

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

P. Kumar Yadav^{1*} , A. Kumar² 

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

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

Recibido: 20/02/2026 | Correcciones recibidas: 03/05/2026 | Aceptado: 08/05/2026 | Publicado: 01/09/2026

RESUMEN

Este estudio tiene como objetivo desarrollar un marco explicable de aprendizaje automático para predecir la resistencia a la compresión a 28 días de hormigón autocompactante con agregados reciclados (RA-SCC), utilizando indicadores de empaquetado de partículas. Se analizó un conjunto de datos compuesto por 365 mezclas de 102 estudios utilizando seis modelos de regresión, aplicando el método SHapley Additives ExPlanations (SHAP) para la interpretación de los resultados. El modelo Extra Trees mostró la mayor precisión ($R^2 = 0,9724$; Error cuadrático medio de raíz (RMSE) = 1,19 MPa). El conjunto de datos está limitado a la intensidad de 28 días y no considera los efectos de los aditivos. El estudio presenta un enfoque combinado que integra conceptos de empaquetado de partículas, evaluación multicolineal e inteligencia artificial explicable. Los resultados demuestran que el aprendizaje automático puede predecir de forma fiable la resistencia mientras captura implícitamente mecanismos relacionados con el empaquetado de partículas.

Palabras clave: hormigón autocompactante con agregados reciclados; aprendizaje automático; predicción de resistencia a la compresión; indicadores de empaquetado de partículas; inteligencia artificial explicable; análisis SHAP; multicolinealidad; hormigón sostenible.

Citar como: Kumar Yadav, P., Kumar, A. (2026), “Predicción explicable basada en aprendizaje automático de la resistencia a la compresión a 28 días de hormigón autocompactante con agregados reciclados utilizando indicadores de empaquetado 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.

Contribución de cada autor

En este trabajo, la contribución de los autores se dividió a partes iguales entre Pintu Kumar Yadav (50%) y el Dr. Ajay Kumar (50%).

Licencia Creative Commons

Los derechos de autor (2026) son propiedad de los autores. Este trabajo es un artículo de acceso abierto publicado bajo los términos y condiciones de una licencia internacional Creative Commons Attribution 4.0 International License ([CC BY 4.0](https://creativecommons.org/licenses/by/4.0/)).

Discusiones y correcciones posteriores a la publicación

Cualquier discusión, incluyendo la réplica de los autores, se publicará en el segundo número del año 2027 siempre y cuando la información se reciba antes del cierre del primer número del año 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.

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.

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.

Información Legal

Revista ALCONPAT es una publicación cuatrimestral de la Asociación Latinoamericana de Control de Calidad, Patología y Recuperación de la Construcción, Internacional, A. C., Km. 6, antigua carretera a Progreso, Mérida, Yucatán, C.P. 97310, Tel. +52 1 983 419 8241, alconpat.int@gmail.com, Página Web: www.alconpat.org

Reserva de derechos al uso exclusivo No.04-2013-011717330300-203, eISSN 2007-6835, ambos otorgados por el Instituto Nacional de Derecho de Autor. Editor responsable: Dr. Pedro Castro Borges. Responsable de la última actualización de este número, Unidad de Informática ALCONPAT, Ing. Elizabeth Sabido Maldonado.

Las opiniones expresadas por los autores no necesariamente reflejan la postura del editor.

La reproducción total o parcial de los contenidos e imágenes de la publicación se realiza en apego al código COPE y a la licencia CC BY 4.0 de la Revista ALCONPAT.

1. INTRODUCCIÓN

La investigación sobre el hormigón de agregados reciclados (RAC) como posible sustituto del hormigón convencional ha aumentado debido a la creciente demanda de materiales de construcción ambientalmente sostenibles. El uso de áridos reciclados reduce el consumo de recursos naturales y la generación de residuos derivados de la construcción y demolición, que representan una carga ambiental importante a nivel global (Behera et al., 2014; Evangelista & de Brito, 2007; Silva et al., 2015). Debido a su excelente fluidez, pasabilidad y resistencia a la segregación, el hormigón autocompactante (SCC) se ha convertido en un material de construcción avanzado, capaz de aplicarse eficientemente a estructuras densamente reforzadas, sin necesidad de vibraciones mecánicas (Khayat, 2016; Okamura & Ouchi, 2003). La incorporación de áridos reciclados en el SCC ofrece beneficios significativos en términos de sostenibilidad; sin embargo, también introduce una marcada variabilidad en las propiedades en estado fresco y endurecido, especialmente en la resistencia a la compresión (Kou & Poon, 2009; Poon et al., 2004).

La resistencia a la compresión sigue siendo uno de los principales parámetros de diseño y control de calidad en la ingeniería del hormigón, influyendo directamente en la seguridad estructural y el rendimiento en servicio. Una compleja interacción entre las propiedades del material, como la tasa de sustitución por agregados reciclados, la composición del aglutinante, la relación agua/aglutinante, la distribución del tamaño de las partículas de los agregados y la dosificación de aditivos, controla el desarrollo de la resistencia del hormigón autocompactante con agregados reciclados (RA-SCC) (Guo et al., 2018; Tam et al., 2007; Xiao et al., 2012). La presencia de mortero adherido a agregados reciclados altera la morfología de partículas, la absorción de agua y el comportamiento de la zona de transición interfacial, afectando en consecuencia la eficiencia del empaquetado de partículas y la estructura interna de los poros (Butler et al., 2013; Etxeberria et al., 2007). Estos efectos combinados hacen especialmente difícil predecir la resistencia de forma fiable mediante formulaciones empíricas convencionales.

La teoría del empaquetamiento de partículas proporciona una base mecanicista para entender la evolución de la resistencia en materiales cementicios. El empaquetamiento denso de componentes sólidos minimiza los vacíos, aumenta el contacto entre partículas y mejora la transferencia de esfuerzos dentro de la matriz endurecida (de Larrard, 1999; Funk & Dinger, 1994). Según investigaciones experimentales, el empaquetado optimizado de agregados puede aumentar significativamente la durabilidad, la resistencia a la compresión y el módulo de elasticidad, al tiempo que reduce la demanda de aglutinante (Cheng et al., 2021; He et al., 2022; Su et al., 2001). En los sistemas de hormigón autocompactante (SCC), la densidad de empaquetado es especialmente importante, ya que la trabajabilidad y la resistencia a la segregación están estrechamente relacionadas con el equilibrio volumétrico entre áridos, materiales finos y suspensión (Felekoğlu et al., 2007; Khayat, 2016). A pesar de estos avances, la mayoría de los enfoques de empaquetado de partículas dependen de procedimientos experimentales laboriosos y presentan dificultades para generalizar a diferentes fuentes de materiales y dosis concretas.

Simultáneamente, el uso de técnicas de aprendizaje automático (ML) para modelar interacciones complejas y no lineales en materiales basados en hormigón se ha extendido cada vez más (Tang et al., 2023). Las redes neuronales artificiales, la regresión de vectores de soporte, los algoritmos de Random Forest y Gradient Boosting se han utilizado con éxito en varios estudios para predecir la resistencia a la compresión con mayor precisión que los modelos convencionales de regresión (Asteris et al., 2020; Liu et al., 2021; Yeh, 1998). La efectividad de las técnicas de aprendizaje en conjunto para tratar la variabilidad de materiales e interacciones no lineales ha sido validada por estudios recientes centrados en hormigón con

agregados reciclados (RAC) (Li et al., 2022; Park et al., 2022). Sin embargo, muchos modelos de aprendizaje automático se consideran "cajas negras", lo que limita su interpretabilidad y reduce la confianza en su uso en los procesos de toma de decisiones de ingeniería (Bardhan et al., 2021; Zhou et al., 2024).

Se han desarrollado métodos explicables de inteligencia artificial, incluyendo SHapley Additive ExPlanations (SHAP), para cuantificar el impacto de variables de entrada individuales en las predicciones de modelos de aprendizaje automático, con el objetivo de superar esta limitación (Lundberg & Lee, 2017). La predicción de la resistencia del hormigón se ha beneficiado del uso efectivo de la interpretación basada en SHAP, que proporciona información físicamente significativa sobre el papel de los componentes de la mezcla como cemento, contenido de agua y materiales cementicios suplementarios (Almutairi, 2025; Fanijo et al., 2023). Aunque la teoría del empaquetamiento de partículas ha sido ampliamente investigada en el contexto del agrupamiento por lotes de hormigón, su integración con enfoques explicables de aprendizaje automático aplicados al hormigón autocompactante con agregados reciclados sigue siendo relativamente limitada.

Además, las variables concretas de dosificación asociadas al empaquetado de partículas suelen ser fuertemente interdependientes. Parámetros como las fracciones agregadas, el contenido de finos y la relación agua-aglutinante presentan una multicolinealidad significativa, lo que compromete la fiabilidad de los modelos estadísticos lineales y dificulta la interpretación física de los resultados (Kim & Lee, 2023; Li et al., 2022; J. Zhang et al., 2019).

Aunque la multicolinealidad tiene implicaciones significativas para la selección e interpretabilidad del modelo, rara vez se cuantifica explícitamente en estudios concretos basados en aprendizaje automático (ML). Sin embargo, la investigación publicada actualmente aún no presenta un enfoque integral para predecir la resistencia del hormigón autocompactante con agregados reciclados (RA-SCC) que considere simultáneamente la precisión de la predicción, la multicolinealidad y la interpretabilidad física.

Por ello, el presente estudio incorpora indicadores de empaquetamiento de partículas derivados de las características de composición de la mezcla en una estructura basada en aprendizaje automático, orientados a predecir la resistencia a la compresión del hormigón autocompactante con agregados reciclados (RA-SCC). A diferencia de los estudios que calculan explícitamente la densidad de empaquetado o el índice de vacío, este trabajo captura implícitamente los mecanismos relacionados con el empaquetado mediante modelado basado en datos e interpreta su influencia mediante análisis SHAP. Además, el análisis del factor de inflación de varianza (VIF) se emplea para cuantificar la multicolinealidad entre las variables relacionadas con el empaquetado, proporcionando una base estadísticamente robusta para la selección de los modelos. Al integrar precisión predictiva, validación estadística e interpretabilidad física, este estudio busca contribuir con una herramienta fiable y explicable para la dosificación sostenible de RA-SCC. Se reconoce que el comportamiento del empaquetado de partículas en el hormigón está gobernado por múltiples variables de dosificación que interactúan entre sí y pueden no estar completamente representadas solo por índices individuales de empaquetado. Varios estudios previos han investigado el efecto de la tasa de sustitución por áridos reciclados sobre la resistencia a la compresión, estableciéndola como un parámetro clave para el diseño de hormigones sostenibles (Kou & Poon, 2009; Pedro et al., 2017; Xiao et al., 2012). Estos estudios destacan de manera constante la influencia del nivel de sustitución en la reducción y variabilidad de resistencia, proporcionando una base para su inclusión como variable de entrada primaria en los modelos predictivos.

2. OBJETIVOS DE INVESTIGACIÓN

El objetivo principal de este estudio es desarrollar una estructura basada en aprendizaje automático capaz de predecir, con precisión e interpretabilidad, la resistencia a la compresión del hormigón autocompactante con agregados reciclados (RA-SCC), mediante la incorporación de indicadores de empaquetado de partículas derivados de características relacionadas con la composición de la mezcla. Para lograr este objetivo, se definen los siguientes objetivos específicos:

1. Compilar y analizar un conjunto de datos basado en la literatura de RA-SCC, incorporando variables de dosificación asociadas al comportamiento de empaquetado de partículas, incluyendo proporciones agregadas, contenido de aglutinante, relación agua/aglutinante, índices volumétricos y nivel de sustitución por agregados reciclados.
2. Crear y evaluar varios modelos de regresión basados en aprendizaje automático para la predicción de la resistencia a la compresión RA-SCC, incluyendo algoritmos *en conjunto*, basados en árboles y lineales.
3. Evaluar el rendimiento de los modelos utilizando indicadores estadísticos integrales, incluyendo coeficientes de determinación, métricas basadas en errores, sesgo de error medio y pruebas de hipótesis, para evaluar la precisión predictiva y la fiabilidad estadística.
4. Utilizar el análisis del factor de inflación de varianza (VIF) para cuantificar el grado de multicolinealidad entre las variables de entrada relacionadas con el empaquetado de partículas e investigar sus implicaciones para la estabilidad de las predicciones y la selección de los modelos.
5. Relacionar los resultados del aprendizaje automático con el comportamiento físico de los materiales, interpretando el impacto de los indicadores de empaquetado de partículas en la predicción de la resistencia a la compresión mediante el método SHapley Additive ExPlanations.
6. Demostrar la capacidad del aprendizaje automático explicable para capturar implícitamente los mecanismos de empaquetamiento de partículas, sin el cálculo explícito de densidad de empaquetamiento o índice de vacío, contribuyendo a una dosis sostenible basada en datos de RA-SCC.

3. MATERIALES Y DESCRIPCIÓN DEL CONJUNTO DE DATOS

3.1 Fuente de datos y estrategia de compilación.

Los resultados experimentales publicados en la literatura científica revisada por pares sobre hormigón con áridos reciclados y hormigón autocompactante con áridos reciclados se compilaron sistemáticamente para constituir el conjunto de datos utilizado en esta investigación. Solo se consideraron estudios que presentaron proporciones de mezcla completas y claramente definidas, características de materiales y valores de resistencia a la compresión. Este enfoque aseguró la fiabilidad de los datos, permitiendo al tiempo que se cubrieron una amplia variedad de combinaciones de materiales y niveles de rendimiento. El conjunto final de datos consta de $N = 365$ mezclas de hormigón, compiladas sistemáticamente a partir de 102 estudios científicos independientes revisados por pares. Cada estudio seleccionado proporcionó información completa y consistente sobre las proporciones de mezcla y la resistencia a la compresión, asegurando la calidad y comparabilidad de los datos entre diferentes programas experimentales. El conjunto de datos abarca una amplia gama de niveles de sustitución por agregados reciclados, composiciones de aglutinantes y configuraciones de mezcla, representando así una variabilidad realista de los materiales. Esta diversidad aumenta la

representatividad estadística y la robustez del marco de aprendizaje automático desarrollado en este estudio. El conjunto de datos está compuesto exclusivamente por valores de resistencia a la compresión obtenidos experimentalmente y reportados en los estudios fuente. No se utilizaron datos sintéticos ni simulados. Además de la división de los datos en conjuntos de entrenamiento y pruebas adoptada en este estudio, se reconoce que el uso de conjuntos de datos experimentales externos para validación independiente constituye un paso importante para futuras investigaciones, permitiendo una evaluación más completa de la generalizabilidad del modelo a sistemas materiales aún no analizados.

Se reconoce que los modelos de aprendizaje automático desarrollados con datos de un único programa experimental bien controlado pueden tener mayor consistencia interna. Sin embargo, el uso, en este estudio, de un conjunto de datos obtenido de la literatura permite incorporar una mayor variedad de propiedades del material, proporciones de mezclas y condiciones experimentales. Esta diversidad aumenta la generalizabilidad del modelo desarrollado y le permite capturar la variabilidad real encontrada en aplicaciones prácticas. Considerando la complejidad inherente y la interdependencia de las variables del hormigón autocompactante con los agregados reciclados, el uso de múltiples fuentes independientes de datos proporciona una representación más completa del comportamiento del material. El enfoque adoptado, por tanto, busca establecer un marco predictivo generalizado, en lugar de un modelo específico para un proyecto dado, calibrado a partir de un único conjunto de datos experimentales. Los estudios seleccionados cubren diferentes fuentes de agregados reciclados, tasas de sustitución, sistemas de aglutinantes y condiciones de curado, proporcionando suficiente variabilidad para un análisis robusto mediante aprendizaje automático (Ajdukiewicz & Kliszczewicz, 2002; Limbachiya et al., 2000; Xiao et al., 2012).

Las fuentes clave incluyen estudios experimentales sobre el comportamiento mecánico del hormigón con agregados reciclados, la influencia del nivel de sustitución de agregados y el papel de materiales cementicios suplementarios como la ceniza volante. En conjunto, estos estudios representan tanto sistemas convencionales de hormigón como de alto rendimiento y reflejan la variabilidad real de los materiales observada en la práctica (Corinaldesi, 2010; Kou & Poon, 2013; Pedro et al., 2017). También se utilizaron revisiones y monografías exhaustivas sobre hormigón con agregados reciclados para asegurar la consistencia en las definiciones de materiales y el alcance experimental (de Brito & Saikia, 2013; Oikonomou, 2005). Una lista detallada de las referencias utilizadas para la compilación del conjunto de datos se presenta en la sección de Referencias. En conjunto, estas fuentes cubren una amplia gama de contenidos de agregados reciclados, sistemas de aglutinantes y condiciones de prueba, asegurando que el conjunto de datos represente adecuadamente la variabilidad experimental real reportada en la literatura.

3.2 Materiales constituyentes.

3.2.1 Agregados reciclados.

Los agregados reciclados utilizados en los estudios compilados se obtuvieron principalmente de residuos de hormigón triturados procedentes de elementos estructurales demolidos. Estos agregados generalmente están compuestos por partículas naturales cubiertas por mortero de cemento adherido, que altera la textura superficial, la porosidad y las características de absorción de agua en comparación con los agregados naturales (Etxeberria et al., 2007; Poon et al., 2004). Varios estudios han informado del uso de métodos de pretratamiento o beneficio, como el procesamiento mecánico o técnicas asistidas por microondas, con el objetivo de mejorar la calidad de los agregados reciclados. Sin embargo, los agregados sin tratar se utilizaron con mayor frecuencia para representar condiciones prácticas para su aplicación en la construcción civil (Akbarnezhad et al., 2011; Tam et al., 2005). La tasa de sustitución por agregados reciclados varió entre estudios, oscilando desde la sustitución parcial hasta total de

los áridos gruesos naturales. Esta variación es esencial para evaluar la influencia del contenido de agregados reciclados en la eficiencia del empaquetado de partículas y el desarrollo de la resistencia a la compresión (Evangelista & de Brito, 2007; Pedro et al., 2017). Estos efectos están ampliamente documentados en estudios previos que analizaron la influencia de los áridos reciclados en el rendimiento mecánico y las características microestructurales del hormigón (Butler et al., 2013; Etxeberria et al., 2007; Poon et al., 2004).

3.2.2 Materiales cementicios.

El cemento Portland común (OPC) fue el principal material aglutinante utilizado en los estudios experimentales empleados para compilar el conjunto de datos. En el conjunto de datos compilado, el término OPC se refiere a los aglutinantes cementantes cementados informados en los estudios originales, que pueden incluir tanto cemento Portland puro (CEM I) como cementos compuestos (como el CEM II), dependiendo de la fuente. Debido a las variaciones en las prácticas de presentación de la información encontradas en la literatura, no siempre fue posible distinguir explícitamente estos tipos de cemento en todas las entradas del conjunto de datos. El OPC desempeña un papel clave en el proceso de hidratación, la formación de lodos y el desarrollo de la resistencia en las primeras edades del RA-SCC. Las variaciones en el contenido de cemento influyen directamente en el volumen de lodo, el grado de recubrimiento de partículas y la capacidad de transferencia de esfuerzos en la matriz endurecida, factores directamente relacionados con el rendimiento de la resistencia a la compresión.

Varios estudios han incorporado materiales cementicios suplementarios, especialmente ceniza volante, como sustituto parcial del cemento. Se ha informado ampliamente que la incorporación de ceniza volante mejora el rendimiento mecánico a largo plazo y la durabilidad del hormigón con agregados reciclados, mediante el refinamiento de la estructura de los poros y la mejora de la zona de transición interfacial entre las partículas de agregado y la suspensión de cemento (Behera et al., 2014; Kou & Poon, 2013). La morfología esférica y el tamaño fino de grano de las partículas de ceniza volante contribuyen a mejorar el empaquetado de partículas al rellenar los microvacíos entre las partículas más gruesas, aumentando así la densidad de empaquetado y reduciendo la porosidad capilar.

Desde la perspectiva del empaquetado de partículas, el cemento y la ceniza volante controlan conjuntamente la disponibilidad de la suspensión y la distribución de las partículas finas en el sistema de hormigón. Un volumen adecuado de suspensión es esencial en el hormigón autocompactante para garantizar una fluidez adecuada, resistencia a la segregación y la completa implicación de partículas de agregados.

En el hormigón autocompactante con áridos reciclados (RA-SCC), este requisito se vuelve aún más crítico debido a la mayor rugosidad superficial y mayor absorción de agua de los agregados reciclados. La ceniza volante contribuye a una mejor trabajabilidad y a la reducción de la demanda de agua al influir indirectamente en la eficiencia del empaquetado y en la estructura de vacíos internos (de Brito & Saikia, 2013). Estos resultados son coherentes con la literatura consolidada sobre el papel de los materiales cementicios suplementarios en la mejora de la durabilidad y el refinamiento microestructural de sistemas de hormigón con agregados reciclados (Behera et al., 2014; Kou & Poon, 2013; Silva et al., 2015).

Cabe señalar que en varios estudios utilizados como fuente, los sistemas de cemento compuesto (por ejemplo, CEM II) pueden ya contener materiales cementicios suplementarios, como la ceniza volante. Cuando estos cementos se reportan junto con contenidos adicionales de ceniza volante, la contribución efectiva de este material puede sobreestimarse. Esto introduce cierto grado de sesgo inherente al conjunto de datos, resultando en altos valores aparentes de ceniza volante y una fuerte interdependencia estadística entre variables relacionadas con los aglutinantes. Este efecto contribuye directamente al extremadamente alto VIF observado para el contenido de ceniza volante, reflejando una severa multicolinealidad entre componentes

cementicios. Sin embargo, esta interdependencia es característica de conjuntos de datos concretos obtenidos de la literatura, en los que la composición de los aglutinantes está determinada por restricciones de dosificación prácticas en lugar de por variables controladas de forma independiente.

El uso, en este estudio, de modelos de aprendizaje automático en conjunto basados en árboles reduce esta limitación, ya que estos algoritmos son intrínsecamente robustos a la multicolinealidad y no dependen de la hipótesis de independencia entre las variables predictoras. En el conjunto de datos compilado, los contenidos de cemento y ceniza volante mostraron una fuerte interdependencia con otros parámetros de dosificación, como la relación agua/aglutinante y las fracciones agregadas. Esta interacción destaca el papel predominante de estos componentes en la definición de la disposición microestructural de partículas sólidas y productos de hidratación. En consecuencia, el contenido de cemento y ceniza volante se consideró indicadores clave del empaquetado de partículas en el marco de aprendizaje automático, permitiendo que los modelos capturasen implícitamente los mecanismos relacionados con el empaquetado que influyen en el desarrollo de la resistencia a la compresión.

3.2.3 Agregados y amasado de agua.

Tanto los agregados gruesos como los finos se incluyeron en el conjunto de datos, y el agregado fino generalmente consistía en arena natural de río. La distribución del tamaño de partículas y las proporciones volumétricas de los agregados se informaron explícitamente en la mayoría de los estudios, lo que permitió determinar fracciones y parámetros de fracción volumétrica (Domingo-Cabo et al., 2009; Limbachiya et al., 2000). El contenido de agua de amasa y la relación agua/aglutinante se extrajeron directamente de los estudios, ya que estos parámetros ejercen una fuerte influencia en la trabajabilidad, la estructura de vacíos y la resistencia a la compresión (Poon et al., 2004; Tam et al., 2005). En el hormigón autocompactante, los aditivos químicos, como los superplastificantes y los agentes modificadores de viscosidad, juegan un papel fundamental en la obtención de la fluidez, pasabilidad y resistencia a la segregación requeridas. Sin embargo, estos parámetros no se reportaron de forma consistente en las fuentes bibliográficas compiladas y, por tanto, no se incluyeron como variables de entrada en el marco de aprendizaje automático utilizado en el presente estudio. Esta omisión representa una limitación, ya que la dosis de aditivos influye significativamente tanto en las propiedades en estado fresco como endurecido del CCE.

En sistemas de hormigón autocompactantes, las propiedades en estado fresco, como la fluyabilidad, la superabilidad y la resistencia a la segregación, son criterios esenciales de rendimiento. Estas características se evalúan típicamente mediante pruebas estandarizadas, incluyendo flujo de hundimiento, *embudo en V* y *L-box*. Las variaciones en la composición de la mezcla, especialmente las proporciones de agregados, el contenido de aglutinante y la relación agua-aglutinante, influyen significativamente en estas propiedades.

Sin embargo, debido a la ausencia de datos sobre las propiedades en estado fresco reportados de forma consistente en las fuentes bibliográficas compiladas, el presente estudio se centra exclusivamente en la predicción de la resistencia a la compresión. Además, los parámetros fundamentales que diferencian el hormigón autocompactante del hormigón convencional, como la dosis de aditivos reductores de agua de alta eficiencia, el contenido de agentes modificadores de viscosidad y las características reológicas relacionadas con el contenido de vacíos, no estaban disponibles de forma consistente en el conjunto de datos compilado. En consecuencia, el modelo no diferencia explícitamente el comportamiento del CCE en función del rendimiento en estado fresco, sino que captura sus efectos indirectamente a través de las variables relacionadas con la composición de la mezcla.

3.3 Variables de entrada y salida.

Se elaboró un conjunto de variables de entrada representativas de la composición del material y las propiedades de empaquetado de partículas basándose en los resultados experimentales publicados. El conjunto de datos utilizado para el desarrollo del modelo consta de 365 muestras, proporcionando una base suficientemente amplia y diversa para un entrenamiento y validación fiables de modelos de aprendizaje automático. Estas variables incluyen contenido de cemento, contenido de cenizas volantes, contenido de agregados gruesos, proporción de agregados finos, porcentaje de sustitución por agregados reciclados, contenido de agua, relación agua/aglutinante, fracción volumétrica de agregados, índice de empaquetado y edad de curado. Aunque la densidad de empaquetado de partículas o el índice de vacíos no se calcularon explícitamente en los estudios utilizados como fuente, estas variables actúan juntas como indicadores del empaquetamiento de partículas, ya que gobiernan la distribución espacial de partículas sólidas y suspensión en la matriz de hormigón (Cheng et al., 2021; de Brito & Saikia, 2013)

La resistencia a la compresión del hormigón, determinada en edades de curado preestablecidas y bajo condiciones de laboratorio controladas, constituye la variable de salida. Los valores de resistencia a la compresión se reportaron de forma consistente en los estudios seleccionados, lo que permite una comparación y agregación significativa de los datos (Ajdukiewicz & Kliszczewicz, 2002; Corinaldesi, 2010). La resistencia a la compresión se consideró como parámetro de salida, mientras que un conjunto de factores relacionados con la composición del material y el empaquetado de partículas se seleccionó como variables de entrada del modelo, basándose en el conjunto de datos compilado de la literatura. La lista completa de variables de entrada y salida, incluyendo sus símbolos, unidades y funciones en el marco de aprendizaje automático, se resume en la Tabla 1. Se calcularon estadísticas descriptivas básicas para el conjunto de datos compilado, con el objetivo de proporcionar información sobre la distribución y variabilidad de las variables de entrada y salida utilizadas en el modelado de aprendizaje automático. Los valores mínimo, máximo, media, desviación estándar y varianza de cada variable se resumen en la Tabla 2.

Tabla 1. Descripción de las variables de entrada y salida utilizadas en el modelado de aprendizaje automático.

Variable	Símbolo	Unidad	Descripción	Papel en el modelo
Edad	Edad	Días	Periodo de curado de muestras de hormigón	Entrada
Cemento	Cemento	kg/m ³	Contenido común de cemento Portland en la mezcla	Entrada
Gris volante	FA	kg/m ³	Contenido de ceniza volante utilizado como material cementicio suplementario	Entrada
Agregado grueso	CA	kg/m ³	Contenido de áridos gruesos en la mezcla de hormigón	Entrada
Agua	Agua	kg/m ³	Contenido de agua para amasar	Entrada
Relación agua/aglutinante	W/B	–	Relación entre la cantidad total de agua de amasado y el contenido total de aglutinante	Entrada
Contenido agregado reciclado	AR (%)	%	Porcentaje de reemplazo de agregado natural por agregado	Entrada
Proporción de agregado fino	LEJOS	–	Proporción de agregado fino respecto al contenido agregado total	Entrada
Fracción volumétrica de agregados	AVF	–	Fracción volumétrica de los agregados en la mezcla de hormigón	Entrada
Índice de Embalaje	IP	–	Índice que representa la eficiencia del empaquetado de partículas	Entrada
Resistencia a la compresión	CS	MPa	Resistencia a la compresión medida del hormigón	Producción

Agua	Agua	kg/m ³	Contenido de agua para amasar	Entrada
Relación agua/aglutinante	W/B	–	Relación entre la cantidad total de agua de amasame y el contenido total de aglutinante	Entrada

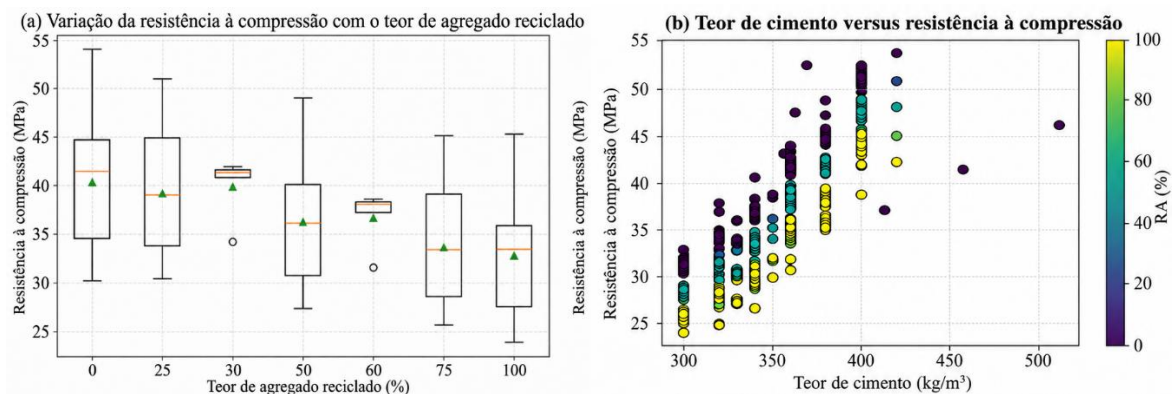
Nota: FAR indica la proporción de agregado fino, AVF indica la fracción volumétrica de agregados, PI indica el contenido de envasado y W/B indica la proporción agua/aglutinante.

Tabla 2. Resumen estadístico de las variables del conjunto de datos utilizado en el estudio.

Parámetro de entrada	Mínimo	Máximo	Promedio	Desviación estándar	Variable
Edad (días)	28	28	28	0	0
Cemento (kg/m ³)	300	511.9	352.249	33.001	1089.069
Ceniza volante (kg/m ³)	511	780	716.027	41.548	1726.209
Ceniza volante (kg/m ³)	980	1330	1069.904	46.106	2125.783
Agua (kg/m ³)	155.1	215	171.202	6.618	43.796
Relación agua/aglutinante (W/B)	0.38	0.58	0.491	0.052	0.003
Contenido agregado reciclado (%)	0	100	48.932	39.757	1580.586
Proporción de agregado fino (FAR)	0.31	0.407	0.401	0.011	0
Fracción volumétrica de agregados (AVF)	0.613	0.728	0.663	0.019	0
Índice de Embalaje (PI)	7.67	12.61	10.475	0.515	0.265
Resistencia a la compresión (MPa)	23.9	54.1	36.591	6.79	46.102

Nota: Es importante señalar que el conjunto de datos utilizado en este estudio está restringido a una única edad de curado de 28 días, por lo que no hay variabilidad en el parámetro de edad. En consecuencia, los modelos de aprendizaje automático desarrollados están específicamente destinados a predecir la resistencia a la compresión a 28 días de hormigón autocompactante con agregados reciclados, y no a representar la evolución de la resistencia a lo largo del tiempo. FAR indica la proporción de agregados finos, AVF indica la fracción volumétrica de agregados, PI indica el índice de empaquetado y W/B indica la relación agua/aglutinante.

Aunque la resistencia a la compresión a 28 días es ampliamente aceptada como referencia estándar para el diseño estructural y el control de calidad en la ingeniería del hormigón, esta limitación implica que el modelo propuesto no considera explícitamente los procesos de hidratación dependientes del tiempo. Se sabe que materiales cementicios suplementarios, como la ceniza volante, contribuyen al desarrollo de resistencia tras 28 días, proporcionando a menudo ganancias significativas de resistencia en edades de curado más avanzadas, como 56 o 90 días. Por ello, el presente estudio se centra en la modelización de la resistencia a la compresión en el periodo estándar de curado de 28 días, que sigue siendo el parámetro más frecuentemente reportado y de mayor relevancia práctica en los conjuntos de datos obtenidos de la literatura. La investigación futura debería incorporar conjuntos de datos con múltiples edades de curado, permitiendo un modelado dependiente del tiempo y una representación más adecuada de la cinética de hidratación a largo plazo de los sistemas cementicios compuestos.



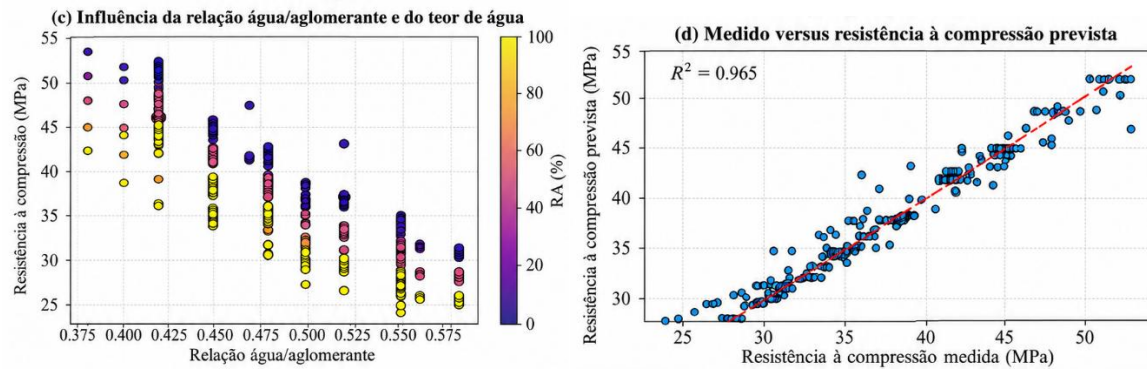


Figura 1. Caracterización del conjunto de datos, que ilustra (a) la variabilidad de la resistencia a la compresión en función del contenido de agregados reciclados, (b) la dependencia de la resistencia a la compresión respecto al contenido de cemento bajo diferentes niveles de agregado reciclado, (c) los efectos combinados de la relación agua/aglutinante y el contenido de agua sobre la resistencia a la compresión, y (d) medidas frente a la resistencia a la compresión predicha basándose en los parámetros de entrada combinados ($R^2 = 0,965$).

La Figura 1 presenta una evaluación visual concisa del conjunto de datos de hormigón autocompactante con agregados reciclados antes del modelado por aprendizaje automático. Los efectos del contenido de agregados reciclados, la dosificación del cemento y los parámetros relacionados con el agua sobre la resistencia a la compresión se presentan en las Figuras 1(a-c), que evidencian la variabilidad de los materiales y los efectos combinados de las variables de dosificación. Con un alto coeficiente de determinación, la Figura 1(d) demuestra una fuerte correlación entre los valores medidos y predichos de resistencia a la compresión, utilizando los parámetros combinados de entrada ($R^2 = 0,965$), confirmando que el conjunto de datos captura eficazmente los mecanismos clave que gobiernan el desarrollo de la resistencia.

3.4 Preprocesamiento y consistencia de los datos.

Para garantizar la compatibilidad entre los diferentes programas experimentales, todas las variables de entrada se convirtieron en unidades consistentes y se normalizaron antes del análisis mediante aprendizaje automático. Se excluyeron los registros de datos que presentaban información incompleta sobre las mezclas o condiciones de prueba ambiguas. El conjunto final presenta una distribución equilibrada de contenidos agregados reciclados, composiciones de aglutinantes y proporciones agua/aglutinante, permitiendo a los modelos de aprendizaje automático identificar tanto los efectos individuales como los comportamientos de interacción asociados al empaquetado de partículas. Este conjunto de datos basado en literatura mantiene una estrecha relación con el comportamiento observado experimentalmente de los sistemas de hormigón de agregados reciclados, a la vez que proporciona una base físicamente significativa para el modelado basado en datos.

4. METODOLOGÍA

El flujo de trabajo del marco propuesto de aprendizaje automático, desarrollado para predecir la resistencia a la compresión del hormigón autocompactante con agregados reciclados (RA-SCC), se presenta en la Figura 2. El proceso comienza con la compilación de un conjunto de datos basado en la literatura y continúa con el preprocesamiento, normalización y partición del conjunto de datos. Antes del desarrollo de los modelos, la multicolinealidad entre las variables relacionadas con el empaquetado de partículas se cuantifica mediante el análisis del factor de inflación de varianza (VIF). Después, se entrenan y evalúan varios modelos de regresión y de aprendizaje automático por *conjunto* utilizando métricas estadísticas de rendimiento y pruebas de hipótesis. Finalmente, se

emplea el método SHapley Additive ExPlanations (SHAP) para interpretar las predicciones de los modelos e identificar los indicadores de empaquetado de partículas más influyentes, asegurando tanto la precisión predictiva como la interpretabilidad física.

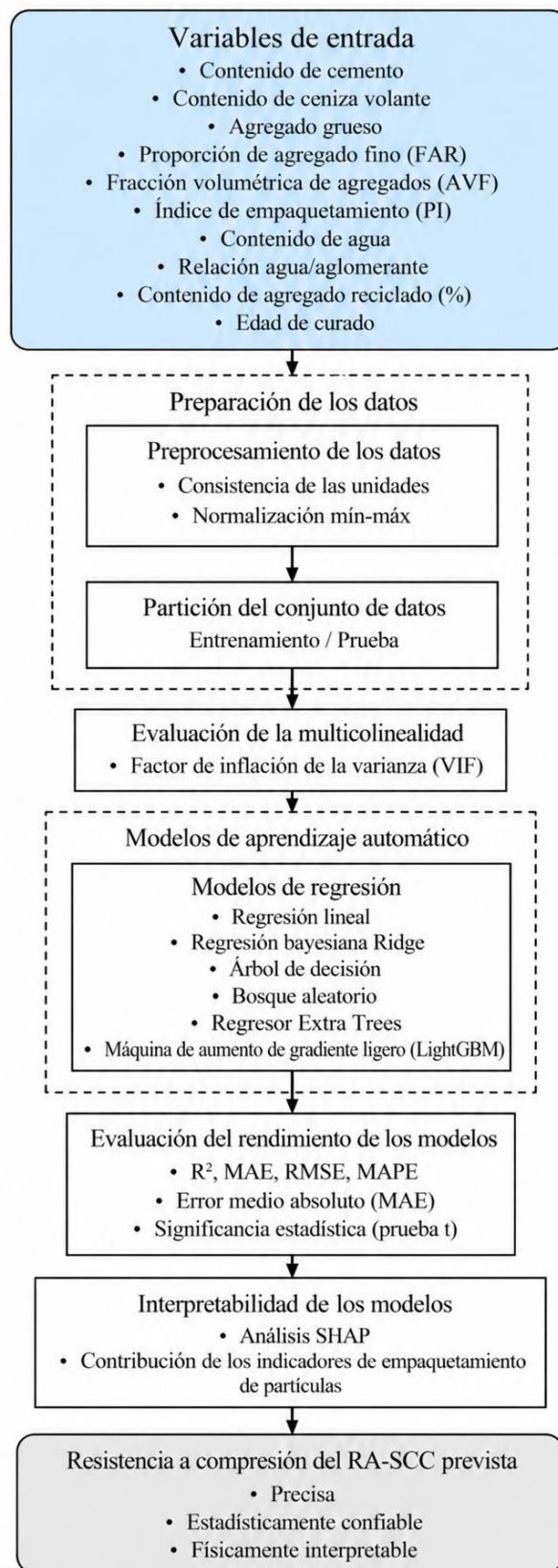


Figura 2. Flujo de trabajo del marco de aprendizaje automático propuesto.

4.1 Visión general de la metodología propuesta

La fuerte no linealidad, multicolinealidad y variabilidad de materiales son características inherentes del hormigón autocompactante con agregados reciclados (RA-SCC), y el enfoque científico utilizado en este estudio se desarrolló para abordar estas dificultades de forma integral. La metodología se dividió en cuatro pasos consecutivos e interconectados: preprocesamiento y normalización de datos, evaluación de la multicolinealidad, desarrollo de modelos de aprendizaje automático y evaluación del rendimiento con análisis explicable. Este enfoque se adoptó para garantizar tanto la precisión predictiva como la interpretabilidad desde una perspectiva de ingeniería.

En el primer paso, el conjunto de datos compilado a partir de la literatura fue sometido a preprocesamiento, para garantizar la consistencia, estabilidad numérica y compatibilidad entre las diferentes fuentes experimentales. Este paso es fundamental en la modelización basada en datos de propiedades concretas, ya que los datos experimentales en bruto suelen mostrar grandes variaciones en escala y distribución, lo que puede afectar negativamente al aprendizaje y convergencia de los modelos (Liu et al., 2021; Yeh, 1998). La normalización de las variables de entrada también permite una comparación adecuada entre los diferentes algoritmos de aprendizaje automático.

Utilizando el análisis del factor de inflación de varianza (VIF), el segundo paso se centró en cuantificar las interdependencias entre las variables de entrada. Los parámetros de dosificación asociados al empaquetado de partículas, como las fracciones agregadas, el contenido de aglutinante y la relación agua/aglutinante, suelen estar altamente correlacionados. La evaluación explícita de la multicolinealidad proporciona una base estadística para comprender las limitaciones de los modelos de regresión lineal y ayuda en la selección de enfoques de aprendizaje no paramétricos y *de conjunto*, que tienen mayor robustez frente a variables de entrada correlacionadas (Kutner et al., 2005; Li et al., 2022).

Para capturar las complejas interacciones no lineales entre los indicadores de empaquetado de partículas y la resistencia a la compresión, se desarrollaron varios modelos de regresión basados en aprendizaje automático en el tercer paso. Se utilizaron tanto métodos tradicionales de regresión como modelos avanzados de aprendizaje en conjunto basados en árboles, *lo que permitió comparar diferentes enfoques e investigar cómo la complejidad de los algoritmos influye en los resultados de predicción* (Asteris et al., 2020).

El paso final integró una evaluación integral del rendimiento con técnicas explicables de inteligencia artificial. Además de las métricas convencionales de precisión y error, se utilizaron evaluaciones estadísticas de sesgo y pruebas de hipótesis para analizar la fiabilidad de los modelos. Además, se empleó el método SHapley Additive ExPlanations (SHAP) para interpretar las predicciones de los modelos y cuantificar la contribución individual de los indicadores de empaquetado de partículas, relacionando así los resultados basados en datos con el comportamiento físicamente significativo de los materiales (Almutairi, 2025; Lundberg & Lee, 2017). Esta metodología integrada garantiza que el marco desarrollado no solo sea preciso, sino también transparente y adecuado para aplicaciones de ingeniería.

4.2 Normalización de datos y partición de conjuntos de datos.

Todas las variables de entrada se normalizaron usando escalado Min-Max, con el objetivo de aumentar la estabilidad numérica durante el entrenamiento del modelo y reducir el impacto de las diferencias de escalado entre variables de entrada.

El procedimiento de estandarización se expresa como:

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

donde representa el valor original de una variable de entrada, y y indican, respectivamente, los valores mínimo y máximo de esa variable en el conjunto de datos. La normalización de variables es especialmente importante para algoritmos de regresión que implican cálculos de distancia o optimización iterativa, ya que evita que variables con amplitudes numéricas mayores dominen el proceso de aprendizaje (Liu et al., 2021; Yeh, 1998).

Tras la normalización, el conjunto de datos se dividió aleatoriamente en subconjuntos de entrenamiento y prueba, con el objetivo de evaluar la generalizabilidad de los modelos de aprendizaje automático desarrollados. En este estudio, el 30% del conjunto de datos se reservó para pruebas independientes, mientras que el 70% restante se utilizó para entrenamiento de modelos y ajuste de hiperparámetros. El subconjunto de entrenamiento se utilizó exclusivamente para calibración de modelos y ajuste de hiperparámetros, mientras que el subconjunto de prueba no se utilizó en ninguna etapa del desarrollo del modelo. Esta estrategia minimiza la fuga de datos y permite una evaluación rigurosa del rendimiento predictivo sobre datos no utilizados durante el entrenamiento, un aspecto particularmente importante para los conjuntos de datos de la literatura y caracterizado por la variabilidad del material.

4.3 Evaluación de la multicolinealidad utilizando el factor de inflación variable.

Los parámetros concretos de dosificación asociados al empaquetado de partículas suelen estar fuertemente correlacionados, lo que puede afectar negativamente a la estabilidad e interpretabilidad de los modelos. Para cuantificar el grado de multicolinealidad entre las variables de entrada, se calculó el factor de inflación de varianza (VIF) para cada variable.

El VIF se define como:

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

donde es el coeficiente de determinación obtenido regresando la variable predictora considerada en relación con todas las demás variables predictoras. Un valor VIF superior a 10 suele indicar una multicolinealidad severa (Kutner et al., 2005; Li et al., 2022). Para cuantificar el grado de multicolinealidad entre las variables de entrada relacionadas con el empaquetado de partículas, se calcularon los valores de TIF. Los valores resultantes para cada variable predictora se presentan en la Tabla 3. Los valores VIF extremadamente altos indican que la interpretación convencional basada en los coeficientes de los modelos es poco fiable, reforzando así la adecuación de los modelos de aprendizaje en conjunto basados en árboles R_i^2 para este problema.

Tabla 3. Valores del factor de inflación de varianza (VIF) de las variables de entrada.

Variable de entrada	Variable de entrada
Cemento	137.495
Gris volante (FA)	1405.819
Agregado grueso (CA)	580.775
Relación agua/aglutinante (W/B)	240.825
Agua	19.978
Contenido agregado reciclado (AR %)	1.025
Porción de Agregado Fino (FAR)	777.145
Fracción volumétrica de agregados (AVF)	35.664
Índice de Embalaje (PI)	12.89

Nota: Los valores VIF superiores a 10 indican una severa multi-colinealidad entre las variables predictoras. FAR indica la proporción de agregado fino, AVF indica la fracción volumétrica de agregados, PI indica el índice de empaquetado y W/B indica la relación agua/aglutinante.

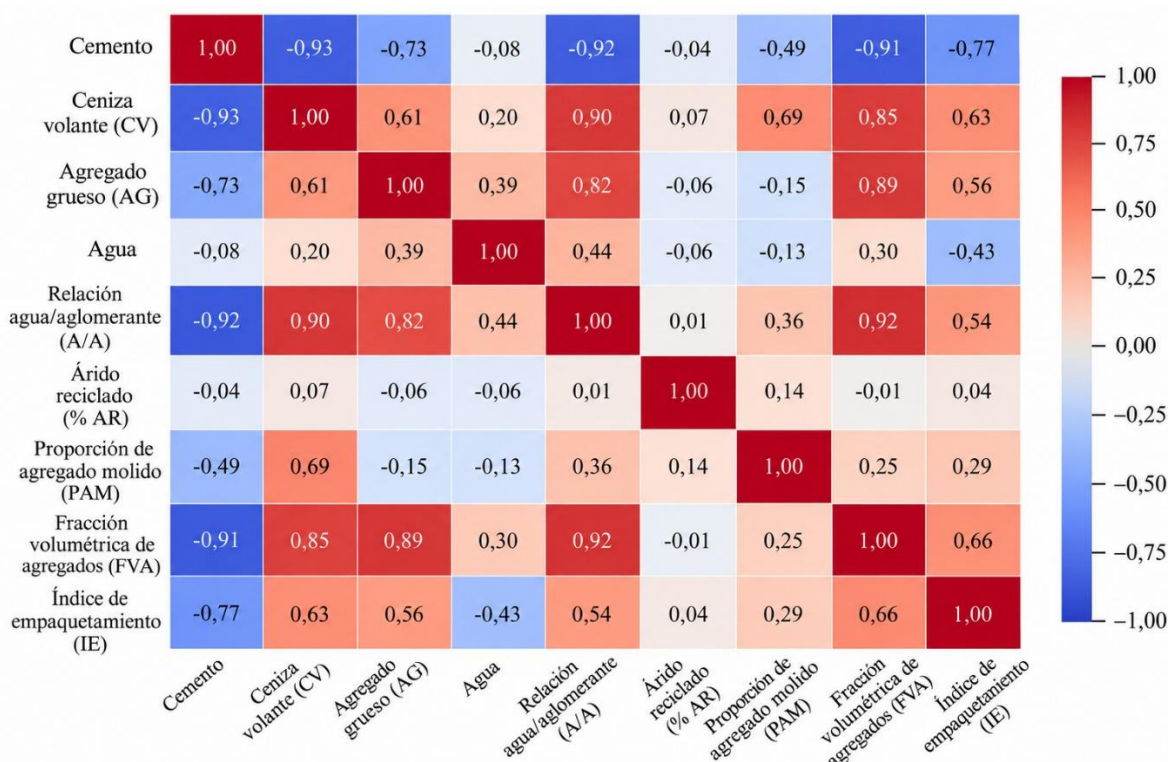


Figura 3. Matriz de correlación de las variables de entrada relacionadas con el empaquetado de partículas utilizada para predecir la resistencia a la compresión del RA-SCC.

La matriz de correlación de Pearson de las variables de entrada relacionadas con el empaquetado de partículas considerada en esta investigación se presenta en la Figura 3. La presencia de multicolinealidad se indica por las fuertes correlaciones observadas entre varias variables relacionadas con ligantes y agregados. Este aspecto se investiga cuantitativamente en la siguiente sección mediante el análisis del factor de inflación de varianza (VIF), para evaluar sus implicaciones para la estabilidad y selección de los modelos. Las fuertes correlaciones negativas observadas entre el contenido de cemento y el contenido de ceniza volante ($r = -0,93$), la relación agua/aglutinante ($r = -0,92$) y la fracción volumétrica agregada ($r = -0,91$) indican una interdependencia significativa entre la composición del aglutinante, la calidad de la pulpa y las características volumétricas del envasado. Por otro lado, el contenido de cenizas volantes presenta fuertes correlaciones positivas con la fracción volumétrica de los agregados ($r = 0,85$) y con la relación agua/aglutinante ($r = 0,90$), lo que indica la acción combinada de dispersión de pasta y contenido fino en mezclas de hormigón autocompactantes.

Las restricciones volumétricas que controlan el empaquetado de partículas se evidencian por las correlaciones moderadas a significativas observadas entre el contenido agregado grueso y la relación agua/aglutinante ($r = 0,82$) y la fracción volumétrica de agregados ($r = 0,89$). La independencia estadística relativa del contenido agregado reciclado en el conjunto de datos se confirma por sus bajas correlaciones con todas las demás variables. Las fuertes correlaciones entre las variables relacionadas con los ligantes y los agregados indican una multicolinealidad significativa. Este resultado apoya el uso de modelos de aprendizaje automático basados en árboles y métodos *de conjunto* para una predicción robusta de la resistencia y justifica el análisis posterior del factor de inflación de varianza (VIF).

4.4 Modelos de regresión por aprendizaje automático.

Para predecir la resistencia a la compresión del hormigón autocompactante con agregados reciclados, se desarrollaron seis modelos de regresión.

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

4.4.1 Regresión lineal y regresión de la dorsal bayesiana.

Según el modelo de regresión lineal, las variables de entrada y la resistencia a la compresión tienen una relación lineal, representada de la siguiente manera:

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

donde representa la resistencia a compresión predicha, representa las variables de entrada y son los coeficientes de regresión. $\hat{y}x_i\beta_i$

La regresión bayesiana amplía esta formulación introduciendo distribuciones probabilísticas a priori en los coeficientes de regresión, con el objetivo de mejorar la estabilidad numérica en presencia de multicolinealidad (Montgomery et al., 2012).

4.4.2 Modelos basados en árboles de decisión.

La regresión por árbol de decisión, un método de aprendizaje supervisado no paramétrico, modela relaciones no lineales dividiendo recursivamente el espacio de variables de entrada en un conjunto de regiones disjuntas. En cada nodo interno del árbol, el conjunto de datos se divide en función de una variable de entrada seleccionada y un valor umbral correspondiente, con el objetivo de minimizar el error de predicción en los subconjuntos resultantes.

Este proceso de división binaria recursiva continúa hasta que se cumplen criterios de parada predefinidos, como el número mínimo de muestras por nodo o la profundidad máxima del árbol. El promedio de los valores observados correspondientes a las muestras contenidas en un nodo terminal dado (hoja) se utiliza para calcular la resistencia a la compresión prevista para ese nodo. La función de predicción de un modelo de árbol de decisión puede expresarse de la siguiente manera:

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

donde representa el número total de regiones terminales, indica la región del espacio variable, representa la resistencia media a la compresión de las muestras contenidas en la región, y es una función indicadora, que asume un valor de 1 cuando se cumple la condición y 0 en caso contrario (Breiman et al., 1984). $MR_m c_m \mathbb{I}(\cdot)$

En problemas de regresión, la división óptima en cada nodo suele determinarse minimizando la varianza o el error cuadrático medio de la variable objetivo en los nodos hijos. Este criterio permite a los árboles de decisión capturar relaciones complejas entre variables de entrada, sin necesidad de establecer hipótesis sobre su estructura funcional. Dado que los indicadores de empaquetado de partículas y los parámetros de dosificación interactúan de forma altamente no lineal en sistemas materiales heterogéneos, como el hormigón autocompactante con agregados reciclados, los modelos de árbol de decisión son por tanto particularmente adecuados para modelar estos sistemas (Asteris et al., 2020; Yeh, 1998).

4.4.3 Regresores de Bosque Aleatorio y Árboles Extra

En la regresión de bosque aleatorio, se construye un *conjunto* de árboles de decisión, entrenados a partir de muestras obtenidas mediante *remuestreo bootstrap*, y la predicción final se obtiene promediando las predicciones de los árboles individuales (Breiman, 2001):

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

donde representa el número de árboles, y representa la predicción hecha por el árbol. $M\hat{y}_m m^{th}$
El regresor Extra Trees aumenta aún más el grado de aleatoriedad seleccionando al azar

umbrales de división, lo que mejora la generalizabilidad y reduce la varianza del modelo, especialmente en conjuntos de datos con variables predictoras correlacionadas.

4.4.4 máquina de aumento de gradiente de luz.

La Light Gradient Boosting Machine (LightGBM) es un marco de aumento de gradiente que construye secuencialmente árboles de regresión aditiva. El resultado previsto se obtiene por:

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

donde f_t representa el árbol de decisión y corresponde al número total de iteraciones de f_t *Impulso* (Ke et al., 2017).

4.5 Métricas de evaluación del rendimiento del modelo.

Para garantizar una evaluación integral y estadísticamente fiable del rendimiento de los modelos, se utilizaron varios indicadores cuantitativos para evaluar la precisión de las predicciones, la magnitud de los errores, el sesgo sistemático y la consistencia estadística entre los valores predichos y experimentales de la resistencia a la compresión. El uso de múltiples métricas de evaluación es esencial para predecir la resistencia del hormigón, ya que la adopción de un único indicador puede llevar a conclusiones erróneas sobre la robustez del modelo (Montgomery et al., 2012; Willmott & Matsuura, 2005).

El coeficiente de determinación (R^2) se utilizó para cuantificar la proporción de la varianza de la resistencia a compresión medida que se explica por las predicciones del modelo. Se define como:

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

donde y_i y $\hat{y}_i$ representan, respectivamente, los valores de resistencia a la compresión medidos experimentalmente y predichos para la muestra, $\bar{y}$ representa la media de los valores medidos y N corresponden al número total de observaciones. Un valor mayor de R^2 indica una mayor concordancia entre las predicciones y los resultados experimentales.

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

El error absoluto medio (MAE) proporciona una medida directa de la magnitud media de los errores de predicción y es menos sensible a valores extremos en comparación con métricas basadas en error cuadrático. Por otro lado, el error cuadrático medio (RMSE) penaliza los errores de magnitud mayor de forma más marcada y se expresa como:

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

En aplicaciones de ingeniería, donde desviaciones mayores respecto a los valores medidos pueden tener importantes implicaciones estructurales, la RMSE es especialmente útil para evaluar el rendimiento del modelo (Willmott & Matsuura, 2005).

Para cuantificar la sobreestimación o subestimación sistemática de las predicciones, el error medio de sesgo (EBM) se calculó de la siguiente manera:

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

Una tendencia a subestimar se indica con un EBM positivo, mientras que un valor negativo indica sobreestimación. Valores cercanos a cero indican predicciones imparciales, una condición necesaria para la toma de decisiones fiable en ingeniería.

Para determinar si las diferencias entre los valores experimental y predicho de resistencia a la compresión eran estadísticamente significativas, se realizó una prueba de hipótesis estadística, además del uso de métricas basadas en errores. Se aplicó una prueba t pareada para evaluar la hipótesis nula de que la diferencia media entre los valores medidos y predichos es igual a cero. Se evaluó la concordancia estadística usando la estadística t resultante y el valor p asociado; para un nivel de confianza del 95%, valores p superiores a 0,05 no indicaron ninguna diferencia estadísticamente significativa (Montgomery et al., 2012). Esta validación estadística aumenta la fiabilidad de las predicciones hechas por los modelos de aprendizaje automático desarrollados.

4.6 Interpretabilidad del modelo basada en SHAP.

Se utilizó el método SHapley Additive Explanations (SHAP) para interpretar el papel de los indicadores de empaquetado de partículas en las predicciones del modelo. Los valores de SHAP satisfacen la propiedad de aditividad y se derivan de la teoría cooperativa de juegos:

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

donde representa la contribución del rasgo a la predicción (Lundberg & Lee, 2017). Este enfoque permite cuantificar tanto la magnitud como la dirección de la influencia de cada variable, permitiendo una relación directa entre los resultados del aprendizaje automático y los mecanismos físicos del empaquetado de partículas. Es importante señalar que los valores SHAP cuantifican la contribución de las variables de entrada a las predicciones del modelo, pero no establecen una relación causal física estricta. En conjuntos de datos caracterizados por una fuerte multicolinealidad, como en el presente estudio, las variables correlacionadas pueden compartir o redistribuir sus contribuciones entre sí. En consecuencia, la importancia atribuida a cada variable individual debe interpretarse con precaución, especialmente cuando se observan valores altos del factor de inflación de la varianza (VIF). En el conjunto de datos utilizado en este estudio, el contenido de cemento, el contenido de ceniza volante y la relación agua/aglutinante muestran una fuerte interdependencia, reflejando restricciones prácticas inherentes a la dosificación del hormigón, en lugar de representar variables físicas independientes. Bajo estas condiciones, la importancia basada en SHAP debe interpretarse como una representación de la influencia conjunta de variables correlacionadas, más que como efectos causales aislados. A pesar de esta limitación, los resultados de SHAP obtenidos en este estudio permanecen físicamente coherentes, ya que las variables identificadas como predominantes—contenido de agregados reciclados, contenido de cemento y relación agua/aglutinante—se sabe que son determinantes que influyen en la resistencia a la compresión y el comportamiento de empaquetado de partículas. La consistencia entre la valoración de importancia obtenida por SHAP y los principios establecidos de la mecánica del hormigón proporciona confianza en que el modelo captura un comportamiento realista del material, incluso en presencia de multicolinealidad. Por lo tanto, en este estudio, el análisis SHAP se interpreta como una herramienta explicativa que proporciona información sobre la influencia de variables dentro de un sistema correlacionado, más que como un método de atribución causal estricta. Esta distinción es esencial para garantizar una interpretación científicamente rigurosa de los resultados. ϕ_i^{th}

4.7 Implementación computacional

Para garantizar la reproducibilidad, transparencia y eficiencia computacional, se utilizaron bibliotecas de computación científica basadas en Python en todas las etapas del procesamiento de datos, el modelado de aprendizaje automático y el análisis estadístico.

Python fue seleccionado debido a su amplio ecosistema de herramientas de código abierto, ampliamente adoptadas en la investigación de materiales basada en datos y en aplicaciones de ingeniería civil (McKinney, 2010; Pedregosa et al., 2011).

El flujo de trabajo computacional comprendía pasos secuenciales, incluyendo la normalización de datos, la partición de conjuntos de datos, el cálculo del factor de inflación de varianza (VIF), el desarrollo de modelos, la evaluación del rendimiento y el análisis explicable de aprendizaje automático. Para proporcionar una comparación justa del rendimiento, los procedimientos de normalización de las variables y preprocesamiento de los datos se aplicaron de forma uniforme a todos los modelos. Se utilizaron funciones estandarizadas de bibliotecas computacionales para el desarrollo de los algoritmos de aprendizaje automático, permitiendo procesos fiables y reproducibles de entrenamiento y prueba de los modelos. El ajuste de los hiperparámetros se realizó en el conjunto de datos de entrenamiento mediante validación cruzada, con el objetivo de mejorar la generalización de los modelos y evitar *el sobreajuste*. Para evitar *fugas de datos* y asegurar una evaluación objetiva del rendimiento predictivo, todos los modelos entrenados se probaron exclusivamente en el conjunto de datos de prueba independiente. Las métricas de evaluación estadística y las pruebas de hipótesis se calcularon utilizando bibliotecas numéricas consolidadas para asegurar la precisión y estabilidad numérica (Montgomery et al., 2012).

La interpretabilidad del modelo se obtuvo utilizando el método SHapley Additive ExPlanations (SHAP), cuyos valores se calcularon mediante algoritmos de aproximación eficientes compatibles con modelos basados en árboles. Este enfoque permite una atribución consistente de las contribuciones de las variables, tanto para predicciones individuales como para el comportamiento general del modelo, permitiendo una interpretación significativa de los indicadores de empaquetamiento de partículas que influyen en la resistencia a la compresión (Lundberg & Lee, 2017). Los pasos intermedios del flujo de trabajo de modelado, incluyendo el preprocesamiento de datos, la evaluación de multicolinealidad, el entrenamiento del modelo y la evaluación del rendimiento, se validaron sistemáticamente mediante el enfoque de división de datos en conjuntos de entrenamiento y prueba. Se empleó la validación cruzada durante el ajuste de los hiperparámetros para asegurar la estabilidad de los modelos y evitar *el sobreajuste*. Por lo tanto, los resultados presentados reflejan tanto la precisión de los modelos como su capacidad de generalización, basándose en datos de prueba no utilizados durante el entrenamiento.

El uso de software de código abierto y prácticas de implementación estandarizadas garantiza que el marco desarrollado pueda ser fácilmente reproducido y ampliado por otros investigadores. Este enfoque abierto facilita tanto el desarrollo futuro de modelos explicables de aprendizaje automático aplicados a materiales de hormigón sostenibles como su validación y comparación.

5. RESULTADOS Y DEBATE

5.1 Comparación del rendimiento de modelos de aprendizaje automático.

Se utilizaron varias métricas estadísticas, como el coeficiente de determinación (R^2), el error absoluto medio (MAE), el error cuadrático medio raíz (RMSE), el error porcentual absoluto medio (MAPE), el error medio de sesgo (MBE) y las pruebas de hipótesis, para evaluar el rendimiento predictivo de los seis modelos de aprendizaje automático. Los resultados demuestran claramente que, en la predicción de la resistencia a la compresión del hormigón autocompactante con agregados reciclados (RA-SCC), los modelos basados en *métodos de*

conjunto tienen un rendimiento superior a los modelos lineales y los modelos basados en un solo árbol de decisión. Con un valor de $R^2 = 0,9724$, el Regresor de Árboles Extra mostró la mayor precisión predictiva entre todos los modelos, indicando que pudo explicar más del 97% de la variabilidad de la resistencia a la compresión. Este modelo también tuvo los valores más bajos de MAE (0,79 MPa) y RMSE (1,19 MPa), demostrando un rendimiento superior en la minimización de errores. Los modelos Random Forest y Light Gradient Boosting Machine también mostraron un rendimiento excelente, con valores de R^2 de 0,968 y 0,965, respectivamente. Estos resultados corroboran las conclusiones de estudios previos sobre hormigón con agregados reciclados, según qué técnicas *de aprendizaje en conjunto* son adecuadas para modelar el comportamiento altamente no lineal del hormigón (Ahmed et al., 2024; Asteris et al., 2020; Liu et al., 2021). La Tabla 4 presenta un resumen del coeficiente de determinación, métricas basadas en errores, error medio de sesgo y resultados de pruebas de hipótesis para cada modelo.

Tabla 4. Comparación del rendimiento de modelos de aprendizaje automático para la predicción de la resistencia a la compresión.

Modelo	R^2	MAE (MPa)	MSE (MPa ²)	RMSE (MPa)	MAPE (%)	MBE (MPa)	t-Estadísticas	valor p
Regresión lineal	0.9526	1.336	2.444	1.563	3.8	-0.042	-0.282	0.779
Regresión de la dorsal bayesiana	0.9146	1.525	4.404	2.099	4.13	-0.042	-0.211	0.833
Regresor del árbol de decisión	0.9363	1.167	3.28	1.811	3.06	0.203	1.177	0.242
Regresor Random Forest	0.968	0.855	1.648	1.284	2.31	0.095	0.771	0.443
Árboles Extra Regressor	0.9724	0.788	1.423	1.193	2.12	0.122	1.074	0.285
Máquina de refuerzo de gradiente de luz	0.965	0.866	1.802	1.343	2.3	0.205	1.61	0.11

Nota: MAE indica el error absoluto medio, MSE indica el error cuadrático medio, RMSE indica el error cuadrático medio raíz, MAPE indica el porcentaje absoluto medio y MBE indica el sesgo medio.

La precisión de las predicciones y los niveles de error de la Regresión Bayesiana de Cresta y la Regresión Lineal fueron significativamente menores. La causa principal de esta reducción del rendimiento es la fuerte interdependencia entre las variables de dosificación, lo que contradice las suposiciones del modelado lineal. Estudios previos sobre la predicción de la resistencia del hormigón basados en aprendizaje automático han documentado limitaciones similares de la regresión lineal en presencia de multicolinealidad (Li et al., 2022).

La Figura 4 presenta diagramas de dispersión que comparan los valores experimentales y predichos de resistencia a la compresión para cada uno de los seis modelos de aprendizaje automático considerados en esta investigación. La agrupación concentrada de puntos de datos alrededor de la línea de referencia de 45° indica alta precisión predictiva y bajo sesgo en modelos basados en métodos *de conjunto*, especialmente en los regresores Extra Trees y Random Forest. La Light Gradient Boosting Machine también muestra un fuerte acuerdo, aunque presenta una dispersión ligeramente mayor para valores de resistencia más altos. Por otro lado, los modelos de Regresión Lineal, Regresión Bayesiana de Cresta y Árbol de Decisión muestran mayor dispersión, reflejando su menor capacidad para capturar las interacciones no lineales entre variables relacionadas con el empaquetado de partículas. Los patrones visuales corroboran la mayor fiabilidad predictiva de las técnicas *de aprendizaje de conjunto* para el

hormigón autocompactante con áridos reciclados y están en consonancia con los indicadores cuantitativos de rendimiento presentados en la Tabla 4.

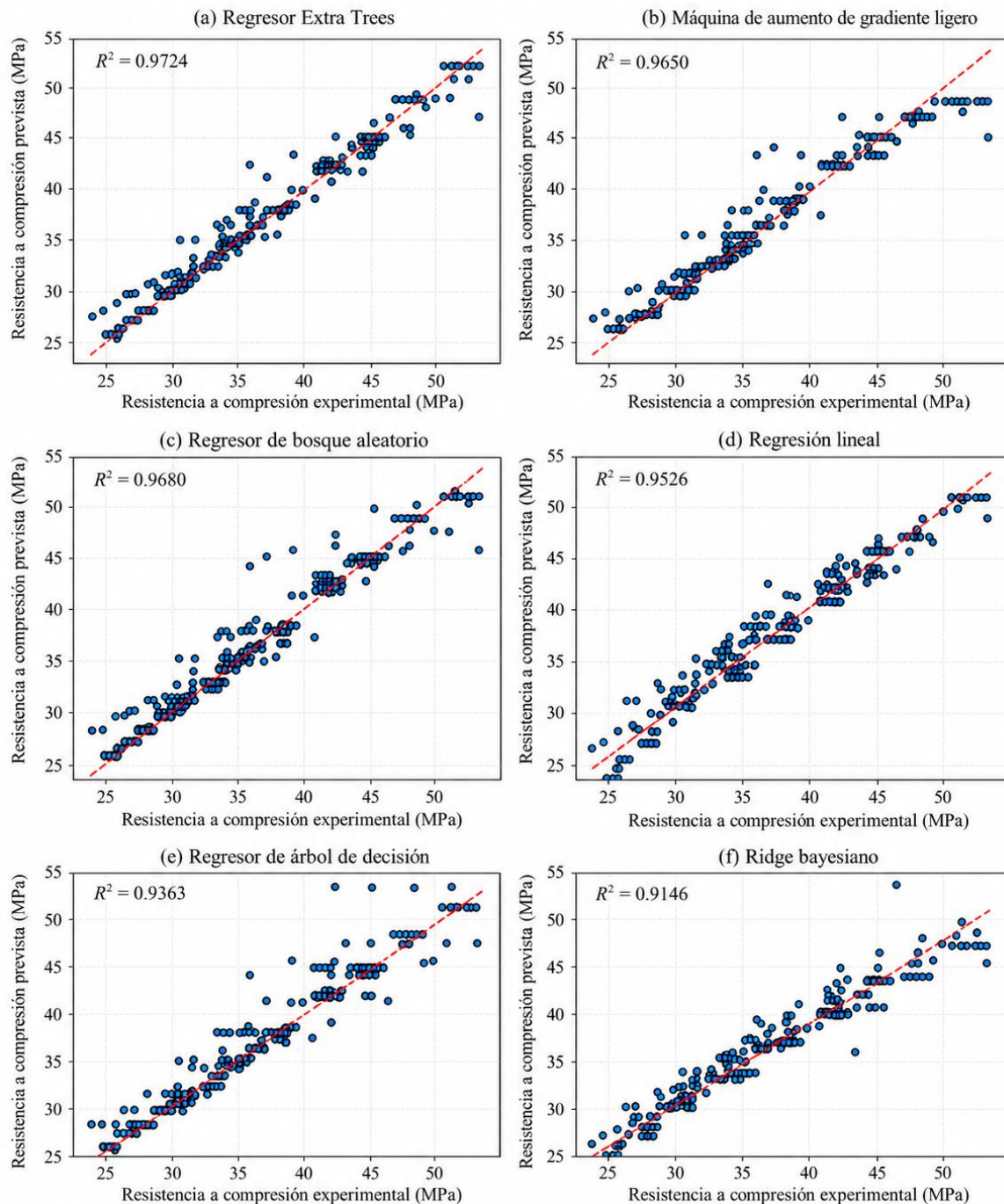


Figura 4. Comparación entre la resistencia a compresión predicha y experimental obtenida mediante diferentes modelos de aprendizaje automático: (a) regresor Extra Trees, (b) Light Gradient Reinforcement Machine, (c) regresor de bosque aleatorio, (d) regresión lineal, (e) regresor de árbol de decisión y (f) regresión Bayesian Ridge.

La Figura 5 presenta las distribuciones residuales obtenidas para los seis modelos de aprendizaje automático, proporcionando información sobre el sesgo de predicción y la dispersión de errores, además de métricas de precisión convencionales. El Regresor de Árboles Extra presenta un perfil residual de distribución estrecha, aproximadamente simétrico y

centrado en cero, consistente con su bajo error medio de sesgo (MBE = 0,122 MPa), RMSE reducido (1,19 MPa) y alto coeficiente de determinación ($R^2 = 0,9724$). La alta estabilidad predictiva también está confirmada por el Regresor de Bosque Aleatorio, que presenta una distribución compacta de residuos y bajo sesgo (MBE = 0,0945 MPa; RMSE = 1,28 MPa). Aunque tiene una dispersión ligeramente mayor, evidenciada por su RMSE de 1,34 MPa, la Máquina de Aumento de Gradiente de Luz también presenta residuos concentrados cercanos a cero (MBE = 0,2046 MPa). Por otro lado, los modelos de Regresión Lineal y Regresión de Dorsal Bayesiana presentan distribuciones residuales más amplias, a pesar de los valores de sesgo cercanos a cero (MBE $\approx -0,04$ MPa), lo que indica un mayor error aleatorio y una menor capacidad para representar las interacciones no lineales entre las variables relacionadas con el empaquetado de partículas. El Regresor del Árbol de Decisión presenta asimetría y dispersión moderadas, en acuerdo con su RMSE comparativamente mayor (1,81 MPa).

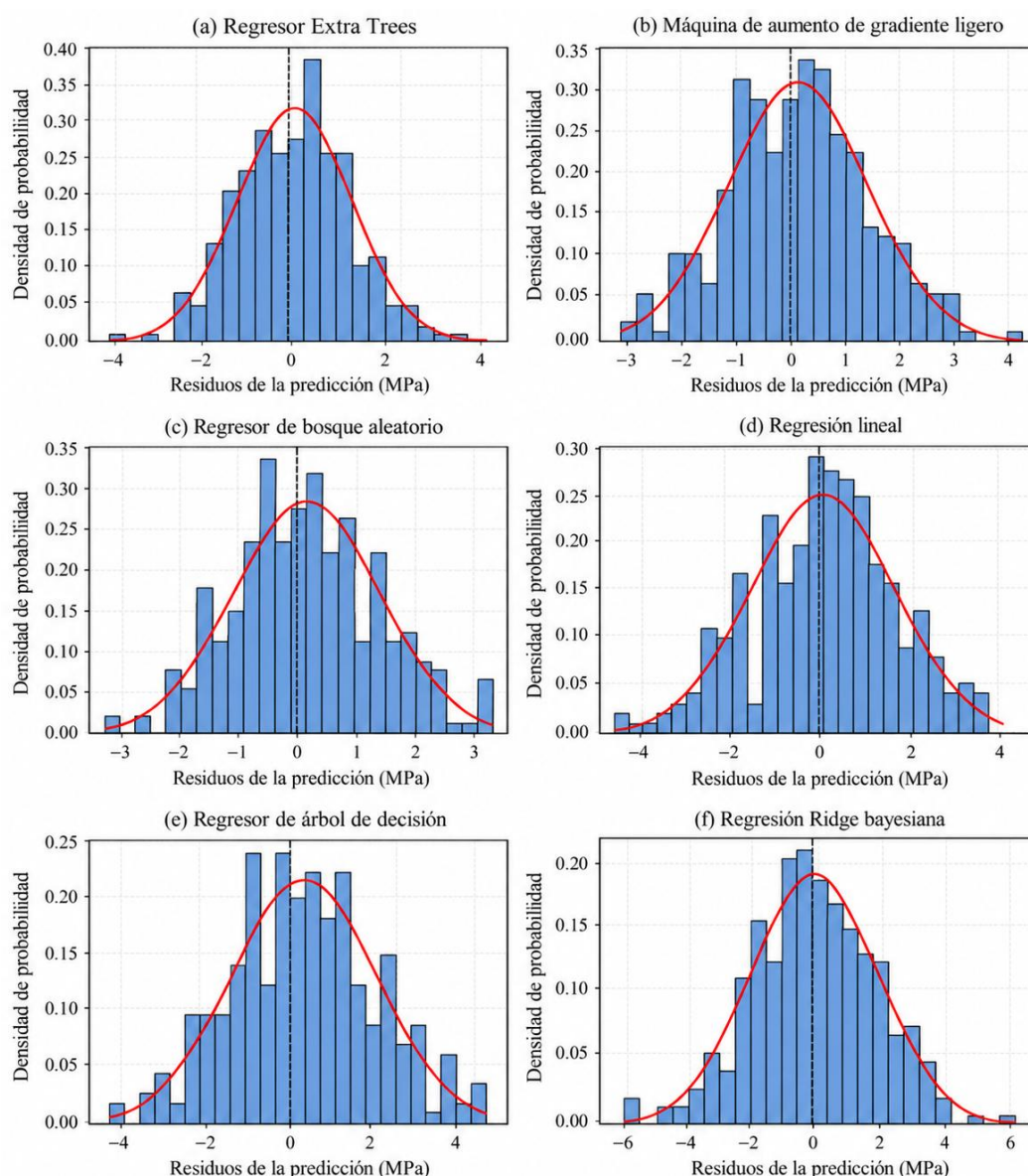


Figura 5. Distribución de los residuos de predicción obtenidos utilizando diferentes modelos de aprendizaje automático: (a) regresor de árboles extra, (b) máquina de refuerzo de gradiente de luz, (c) regresor de bosque aleatorio, (d) regresión lineal, (e) regresor de árbol de decisión y (f) regresión de dorsales bayesianas.

5.2 Fiabilidad estadística y evaluación del sesgo.

Además de las métricas de precisión convencionales, la fiabilidad estadística se evaluó mediante el error de sesgo medio (EBM) y pruebas de hipótesis. Los valores de EBM para todos los modelos fueron cercanos a cero, lo que indica la ausencia de sobreestimación o subestimación sistemática de la resistencia a la compresión. Cabe destacar que el Regresor de Árboles Extra presentó un EBM de 0,12 MPa, confirmando su capacidad para hacer predicciones sin sesgo significativo. El estadístico t y los respectivos valores p refuerzan esta observación. Para todos los modelos, los valores p fueron superiores al nivel de significación de 0,05, lo que indica que no hubo diferencia estadísticamente significativa entre los valores predichos y experimentales de resistencia a la compresión. La consistencia estadística observada en *los modelos de conjunto* aumenta la confianza en su aplicabilidad en la toma de decisiones de ingeniería, en la que las predicciones fiables e imparciales son esenciales (Montgomery et al., 2012).

5.3 Influencia de la multicolinealidad en el rendimiento del modelo.

El análisis del factor de inflación de varianza (VIF) reveló una multicolinealidad severa entre varias variables de entrada relacionadas con el empaquetado de partículas. El contenido de cenizas volantes, el contenido de agregados gruesos, la proporción de agregado fino y la relación agua/aglutinante mostraron valores VIF extremadamente altos, lo que confirma la fuerte interdependencia entre estos parámetros. Esta multicolinealidad explica el rendimiento inferior de los modelos de regresión lineal y regresión bayesiana, ya que la inestabilidad de los coeficientes y el aumento de la varianza comprometen la fiabilidad de las predicciones. Como se muestra en la Tabla 3, varias variables relacionadas con el empaquetado de partículas presentan una multicolinealidad severa, particularmente el contenido de cenizas volantes, las fracciones agregadas y la relación agua/aglutinante, mientras que el contenido agregado reciclado sigue siendo estadísticamente independiente. La tasa de sustitución por agregado reciclado presentó un valor VIF cercano a la unidad, indicando independencia estadística. Esta característica convierte el contenido agregado reciclado en una variable predictora robusta tanto en estructuras de modelado lineales como no lineales. El rendimiento superior de *los modelos de conjunto* basados en árboles observados en este estudio está en línea con resultados de investigaciones previas que demuestran que estos algoritmos son inherentemente robustos a la multicolinealidad y capaces de capturar interacciones complejas entre variables sin requerir una decorrelación explícita de las variables de entrada (Breiman, 2001; Li et al., 2022). Parte de la multicolinealidad observada también puede atribuirse a inconsistencias en la presentación de tipos de cemento y materiales cementicios suplementarios entre los diferentes estudios. En particular, el uso de cementos compuestos que contienen ceniza volante puede dar lugar a una representación superpuesta de los componentes del aglutinante, ampliando así las correlaciones estadísticas entre las variables de entrada.

5.4 Interpretación de indicadores de empaquetado de partículas basados en SHAP.

Para determinar la importancia relativa de los indicadores de empaquetado de partículas e interpretar las predicciones del modelo con mejor rendimiento, se realizó un análisis utilizando el método SHapley Additive ExPlanations (SHAP). Los resultados muestran que el factor más importante es el porcentaje de sustitución por agregado reciclado, seguido del contenido de cemento y la proporción agua/aglutinante. Esta clasificación destaca el papel predominante de la calidad de los agregados, la disponibilidad de lodos y la estructura interna de vacíos en el desarrollo de la resistencia a la compresión RA-SCC (Tang et al., 2023). La fuerte influencia del contenido de agregados reciclados puede atribuirse a su impacto en la eficiencia del empaquetado de partículas y en las características de las zonas de transición interfaciales. Los agregados reciclados introducen porosidad adicional y mayor rugosidad superficial debido a la

presencia de mortero adherido, lo que afecta tanto a la densidad del envase como a los mecanismos de transferencia de tensión (Etxeberria et al., 2007; Poon et al., 2004; H. Zhang et al., 2022).

El contenido de cemento y la proporción agua/aglutinante controlan directamente el volumen de la suspensión y los productos de hidratación, influyendo aún más en el relleno de vacíos y la conectividad de partículas. Para el modelo de aprendizaje automático con mejor rendimiento, se utilizó el método SHapley Additive exPlanations para analizar la contribución individual de los indicadores de empaquetado de partículas a la predicción de la resistencia a la compresión. El gráfico resumen de SHAP muestra cómo se distribuyen los valores individuales de SHAP a lo largo del conjunto de datos de entrenamiento, mientras que el gráfico de barras de importancia variable se elaboró usando los valores absolutos medios finales de SHAP obtenidos del Regresor de Árboles Extra entrenado. La Tabla 5 presenta un resumen de los valores absolutos medios de SHAP y las respectivas clasificaciones de importancia de las variables. Esta consistencia entre los resultados obtenidos mediante aprendizaje automático explicable y los principios consolidados de la mecánica del hormigón aumenta la confianza en la aplicabilidad práctica de la estructura propuesta.

Tabla 5. Clasificación de la importancia de las variables basándose en SHAP para el modelo con mejor rendimiento.

Clasificación	Variable	Valor absoluto medio de SHAP
1	Contenido agregado reciclado (AR %)	2.476
2	Cemento	1.483
3	Relación agua/aglutinante (W/B)	1.401
4	Gris volante (FA)	0.728
5	Agregado grueso (CA)	0.703
6	Fracción volumétrica de agregados (AVF)	0.694
7	Índice de Embalaje (PI)	0.157
8	Agua	0.088
9	Proporción de agregado fino (FAR)	0.08
10	Edad	0

Nota: Los valores absolutos medios de SHAP representan la magnitud media de la contribución de cada variable a las predicciones del modelo. Valores más altos indican una mayor influencia en la resistencia a la compresión. FAR indica la proporción de agregados finos, AVF indica la fracción volumétrica de agregados, PI indica el índice de empaquetado y W/B indica la relación agua/aglutinante.

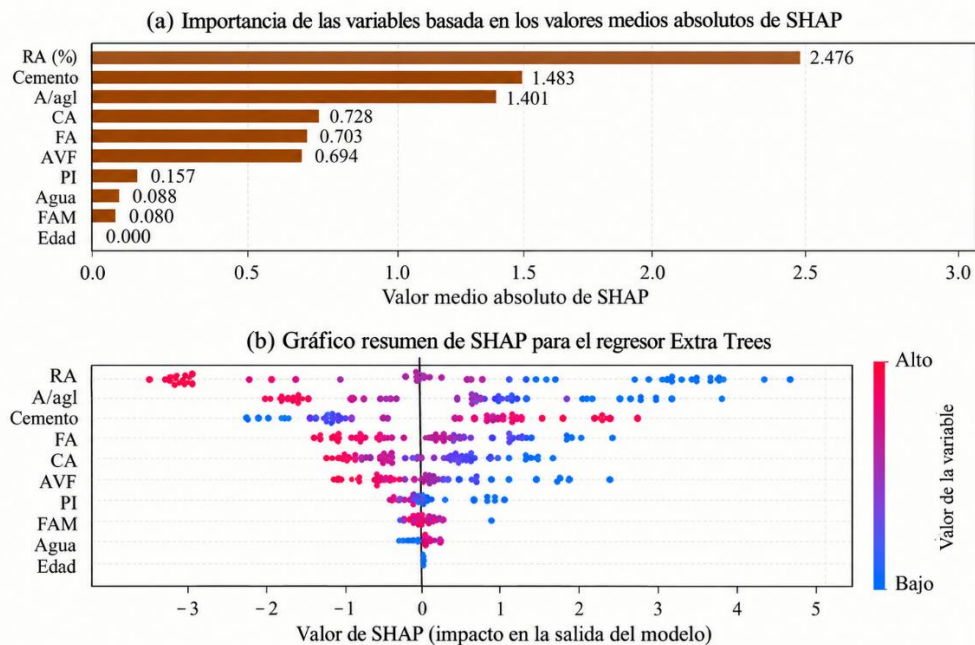


Figura 6. Interpretación basada en SHAP del Regresor de Árboles Extra, mostrando (a) la importancia de las variables clasificadas en función de los valores absolutos medios de SHAP y (b) el gráfico resumen de SHAP, que ilustra la influencia de las variables de entrada relacionadas con el empaquetado de partículas en la predicción de la resistencia a la compresión.

La Figura 6 presenta el análisis de explicabilidad basado en SHAP para el Regresor de Árboles Extra, proporcionando una evaluación cuantitativa de la contribución relativa de variables relacionadas con el empaquetado de partículas a la predicción de la resistencia a la compresión. Según la importancia del gráfico de variables presentado en la Figura 6(a), la tasa de sustitución por agregado reciclado es la variable predictora más significativa, presentando el valor medio absoluto más alto de SHAP (2,48), seguida del contenido de cemento (1,48) y la relación agua/aglutinante (1,40). También debe señalarse que la importancia relativamente mayor del contenido de agregados reciclados puede estar parcialmente influenciada por su amplio rango de variación (0–100%) en comparación con las otras variables de entrada. Esta mayor variabilidad incrementa su contribución estadística dentro del modelo y debe considerarse en la interpretación de la clasificación de importancia de las variables. Estas variables controlan directamente la calidad de los agregados, la disponibilidad de suspensión y la estructura interna de los huecos, factores que, en conjunto, gobiernan la eficiencia del empaquetado de partículas en hormigón autocompactante.

El contenido de ceniza volante (0,73), el contenido de agregados gruesos (0,70) y la fracción volumétrica de los agregados (0,69) muestran contribuciones moderadas, lo que indica su papel secundario en el ajuste de la densidad de empaquetado y la dispersión de la pulpa. Por otro lado, el índice de empaquetado (0,16), el contenido de agua (0,09) y la proporción de agregado fino (0,08) tienen una influencia relativamente limitada, mientras que la edad de curado tiene una contribución insignificante al conjunto de datos analizado.

El gráfico resumen SHAP presentado en la Figura 6(b) ilustra aún más la dirección de la influencia de estas variables en las predicciones individuales. Un mayor contenido de agregados reciclados y un mayor valor de la relación agua/aglutinante generalmente tienen un efecto negativo en la resistencia a la compresión, mientras que un mayor contenido de cemento contribuye positivamente.

Los resultados del análisis SHAP confirman que el modelo Extra Trees es capaz de capturar Predicción explicable basada en aprendizaje automático de la resistencia a la compresión a 28 días de hormigón autocompactante con agregados reciclados utilizando indicadores de empaquetado de partículas.

mecanismos de empaquetado de partículas físicamente coherentes, proporcionando un soporte transparente y basado en datos para el marco predictivo propuesto. También debe señalarse que, debido a la multicolinealidad entre las variables relacionadas con los conglomerados, los valores SHAP reflejan contribuciones distribuidas entre variables de entrada correlacionadas. En consecuencia, la importancia relativa de las variables individuales debe interpretarse junto con el conocimiento técnico del dominio, y no como una representación de efectos causales independientes.

El contenido de materiales cementicios suplementarios y los parámetros volumétricos de los agregados mostraron contribuciones moderadas, indicando su papel secundario pero no despreciable en la mejora del empaquetamiento de partículas y la microestructura. Por otro lado, la edad de curado mostró una influencia insignificante en el conjunto de datos compilado, lo que sugiere que los parámetros de composición de la mezcla predominan en la variación de resistencia dentro del rango de datos seleccionado. Es importante interpretar críticamente la importancia relativamente baja del índice de empaquetamiento (PI) y la relación de agregados finos (FAR), a pesar de su relevancia conceptual para el comportamiento del empaquetamiento de partículas. Aunque estos parámetros pretenden representar características de empaquetado, constituyen indicadores derivados o secundarios, dependientes de las variables subyacentes de composición de la mezcla. Como resultado, su contribución independiente a las predicciones del modelo es comparativamente limitada. En contraste con la variabilidad de dosis, resultados similares se han documentado en conjuntos de datos procedentes de la literatura con distribución por edades restringida (Asteris et al., 2020; Fanijo et al., 2023). El análisis SHAP indica que las variables primarias de dosificación, como el contenido de agregados reciclados, el contenido de cemento y la relación agua/aglutinante, ejercen una mayor influencia en la resistencia a la compresión. Esto se debe a que estos parámetros controlan directamente el volumen de la suspensión de lodos, la interacción de partículas y la estructura de vacíos, que en conjunto determinan la eficiencia del envasado. Por otro lado, PI y FAR representan estos efectos de forma agregada o indirecta, resultando en una menor importancia individual en el modelo de aprendizaje automático. Esta observación no contradice la estructura basada en el empaquetado de partículas propuesta en este estudio. Al contrario, demuestra que el modelo de aprendizaje automático captura implícitamente los mecanismos de envasado a través de las variables de dosificación predominantes, sin depender en exceso de índices de envasado explícitamente definidos. Los resultados sugieren que el comportamiento del envasado está intrínsecamente incrustado en las relaciones entre el contenido de material cementicio, las proporciones de agregados y la demanda de agua, en lugar de estar gobernado por una única variable indicadora. Por lo tanto, la clasificación relativamente baja de PI y RF debe interpretarse como una indicación de que los efectos del empaquetado de partículas están distribuidos entre múltiples variables interactuantes, en lugar de estar aislados en índices específicos. Esto refuerza la capacidad del aprendizaje automático explicable para capturar implícitamente mecanismos físicos complejos, en línea con el objetivo central de este estudio.

5.5 Implicaciones para la dosificación impulsada por empaquetado de partículas.

Los resultados combinados de los análisis VIF y SHAP muestran que el aprendizaje automático puede capturar implícitamente los mecanismos del empaquetado de partículas a través de las variables de composición de la mezcla, incluso si la densidad de empaquetado de partículas o la razón de vacíos no se ha determinado explícitamente. El fuerte acuerdo entre la importancia de las variables obtenidas mediante SHAP y la teoría del empaquetado consolidado confirma la consistencia física del modelo desarrollado. Esta representación implícita ofrece una alternativa práctica a las pruebas experimentales de envasado, que suelen ser laboriosas y específicas de cada material (Cheng et al., 2021; de Larrard, 1999). Desde una perspectiva práctica, los resultados sugieren que controlar el contenido de agregados reciclados, la dosis de

aglutinante y la relación agua/aglutinante es fundamental para optimizar la resistencia a la compresión de RA-SCC. El marco desarrollado proporciona tanto capacidad predictiva como información interpretativa, contribuyendo a una dosificación concreta sostenible y basada en datos.

6. CONCLUSIONES Y DIRECCIONES PARA FUTURAS INVESTIGACIONES

6.1 Conclusiones.

Mediante el uso de indicadores de empaquetado de partículas derivados de las características de composición de la mezcla, este estudio desarrolló un marco explicable de aprendizaje automático para la predicción de la resistencia a la compresión del hormigón autocompactante con agregados reciclados (RA-SCC). Desde una perspectiva ingenierina, el modelo desarrollado puede ayudar a los profesionales en la selección del contenido de agregados reciclados y las proporciones de aglutinantes, buscando equilibrar los requisitos de sostenibilidad y resistencia, sin recurrir a pruebas de laboratorio basadas en prueba y error. Basándose en un conjunto de datos de la literatura y en un análisis estadístico exhaustivo, se pueden extraer las siguientes conclusiones:

1. Entre los seis modelos evaluados, los Árboles Extra Regresores presentaron la mayor precisión predictiva, alcanzando R^2 de 0,9724, MAE de 0,79 MPa, RMSE de 1,19 MPa y MAPE del 2,12%. Estos resultados demuestran una alta concordancia entre los valores experimentales y predichos de resistencia a la compresión.
2. Los modelos basados en métodos *de conjunto*, incluyendo Extra Trees, Random Forest ($R^2 = 0,968$) y Light Gradient Boosting Machine ($R^2 = 0,965$), superaron consistentemente a los modelos de regresión lineal y regresión bayesiana. Esta ventaja de rendimiento pone de manifiesto la capacidad de los *modelos de conjunto* para capturar interacciones no lineales entre las variables de dosificación que gobiernan el comportamiento de empaquetado de partículas.
3. Los modelos lineal y bayesiano mostraron menor precisión debido a la severa multicolinealidad entre los parámetros relacionados con el empaquetado de partículas. El análisis del factor de inflación de varianza (VIF) confirmó una fuerte interdependencia entre el contenido de ceniza volante, las fracciones agregadas y la relación agua/aglutinante, mientras que la tasa de sustitución por agregado reciclado permaneció estadísticamente independiente.
4. El análisis de fiabilidad estadística reveló que todos los modelos presentaron un sesgo de predicción muy bajo. La robustez de la estructura desarrollada fue corroborada por el error medio de sesgo de 0,12 MPa de los Árboles Extra Regresor y por las pruebas de hipótesis, que no identificaron una diferencia estadísticamente significativa entre las resistencias predichas y medidas ($p > 0,05$).
5. El análisis mediante el método SHapley Additive ExPlanations (SHAP) identificó la tasa de sustitución por agregado reciclado como el parámetro más influyente, seguida del contenido de cemento y la relación agua/aglutinante. Esta clasificación es físicamente consistente con la teoría del empaquetado de partículas, ya que la calidad del agregado, la disponibilidad de suspensión y la estructura interna del vacío gobiernan conjuntamente el desarrollo de la resistencia a la compresión.
6. El acuerdo significativo entre la importancia de las variables determinadas por SHAP y los principios consolidados de empaquetado demuestra que el aprendizaje automático puede capturar implícitamente los mecanismos del empaquetado de partículas a través de las variables de composición de la mezcla, incluso si la densidad de empaquetado de

partículas o la razón de vacíos no se han estimado explícitamente.

7. La estructura propuesta contribuye al desarrollo sostenible del hormigón y a la evaluación de su rendimiento, al ofrecer un método preciso, estadísticamente robusto y físicamente interpretable para la predicción de la resistencia a la compresión del hormigón autocompactante con agregados reciclados (RA-SCC). Los resultados demuestran que el aprendizaje automático explicable puede actuar como una herramienta fiable para apoyar la toma de decisiones para la dosis sostenible de RA-SCC, tanto en el ámbito de la investigación como en la práctica profesional. El uso de un conjunto de datos compuesto por 365 mezclas, compilado a partir de 102 estudios independientes, refuerza aún más la fiabilidad estadística y la aplicabilidad práctica de la estructura predictiva propuesta.

6.2 Directrices para futuras investigaciones.

Aunque el presente trabajo demuestra que el aprendizaje automático explicable es una herramienta útil para predecir la resistencia a la compresión, existen varias áreas que pueden ser objeto de una investigación más profunda:

- Estudios futuros podrían incorporar la densidad de empaquetamiento de partículas o la razón de vacíos medida experimentalmente como variables de entrada explícitas para fortalecer la relación mecanicista entre las predicciones de aprendizaje automático y el comportamiento físico del empaquetado.
- Para ampliar la generalización de los modelos en diferentes escenarios constructivos, el conjunto de datos puede ampliarse para cubrir un rango más amplio de edades de curado, clases de agregados reciclados y características reológicas del hormigón autocompactante.
- Se pueden investigar enfoques de modelado híbrido, que combinan la teoría del empaquetamiento de partículas con marcos de aprendizaje automático basados en datos, con el objetivo de mejorar aún más la precisión de las predicciones y la interpretabilidad de los modelos.
- Ampliar la metodología propuesta para predecir propiedades relacionadas con la durabilidad, como permeabilidad, contracción y desarrollo de resistencia a largo plazo, ampliaría su aplicabilidad en el diseño de infraestructuras sostenibles.
- Finalmente, los modelos explicables de aprendizaje automático pueden asociarse con técnicas de cuantificación de incertidumbre, permitiendo establecer intervalos de confianza para las predicciones y contribuyendo a una práctica de ingeniería basada en la evaluación de riesgos.
- La investigación futura debería centrarse en validar el marco propuesto de aprendizaje automático utilizando conjuntos de datos experimentales controlados obtenidos de programas de investigación individuales. Estos estudios permitirían una evaluación más precisa del rendimiento de los modelos bajo condiciones consistentes de material y prueba, además de contribuir al desarrollo de herramientas predictivas específicas para cada proyecto.
- Además, los estudios futuros deberían incorporar conjuntos de datos que abarcan múltiples edades de curado, lo que permite predecir la resistencia en función del tiempo. Este aspecto es especialmente importante para sistemas de hormigón que incorporan materiales cementicios suplementarios, como la ceniza volante, en los que el desarrollo de resistencia a largo plazo se extiende más allá del periodo convencional de 28 días.

7. REFERENCIAS

- 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>
- Etxeberria, 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>