82–101. <https://doi.org/10.1016/j.resconrec.2006.05.012>
- Tang, Z., Wang, Y., Li, L., Zhang, J. (2023). *Machine Learning Techniques for Prediction of Mechanical Properties of Recycled Aggregate Concrete: A State-of-the-Art Review*. *Construction and Building Materials*, 366, 130203. <https://doi.org/10.1016/j.conbuildmat.2023.130203>
- Willmott, C. J., Matsuura, K. (2005). *Advantages of the Mean Absolute Error over the Root Mean Square Error in Assessing Average Model Performance*. *Climate Research*, 30(1), 79–82. <https://doi.org/10.3354/cr030079>
- Xiao, J., Li, W., Fan, Y., Huang, X. (2012). *An Overview of Study on Recycled Aggregate*

- Concrete in China (1996–2011)*. Construction and Building Materials, 31, 364–383. <https://doi.org/10.1016/j.conbuildmat.2011.12.074>
- Yeh, I.-C. (1998). *Modeling of Strength of High-Performance Concrete Using Artificial Neural Networks*. Cement and Concrete Research, 28(12), 1797–1808. [https://doi.org/10.1016/S0008-8846\(98\)00165-3](https://doi.org/10.1016/S0008-8846(98)00165-3)
- Zhang, H., Shi, C., Ding, Y., Guo, H. (2022). *Interfacial Transition Zone Characteristics of Recycled Aggregate Concrete and Their Influence on Mechanical Performance*. Cement and Concrete Research, 154, 106730. <https://doi.org/10.1016/j.cemconres.2022.106730>
- Zhang, J., Li, J., Zhao, Y. (2019). *Prediction of Recycled Aggregate Concrete Strength Using Machine Learning Approaches*. Journal of Cleaner Production, 241, 118218. <https://doi.org/10.1016/j.jclepro.2019.118218>
- Zhou, S., Chen, Y., Zhang, R., Shi, C. (2024). *Explainable Artificial Intelligence for Performance Prediction of Cement-Based Materials: Progress and Challenges*. Cement and Concrete Composites, 147, 105215. <https://doi.org/10.1016/j.cemconcomp.2023.105215>