

Estratégias de Machine Learning na Otimização de Geopolímeros: Uma Revisão Integrativa na Engenharia Química

Elena Luiza Cezar e Silva; Jorge David Alguiar Bellido

Resumo

A complexidade química inerente e a não-linearidade das reações de álcali-ativação impõem barreiras significativas à otimização experimental de geopolímeros, impulsionando a transição para abordagens computacionais avançadas. Esta revisão integrativa investigou o estado da arte das estratégias de Machine Learning (ML) aplicadas à otimização do processo e das propriedades funcionais de ligantes geopoliméricos. O estudo refinou 2.328 registros brutos para um corpus analítico de 12 estudos de alta fidelidade, triados sistematicamente na base ScienceDirect (2020–2026). A síntese das evidências revelou que algoritmos de Ensemble Learning ($R^2 > 0.96$) estabeleceram o padrão para a predição de propriedades estruturais e ambientais, como a resistência e a Pegada de CO₂. A virada de paradigma reside no uso de modelos híbridos e interpretáveis (SHAP), que permitiram o design prescritivo ao mapear a relação de causa-e-efeito entre as variáveis de Composição Estequiométrica (e.g., relação Si/Al) e o desempenho final, viabilizando a otimização multiobjetivo (NSGA-II) para o design eco-eficiente. Os resultados confirmam que o ML é crucial para o controle de processos, como a reologia (Yield Stress) e o desempenho funcional (adsorção de metais pesados). No entanto, o avanço do campo é limitado pela escassez de datasets padronizados e globais. As perspectivas futuras convergem para a expansão da interpretabilidade, o desenvolvimento de softwares GUI para a acessibilidade industrial e a aceleração da simulação em escala atômica como via para a descoberta de novos materiais sustentáveis.

Palavras-chave: Geopolímeros; Aprendizado de Máquina; Inteligência Artificial; Otimização; Geopolimerização.

Introdução

O setor global de produção de ligantes inorgânicos enfrenta uma pressão sem precedentes para mitigar sua pegada ambiental, dado que a manufatura do Cimento Portland Comum (OPC) é responsável por aproximadamente 8% das emissões globais de CO₂ antropogênico (1)(2). Neste cenário, os geopolímeros e ligantes álcali-ativados consolidaram-se como a alternativa mais promissora. Estes materiais não apenas oferecem reduções de emissões que podem superar 70%, dependendo do precursor utilizado, mas também valorização de múltiplos resíduos industriais, como a lama vermelha (3)(4), rejeitos de cobre (5) e rejeitos de mineração em geral (6), conferindo-lhes novas funcionalidades que transcendem a construção civil. (7,8). Os geopolímeros têm demonstrado potencial como adsorventes ambientais de alta performance (9), filtros para efluentes (10) e agentes de imobilização de metais pesados (11). Entretanto, a transição destes materiais da escala laboratorial para a aplicação industrial em larga escala esbarra na complexidade cinética e estequiométrica de sua síntese. Diferentemente do cimento Portland, cuja hidratação é bem compreendida, a geopolimerização envolve a formação de géis aluminossilicatos (N-A-S-H)

através de reações multicomponentes sensíveis a variáveis de processo, como a molaridade e tipo do ativador, a relação Si/Al, a temperatura e o tempo de cura. Esta complexidade se estende às propriedades funcionais, onde a porosidade, a reologia (viscosidade/tensão de escoamento) e a cinética de adsorção são críticas para aplicações de engenharia, como o bombeamento em minas (12) ou a filtração dinâmica (10).

Tradicionalmente, a otimização dessas formulações tem dependido de abordagens empíricas baseadas em tentativa e erro ou em métodos estatísticos clássicos, como a Metodologia de Superfície de Resposta (RSM). Embora a RSM seja eficaz para sistemas lineares com poucas variáveis de controle, ela frequentemente falha em capturar as não-linearidades e interações de alta ordem inerentes à cinética de ativação alcalina dos geopolímeros, resultando em um consumo excessivo de reagentes e em designs subótimos (7). A variabilidade na composição química das matérias-primas, inerente a resíduos industriais, exacerba este desafio, tornando os modelos cinéticos tradicionais pouco robustos e limitados a condições de contorno específicas.

A última década testemunhou uma mudança de paradigma com a introdução de estratégias de Machine Learning (ML) e Inteligência Artificial na engenharia e ciência de materiais. Impulsionada pelo crescimento de bases de dados (Big Data) e pelo poder de processamento, o ML demonstrou uma capacidade superior de modelar sistemas complexos, superando as barreiras dos modelos físicos e estatísticos tradicionais. Em diversas áreas da Engenharia, algoritmos inteligentes têm revolucionado processos, desde o monitoramento de consistência em linhas de produção de alimentos (13) até a predição rápida de riscos e explosões em clusters químicos (14). Na ciência de materiais, essa revolução se manifestou inicialmente na capacidade de correlacionar a estequiometria dos reagentes e as propriedades finais dos materiais, como resistência mecânica e capacidade de adsorção (15)(9).

Entretanto, apesar da clara promessa e do crescimento exponencial das publicações, a literatura sobre geopolímeros e ML ainda carece de uma análise integrativa que mapeie as estratégias de otimização de forma específica e com foco no processo. Muitas revisões existentes concentram-se predominantemente na comparação de métricas de acurácia preditiva (R^2), dedicando menos atenção à interpretação físico-química dos modelos e à aplicação do ML para otimizar as variáveis de processo em si. A ausência de uma síntese clara sobre como as diversas estratégias de Machine Learning, desde modelos híbridos de otimização até o Deep Learning e análise de interpretabilidade via SHAP, têm sido utilizadas para resolver desafios práticos de engenharia cria uma lacuna para o design de novos materiais. Esta revisão integrativa busca, portanto, discorrer sobre o tema, analisando as estratégias de Machine Learning aplicadas especificamente à otimização das variáveis e dos processos de geopolimerização, fornecendo uma base para o entendimento e apontando caminhos futuros na área da Engenharia Química.

Metodologia

A presente investigação foi delineada como uma Revisão Integrativa (IR) da literatura, método que melhor se adequa à síntese de conhecimentos teóricos e empíricos para proporcionar uma compreensão abrangente de fenômenos complexos, como a otimização de materiais multicomponentes (16). Para a etapa de busca e seleção dos estudos, adotou-se o fluxograma e as diretrizes de reportabilidade do PRISMA 2020 (17), adaptados ao contexto de uma revisão integrativa, a fim de garantir a transparência, rastreabilidade e replicabilidade do processo bibliográfico.

Estratégia de Busca e Fontes de Dados

A busca foi realizada na base de dados ScienceDirect (18), escolhida por sua abrangência e relevância no nicho de periódicos de engenharia química e ciência dos materiais. Para garantir a rastreabilidade e a profundidade da investigação, a estratégia de busca foi formulada com base nos termos booleanos que envolvem ML e geopolímeros, o que retornou um total de 2.328 registros brutos.

A este conjunto inicial, foram aplicados filtros de refinamento automáticos na plataforma, que incluíram o recorte temporal de artigos, a restrição ao tipo de publicação e a seleção da área temática. A aplicação desses filtros reduziu drasticamente o volume de dados para 82 registros, que constituíram a amostra inicial para a triagem manual. Os termos booleanos e os filtros aplicados na plataforma estão descritos na Tabela 1.

Tabela 1. Parâmetros de Busca e Filtros na Base ScienceDirect.

Base de Dados	ScienceDirect (Elsevier)
Termos Booleanos	("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural networks") AND (geopolymer OR "alkali activated")
Área Temática	Chemical Engineering
Recorte Temporal	2020-2026

Processo de Seleção e Fluxo PRISMA

O processo de seleção subsequente seguiu um fluxo de triagem e elegibilidade qualitativas. Inicialmente, os 82 registros foram submetidos à Triagem (Screening) por análise de título e resumo. Nesta fase, aplicou-se os dois critérios de inclusão essenciais: o material principal sob estudo deveria ser um geopolímero e o estudo deveria utilizar ativamente técnicas de Machine Learning ou Inteligência Artificial para modelagem de dados. No entanto, grande parte dos estudos indexados, apesar de satisfazerem os filtros automáticos, não atendiam aos critérios temáticos. Estes foram classificados como "falsos positivos" e excluídos, pois o foco era em cimentos convencionais, híbridos não-geopoliméricos, adsorventes de outra natureza química (como carvão ativado ou biopolímeros) ou zeólitas puras.

Os artigos aprovados na triagem seguiram para a análise de Elegibilidade (Eligibility), realizada por meio de leitura na íntegra. Nesta etapa, foram excluídos os artigos que não

respondiam à pergunta norteadora da revisão, pois, apesar de utilizarem geopolímeros, o estudo não estava voltado à otimização da geopolimerização. Após a aplicação rigorosa destes critérios de fidelidade química e

metodológica, restou um corpus analítico final de 12 artigos. A Figura 1 ilustra o diagrama de fluxo PRISMA, detalhando a redução quantitativa em cada fase do processo de seleção.

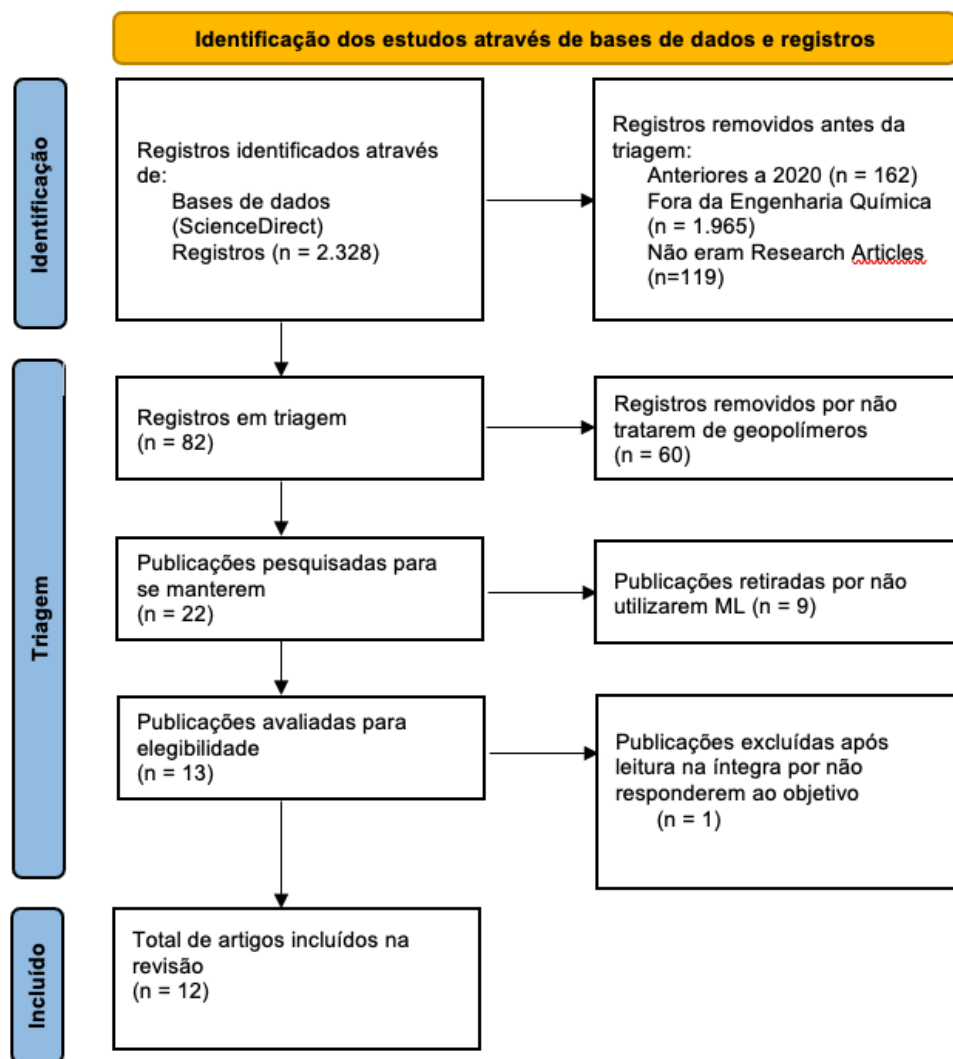


Figura 1. Fluxograma PRISMA 2020 para seleção de estudos (adaptado de Page et al. (3); tradução portuguesa de Abreu et al.).

Resultados e Discussão

Panorama Bibliométrico

Para contextualizar a análise qualitativa e mapear as tendências conceituais dominantes, conduziu-se uma análise bibliométrica assistida pelo software VOSviewer (19) sobre a amostra ampliada de 82 registros, que estão referenciadas. O objetivo desta análise foi visualizar a topologia intelectual do campo de pesquisa após a aplicação dos filtros automáticos.

O mapeamento foi baseado na co-ocorrência de termos (palavras-chave e expressões) extraídos exclusivamente de títulos e resumos. Para o processamento, foi utilizado o método de contagem completa (full counting), sendo considerado termos com frequência mínima de cinco ocorrências. A rede foi construída selecionando-se os 60% mais relevantes segundo o algoritmo de força de associação do software, após a depuração manual de termos genéricos ou metodológicos sem valor semântico específico. A Figura 2 apresenta a rede bibliométrica resultante, revelando os clusters temáticos que dominam a produção científica no nicho de otimização de

geopolímeros por Machine Learning.

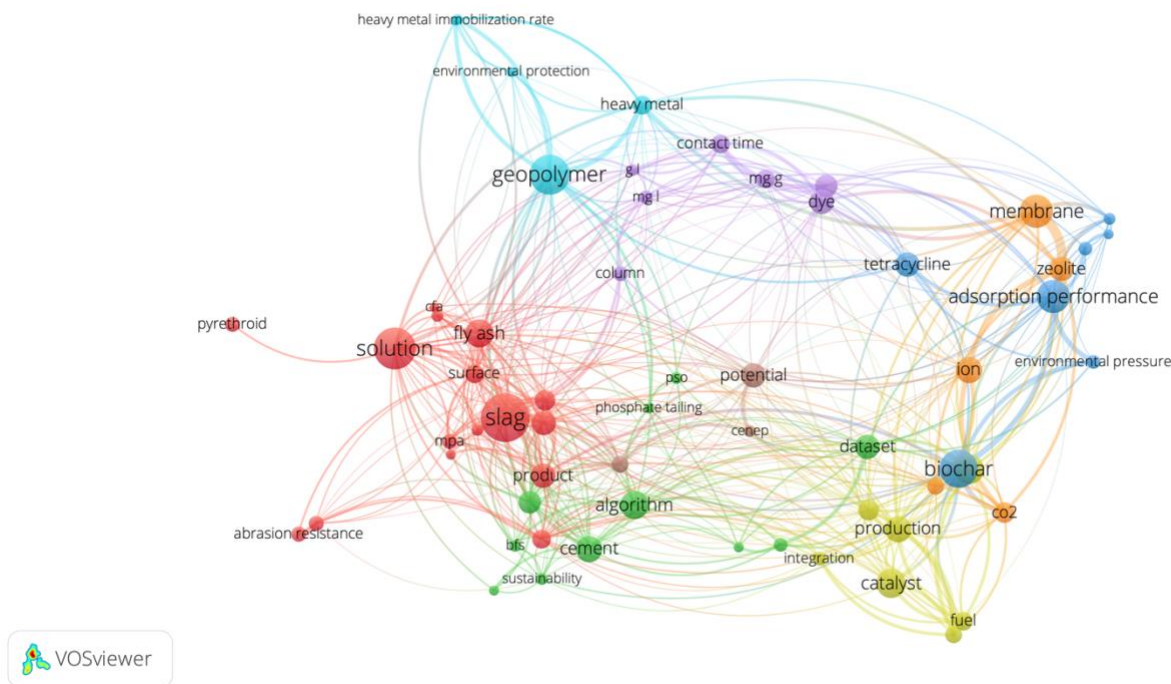


Figura 2. Análise de co-ocorrência de termos por meio do VOSviewer (19).

A topologia da rede bibliométrica, analisada através da métrica de força de associação (Link Strength) entre os termos, revela que a aplicação de Machine Learning na geopolimerização não é dispersa, mas sim ancorada em nós centrais de alta conectividade que unem a química de síntese à engenharia de processos. A estrutura da rede evidencia que o campo evoluiu da simples caracterização de materiais para uma abordagem integrada de otimização multivariável.

A análise de centralidade aponta o termo *solution* (solução) como o nó de maior influência em toda a rede, apresentando a maior força total de ligação (207). A forte co-ocorrência deste termo com variáveis como *activator*, *fly ash* e *slag* confirma que a estequiometria da solução ativadora é a variável de processo dominante. Diferentemente de abordagens estatísticas simples, a densidade de conexões em torno de *solution* sugere que os modelos de Machine Learning estão sendo treinados prioritariamente para decodificar as não-linearidades complexas da ativação alcalina, validando a premissa de que a otimização da mistura química é o gargalo que a inteligência artificial busca resolver.

No que tange à aplicação metodológica, o termo *algorithm* emerge como um conector recente e vital (ano médio de publicação ≈ 2024.8), estabelecendo pontes diretas tanto com propriedades mecânicas tradicionais quanto com novos requisitos de dados (*dataset*). Entretanto, a análise das conexões individuais revela uma tendência surpreendente e específica para a Engenharia Química: a ligação mais forte de toda a rede não está relacionada ao concreto, mas sim à conexão entre *membrane* e *zeolite* (Link Strength = 99). Este dado quantitativo prova que existe um nicho de pesquisa consolidado e de alta intensidade focado na síntese de materiais porosos para processos de separação, desviando o foco exclusivo da construção civil para aplicações funcionais de alto valor agregado.

Esta vertente funcional é corroborada pela alta conectividade do termo *adsorption performance* (Força Total: 191), que atua como um hub ligando o geopolímero a termos ambientais como *biochar*, *heavy metal* e *carbon capture*. A topologia da rede demonstra, portanto, que o Machine Learning atua como a ferramenta transversal que viabiliza duas frentes simultâneas: a otimização de matrizes densas para resistência mecânica e, de forma cada

aluminossilicatos reativos mais comuns. Contudo, o corpus avança para precursores de maior complexidade e variabilidade química, incluindo Rejeitos de Fosfato, Cinzas Volantes de Incineração de Resíduos Sólidos Urbanos (MSWIFA) e Rejeitos de Carbetto. A adoção desses resíduos complexos, cuja composição química é inerentemente variável, reforça a premissa de que a otimização por métodos empíricos é inviável, tornando a modelagem via Machine Learning uma necessidade para garantir a consistência do produto. Um agrupamento dos materiais precursores, função e foco da otimização pode ser visto na Tabela 2.

Tabela 2. Agrupamento do Corpus Analítico por Material Precursor, Função e Foco da Otimização.

Artigos	Categoria Principal	Precursos Dominantes	Foco da Otimização com ML
(9) (10) (20) (21) (22)	Adsorvente/ Funcional	Aluminossilicatos Diversos, Escória BOFS, Metacaulim/Zeólita	Capacidade de Adsorção, Eficiência de Degradação, Parâmetros de Breakthrough (Filtração), Porosidade Microestrutural
(7)(15) (8)	Cimento Estrutural	Cinza Volante (FA), Rejeitos de Fosfato, Metacaulim	Resistência à Compressão (CS), Pegada de CO ₂ , Resistência à Flexão
(12) (23)	Engenharia Especializada	Escória (CS, BFS), Múltiplos Resíduos Industriais	Reologia (Tensão de Escoamento τ_0), Cinética de Cura (UCS)
(11)	Estabilização/ Imobilização	Cinzas Volantes de Incineração (MSWIFA)	Taxa de Imobilização de Metais Pesados

(24)	Simulação/ Mecanismo	(Teórico, Simulação de Átomos: Ca, Si, Al, Na, K)	Mecanismo de Polimerização, Energia Atômica (Validação de Modelos Surrogados DFT)
------	-------------------------	---	---

Enquanto 41,7% dos estudos se concentram no Geopolímero para aplicações de Cimento Estrutural ou Engenharia Especializada, como a Pasta de Retaludamento Cementada (CUFTB) (12) (23), onde a resistência é ligada à bombeabilidade, a maioria, correspondendo a 50,0% dos artigos, visa aplicações funcionais e ambientais. Esta vertente funcional é a mais alinhada com as tendências do VOSviewer, utilizando o geopolímero para adsorção ambiental de metais pesados (9), Solidificação/Estabilização de resíduos tóxicos (11), Filtração em Coluna (Parâmetros de Breakthrough) (10) e até mesmo como Adsorvente-Fotocatalisador (GP/ASN) (22). A emergência do Geopolímero Impresso em 3D para aplicações de adsorção dinâmica ilustra o avanço no design estrutural para funcionalidade (20).

Um único estudo, representando 8,3%, foca na Simulação e Mecanismo de polimerização em escala atômica (24). Embora numericamente pequeno, este estudo é metodologicamente crucial, pois visa decodificar a formação do gel aluminossilicato (N-A-S-H/C-A-S-H), a base de todos os materiais estruturais e funcionais. Deste modo, o mapeamento funcional valida a intersecção da pesquisa com as demandas da Engenharia Química, onde o ML é aplicado para resolver não apenas o desafio da resistência, mas sim o controle preciso das propriedades funcionais complexas, como a reologia (tensão de escoamento) e a porosidade microestrutural, essenciais para a performance ambiental e industrial. A distribuição percentual do foco de pesquisa está sumarizada na Figura 4.

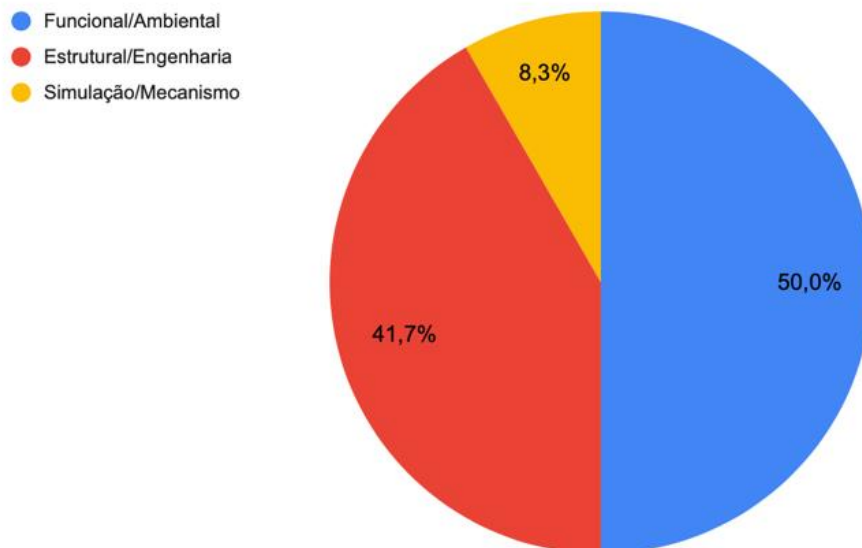


Figura 4. Gráfico da Síntese do Foco de Pesquisa.

Datasets

A análise dos estudos que aplicam Machine Learning (ML) na otimização de geopolímeros revela que a escassez, a inconsistência e a origem dos datasets representam um dos gargalos metodológicos para a plena consolidação e generalização dos modelos preditivos. O desafio não reside apenas no volume de dados, mas na sua qualidade e na intensa curadoria necessária, o que limita o avanço do campo e a transição para aplicações em larga escala.

A maioria das investigações com foco em modelos de ML que buscam generalização baseia-se em datasets de pequeno ou médio porte, compilados reativamente a partir de publicações acadêmicas abertas. Exemplos disso incluem estudos sobre a otimização da capacidade de adsorção de metais pesados, que utilizaram 267 amostras (9), e a previsão da taxa de imobilização em resíduos tóxicos, com 196 amostras (11). Tais volumes são insuficientes para que o ML capture a vasta complexidade e as interações não-lineares inerentes a um sistema multicomponente como a geopolimerização. A fonte destes dados, muitas vezes, é a extração manual e visual de informações de gráficos e figuras publicadas, utilizando ferramentas como o Web Plot Digitizer. Esta falta de dados robustos não é exclusiva dos geopolímeros; desafios similares na modelagem de propriedades, como a condutividade térmica de materiais de construção porosos, também são limitados pela falta de datasets abrangentes (25).

Para tornar estes dados dispersos utilizáveis, um esforço substancial de curadoria e normalização é exigido, como a conversão de unidades ou o cálculo de razões molares essenciais (como a razão $\text{SiO}_2/\text{Al}_2\text{O}_3$) e de proporções de ativadores, uma etapa fundamental para a engenharia do dataset (9)(11). O maior dataset referenciado no corpus, utilizado para otimização multiobjetivo de concreto geopolimérico, contendo 799 misturas (7), ilustra a limitação mais crítica: a esparsidade. O estudo revela uma carência aguda de dados para propriedades de longo prazo, com apenas 38 amostras disponíveis para a resistência à compressão em 90 dias e um volume crítico de apenas 7 amostras para 365 dias. Essa falta de dados de durabilidade é um problema crônico na ciência de materiais e leva à dependência de técnicas de imputação (preenchimento de dados faltantes com algoritmos como KNN e XGBoost) (7), o que pode introduzir incertezas e reduzir a robustez dos modelos preditivos.

Finalmente, a aplicabilidade industrial do ML é diretamente mitigada pelo fato de que os datasets proprietários de alta fidelidade e grande escala, gerados em ambientes industriais e por empresas, permanecem inacessíveis ao meio acadêmico. Estes dados, que englobam o monitoramento rigoroso e de longo prazo das variáveis de processo e do desempenho do material em escala real, são frequentemente mantidos em sigilo devido à sua vantagem competitiva. A ausência de um Repositório de Dados de Acesso Aberto e Padronizado (Open-Access Standardized Repository) impede que a comunidade científica possa treinar modelos robustos e generalizáveis.

No entanto, a recente ascensão do termo dataset como um nó metodológico na análise bibliométrica aponta para o reconhecimento da comunidade sobre esta limitação, sugerindo que a criação e a padronização destas bases de dados será a próxima fronteira de inovação para sustentar a fase de crescimento exponencial do ML em Engenharia Química.

Modelos de ML Testados

A estratégia de modelagem no campo de geopolímeros assistida por Inteligência Artificial (IA) é marcada por uma intensa e diversificada experimentação de algoritmos. A escolha dos modelos de Machine Learning (ML) transcende as abordagens estatísticas tradicionais, focando em arquiteturas robustas para lidar com a não-linearidade química, a alta dimensionalidade e a variabilidade temporal dos dados de síntese. A seleção de algoritmos reflete uma busca contínua pelo melhor equilíbrio entre acurácia preditiva, interpretabilidade e aplicabilidade industrial. A Tabela 3 detalha a taxonomia dos modelos de Machine Learning testados nos estudos, agrupados conforme suas metodologias.

Tabela 3. Taxonomia e Descrição dos Algoritmos de Machine Learning Aplicados à Modelagem e Otimização de Geopolímeros na Literatura Revisada.

Grupo Metodológico	Algoritmo	Descrição e Aplicação no Estudo de Geopolímeros		
Ensemble Learning (Modelos Baseados em Árvores)	Random Forest Regression (RF)	Modelo Ensemble que utiliza múltiplas árvores de decisão treinadas de forma independente, combinando seus resultados para melhorar a precisão e evitar o overfitting.		Boosting (XGBoost)
	Extremely Randomized Trees Regression (ET)	Variação do RF que introduz maior aleatoriedade nas divisões das árvores. É conhecido por melhorar a velocidade de treinamento e a generalização do modelo.		Decision Tree Regression (DTR)
	Gradient Boosting (GB)	Método Ensemble que constrói árvores sequencialmente, onde cada nova árvore é ajustada para corrigir os erros (residuals) das árvores anteriores. É altamente eficaz em problemas de regressão.		Voting/ Stacking Regressor
	Extreme Gradient	Estrutura otimizada e eficiente do Gradient Boosting, conhecida por sua escalabilidade e		Bootstrap Aggregating Regression (BA)
			Modelos Kernel, Probabilísticos e de Instância	AdaBoost Regression
				Random Forest Classifier
				Support Vector Machine Regression (SVR)
				Gaussian Process Regression (GPR)

velocidade de processamento, frequentemente usada para competir em benchmarks de alta performance.

Modelo fundamental que estabelece uma série de regras de decisão hierárquicas, ramificando-se a partir dos features de entrada para prever um valor alvo.

Meta-heurística que combina as previsões de múltiplos modelos subjacentes (ou preditores fracos) por meio de voto ou aprendizado para obter um resultado final mais acurado e estável.

Técnica de reamostragem (com substituição) usada para criar múltiplas versões de um conjunto de dados. É a base para o Random Forest e é usada para reduzir a variância do modelo.

Algoritmo de Boosting que ajusta o peso das amostras incorretamente previstas em iterações subsequentes para concentrar o aprendizado onde o modelo é mais fraco.

Versão do RF projetada para problemas de classificação, ou seja, prever uma classe ou categoria de saída em vez de um valor contínuo.

Versão de regressão da Máquina de Vetores de Suporte (SVM). Utiliza funções de kernel para mapear dados em um espaço de alta dimensão, o que permite modelar relações não-lineares complexas.

Modelo de regressão probabilístico que assume que as variáveis seguem um processo gaussiano. Fornece uma previsão com

Redes Neurais (ANN e Deep Learning)		uma medida de incerteza (variância) associada.	Otimização e Híbridos (Prescritivos)		Algoritmo Evolucionário (Genético) de Otimização Multiobjetivo (MOO), focado em encontrar o conjunto de soluções ideais (Pareto Front) para problemas com objetivos conflitantes.
	K-Nearest Neighbors Regression (KNN)	Método não-paramétrico que prevê o valor de um ponto com base na média dos valores dos K pontos de dados mais próximos no espaço de features.		NSGA-II (Non-dominated Sorting Genetic Algorithm II)	
	Long Short-Term Memory (LSTM)	Rede Neural Recorrente (RNN) especializada em modelar e prever sequências temporais e dependências de longo prazo em dados sequenciais.		Particle Swarm Optimization (PSO) / Genetic Algorithm (GA)	Meta-heurísticas de otimização que simulam o comportamento de grupos ou a evolução biológica para encontrar soluções ideais em um grande espaço de busca (search space).
	Feedforward Neural Network (FNN) / Artificial NN (ANN)	Rede Neural Artificial (ANN) simples, onde a informação se move apenas em uma direção (para frente), passando por camadas de entrada, ocultas e de saída. Usada como preditor básico ou componente híbrido.		LSTM + XGBoost (Híbrido)	Combinação de dois modelos que explora a capacidade do LSTM de modelar sequências temporais e a alta acurácia de regressão do XGBoost para melhorar o desempenho preditivo em cenários complexos.
	Adaptive Neuro-Fuzzy Inference System (ANFIS)	Sistema híbrido que combina a estrutura de uma Rede Neural com a lógica Fuzzy (conjuntos nebulosos e regras "se-então") para maior interpretabilidade e capacidade de modelar sistemas baseados em regras linguísticas.			
	Convolutional Neural Network (CNN)	Arquitetura de DL com camadas convolucionais, mais conhecida por processamento de imagem, mas aplicada a dados sequenciais para extrair features hierárquicos.			
	Gated Recurrent Unit (GRU) / Transformer / Dense NN	Outras arquiteturas de redes neurais profundas testadas em benchmarks para modelagem de alta complexidade em dados de materiais.			
	CeNEP (Neuroevolution Potential - MLIP)	Modelo de Potencial Interatômico baseado em Machine Learning (MLIP), focado em simulação em escala atômica para decodificar a energia de formação e as interações entre átomos.			

A modelagem dos sistemas geopoliméricos, que possuem dependência complexa de múltiplas variáveis estequiométricas e de processo e, frequentemente, existe a carência de dados de longo prazo (7), exigiu a mobilização de um arsenal diversificado de algoritmos de ML. Esta busca por modelos que pudessem abstrair a complexa cinética da ativação alcalina é, em parte, uma resposta direta à ameaça do overfitting. Este fenômeno ocorre quando o modelo, em vez de aprender a relação funcional generalizável, memoriza o ruído e os padrões altamente específicos do conjunto de treinamento. Em um campo onde os datasets agregam informações de precursores quimicamente variados e protocolos laboratoriais distintos, como na síntese de geopolímeros a partir de resíduos industriais, o overfitting resulta em modelos que falham criticamente ao serem aplicados a novas misturas reais, inviabilizando a transição para a escala industrial.

Para mitigar esse efeito e garantir a generalização, a literatura priorizou os Modelos de Ensemble e Boosting. Esses algoritmos, como o Random Forest, Extra Trees (ET) e Gradient Boosting (GB), são inerentemente robustos porque combinam múltiplas árvores de decisão, estabilizando a função preditiva e filtrando o ruído inconsistente. O Extra Trees (ET), por exemplo, foi selecionado por demonstrar desempenho superior na previsão da capacidade de adsorção de metais pesados (9), uma área de alta não-linearidade, devido à sua capacidade de generalização aprimorada. O XGBoost e o Gradient Boosting (GB) provaram ser cruciais para atingir alta

acurácia na previsão da taxa de imobilização de metais pesados (11) e na modelagem da pegada de CO₂ (7)(15). A escolha do Voting Ensemble para prever a resistência à compressão ilustra a estratégia de diversificação para maximizar a estabilidade do preditor final (7).

Entretanto, os modelos Ensemble e Boosting não são suficientes para resolver problemas que envolvem dependências temporais ou análise em múltiplas escalas. Para isso, as Redes Neurais Artificiais (ANN) e o Deep Learning (DL) foram essenciais. Arquiteturas de DL mais complexas, como Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), Transformers e Dense Neural Networks, foram testadas em benchmarks (7) para capturar features hierárquicos e não-lineares, mas as redes com capacidade de memória se destacaram. A Long Short-Term Memory (LSTM), especializada em modelar sequências temporais, foi fundamental em modelos híbridos que previram o desempenho ao longo do tempo de cura, atuando em conjunto com o XGBoost para modelar a pegada de CO₂ com R² de 0.996 (7). A simplicidade da Feedforward Neural Network (FNN) também foi capitalizada em modelos híbridos, como o acoplamento com o Particle Swarm Optimization (PSO-FNN), para otimizar propriedades do processo, como o estresse de escoamento, vital para a reologia e o bombeamento em minas (12). Em uma aplicação de ponta na Engenharia Química, o potencial CeNEP (Cementitious Neuroevolution Potential), um modelo de Potencial Interatômico foi desenvolvido para realizar simulações atômicas e em escala quântica da formação dos produtos de hidratação (como C-S-H e geopolímeros) de forma muito mais eficiente do que a Teoria do Funcional da Densidade (DFT), resolvendo o problema da escala de simulação (24).

Além dos modelos de DL, os Modelos Probabilísticos e Kernel também foram estratégicos. A Regressão por Processo Gaussiano (GPR) é escolhida quando o problema transcende a mera acurácia e exige a quantificação da incerteza da previsão, fornecendo o intervalo de confiança ao lado da previsão da Resistência à Compressão Não Confinada (UCS) (23) ou da pegada de CO₂ (15). Já o SVR (Support Vector Regression) é valorizado pela sua capacidade, via kernel trick, de mapear relações altamente não-lineares em espaços de características de alta dimensão, garantindo que as interações complexas sejam modeladas (9)(7). Finalmente, a fusão dessas abordagens culmina nas Estratégias Híbridas de Otimização, onde modelos de ML preditivos (como LSTM + XGBoost) agem como substitutos (surrogate models) para a complexa função de desempenho do material. Essa combinação, alimentando meta-heurísticas como o NSGA-II (Non-dominated Sorting Genetic Algorithm II), transforma o ML de uma ferramenta de previsão para uma ferramenta de prescrição, permitindo a geração automatizada de

Soluções de Pareto-Ótimo que solucionam os trade-offs entre objetivos conflitantes (ex: resistência versus CO₂) (7). Assim, a escolha dos algoritmos é uma resposta tática e diversificada às múltiplas complexidades, visando não apenas a acurácia, mas também a generalização, a modelagem temporal, a quantificação de incerteza e, sobretudo, o design prescritivo de materiais sustentáveis.

Algoritmos de Interpretabilidade

Os Algoritmos de Interpretabilidade representam uma etapa de enorme valor agregado do Machine Learning na Engenharia Química, pois resolvem o dilema do "modelo caixa-preta". Um modelo de ML, mesmo que alcance um R² próximo de 1, é de utilidade limitada se não puder explicar por que chegou àquela previsão. Em sistemas complexos como os geopolímeros, onde a otimização exige a manipulação de variáveis químicas e de processo, a capacidade de interpretar o modelo é essencial para a descoberta científica e para a confiança regulatória; ou seja, saber a causa de um desempenho ótimo permite aos engenheiros químicos criar novas misturas sem depender apenas da ferramenta computacional.

O grande ponto de virada nessa área é o algoritmo SHAP (SHapley Additive Explanations). O SHAP baseia-se na Teoria dos Jogos e atribui a cada variável de entrada (feature) um valor de contribuição para a previsão final. O seu valor não está apenas em ranquear a importância dos fatores (Feature Importance), mas em mostrar se a contribuição foi positiva ou negativa e qual a sua magnitude, permitindo uma interpretação fina. Nos artigos analisados, o SHAP foi crucial para revelar que a relação Si/Al e a temperatura inicial de cura são os fatores dominantes na determinação da resistência precoce dos geopolímeros, enquanto métricas relacionadas ao sódio controlam a pegada de CO₂ (7). No estudo de adsorção, o SHAP (juntamente com o Partial Dependence Plot - PDP, que mostra o efeito marginal de uma variável no resultado) confirmou o tempo de cura e a dosagem do adsorvente como os mais influentes na capacidade de adsorção de metais pesados (9).

Além do SHAP, outras estratégias de interpretabilidade foram empregadas. A Análise de Sensibilidade é uma abordagem mais tradicional, vista em modelos como GPR, SVR ou ANN/ANFIS (8). Ela mede o quão sensível é a previsão de saída a pequenas variações nas variáveis de entrada. Embora não seja tão rica quanto o SHAP, ela é fundamental para identificar quais parâmetros requerem controle mais rigoroso durante o processo de fabricação.

Por outro lado, alguns modelos de alta performance, como as arquiteturas de Deep Learning (LSTM, FNN), são historicamente mais difíceis de interpretar. Entretanto, a pesquisa avançou para integrar a interpretabilidade nesses

modelos. Um exemplo notável é o potencial CeNEP (Cementitious Neuroevolution Potential) (24), que, embora seja um modelo de ML para simulação atômica, fornece interpretabilidade na forma de Análise de Força/Energia com fidelidade de Density Functional Theory (DFT). Essa é uma forma de interpretabilidade em escala atômica, mostrando como as forças interatômicas e as energias de formação contribuem para a estabilidade estrutural do geopolímero, ligando a macropropriedade (resistência) à nanoestrutura.

Finalmente, o Conjunto Pareto-Ótimo (Pareto-Optimal Set) gerado pelo algoritmo de otimização multiobjetivo NSGA-II (7), embora seja um output de otimização, funciona como uma interpretabilidade prescritiva. Ele não explica por que um modelo preditivo funcionou, mas sim qual trade-off o engenheiro deve aceitar entre objetivos conflitantes. Este conjunto de soluções fornece um mapa de decisão claro, traduzindo as complexas relações do modelo de ML em um guia de design de material acionável. Em suma, o SHAP e o PDP são essenciais para entender quais e como as variáveis afetam a previsão (causa e efeito), enquanto o NSGA-II e os Potenciais

Interatômicos fornecem a interpretabilidade prescritiva e a atômica, respectivamente.

Inputs e Outputs

As variáveis de controle, que constituem as Variáveis de Entrada (Inputs), são agrupadas em categorias que representam as etapas manipuláveis do processo, como a Composição Estequiométrica (o design químico da mistura), as Condições de Processo e Cura (a cinética da reação) e as Condições de Aplicação (o ambiente de uso final). Por sua vez, as Variáveis de Saída (Outputs) traduzem o valor do material em métricas de engenharia mensuráveis, englobando o Desempenho Mecânico e Estrutural, o Desempenho Funcional e Ambiental e as Propriedades de Processo e Reologia (necessárias para o scale-up industrial). Essa categorização, ilustrada na Figura 5, serve como base para a análise preditiva e prescritiva, permitindo que os modelos de ML mapeiem precisamente a influência de cada parâmetro de controle sobre as propriedades finais, conforme explorado detalhadamente nos estudos revisados.

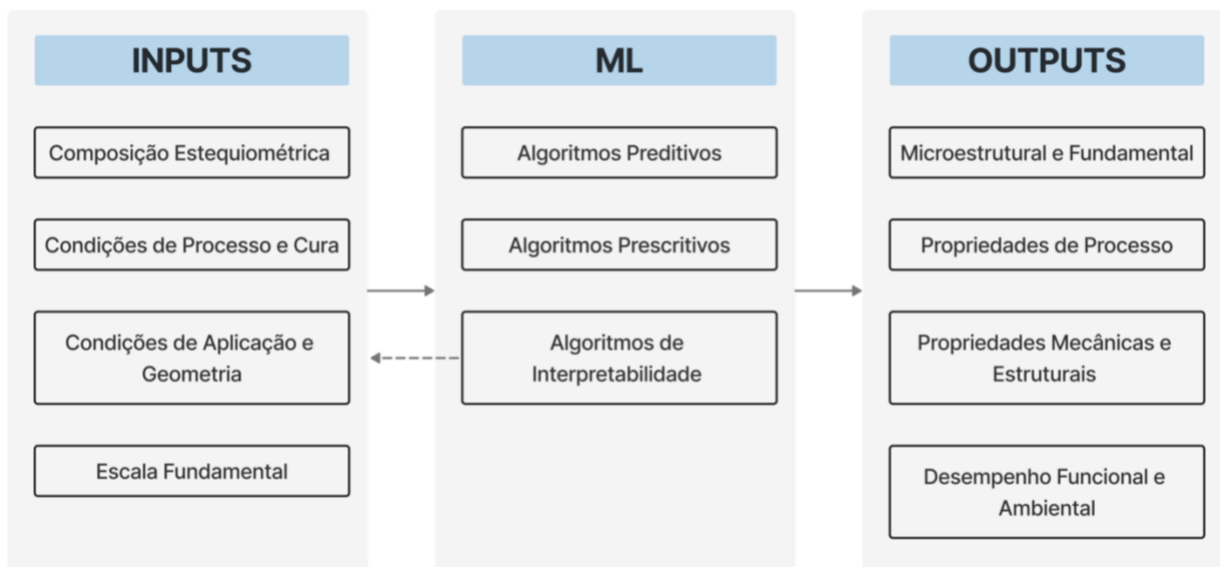


Figura 5. Fluxograma Conceitual: Mapeamento dos Inputs e Outputs de Otimização de Geopolímeros via Machine Learning.

A macro-categoria Composição Estequiométrica reflete o coração do design do material, pois define a reatividade e a estrutura final. As variáveis mais utilizadas são as relações molares, que governam a polimerização. Os modelos incluem as relações $n(\text{SiO}_2)/n(\text{Al}_2\text{O}_3)$, $n(\text{Na}_2\text{O}/\text{SiO}_2)$ e $n(\text{H}_2\text{O}/\text{Na}_2\text{O})$ (9). Os modelos de otimização de concreto geopolimérico incorporam até 11 variáveis de traço (como a relação Si/Al e a relação Na₂O/aglomerante) (7), e integram a composição detalhada dos resíduos, como a porcentagem de Fe₂O₃ e Na₂O nas matérias-primas (11), bem como as proporções

de escória de aço, cinza volante e gesso de dessulfuração utilizadas no cimento de preenchimento (23).

As Condições de Processo e Cura controlam a cinética de reação e são vitais para o desenvolvimento de propriedades. O Tempo de Cura (t) e a Temperatura de Cura (T. Cura) são essenciais, pois influenciam a taxa de dissolução das fases precursoras e a polimerização, estando presentes na maioria dos modelos (9)(11)(7). Variáveis de processo físico, como o Conteúdo Sólido (%) e a Proporção Aglomerante/Rejeito (B/T), são inputs

cruciais para a otimização da reologia e fluidez de pastas cimentícias utilizadas em mineração (12) (23).

Nas Condições de Aplicação e Geometria, as variáveis são específicas para o ambiente de uso final. Em remediação, isso inclui o pH da solução e a Concentração Inicial do contaminante (mg/L), fatores que governam a eletrostática da superfície do adsorvente (9) (11) (22). A Dosagem do Adsorvente e a Morfologia (3D vs. Pó) do compósito também são inputs importantes, afetando a eficiência de degradação em sistemas dinâmicos (9) (22).

A Escala Fundamental destina-se a modelos avançados que exploram as interações em nível atômico. Aqui, os inputs são as Posições e Forças Atômicas derivadas de cálculos quânticos, utilizadas para treinar potenciais interatômicos de Machine Learning (como o CeNEP) que simulam a energia e a estabilidade dos produtos de hidratação (24).

As variáveis de saída são as metas de previsão e otimização. A macro-categoria Desempenho Mecânico e Estrutural foca na validação para a engenharia civil e construção. O principal output é a Resistência à Compressão (CS) em diferentes idades de cura (7) e a Resistência à Compressão Não Confinada (UCS), métrica crítica para o controle de qualidade de materiais de preenchimento de minas (23).

O Desempenho Funcional e Ambiental quantifica a sustentabilidade e a capacidade de remediação. O output de maior impacto é a Pegada de CO₂ (em kg/m³) para otimização multiobjetivo, visando a redução de emissões (7)(15). Em remediação, os modelos preveem a Capacidade de Adsorção (Q_e) de metais pesados (9) e a Taxa de Imobilização (%) em resíduos perigosos (11). Para novos fotocatalisadores, a Eficiência de Degradação de antibióticos é o output funcional de interesse (22).

As Propriedades de Processo e Reologia são essenciais para o scale-up industrial e a Engenharia de Processos. O Yield Stress é o output reológico modelado para garantir a fluidez e a capacidade de bombeamento de pastas concentradas (12). Em filtração, a previsão do Break Point e da Capacidade de Saturação é vital para determinar a vida útil e a eficiência operacional de filtros e adsorventes (10).

Por fim, a Microestrutural e Fundamental abrange os outputs em escalas não-macroscópicas. O ML pode prever a Energia Total e Força Atômica em simulações, validando a precisão dos modelos de potencial interatômico (24). Além disso, a análise de imagens (por exemplo, SEM e MIP) é traduzida pelo ML para outputs como Pixels Segmentados (poro vs. sólido), auxiliando na análise da microestrutura e porosidade do material (10).

Impacto na Otimização

A aplicação de Machine Learning na otimização de geopolímeros resultou em um salto de precisão e insight científico que está redefinindo as fronteiras da Engenharia Química e de Materiais. De forma geral, os melhores modelos alcançaram uma precisão preditiva que consistentemente superou R-quadrado de 0.96 (7)(15) (23)(9), com alguns modelos de Deep Learning atingindo R-quadrado próximo da unidade em tarefas específicas (22)(24). Essa alta acurácia permitiu uma redução significativa do erro, como demonstrado na otimização de objetivo duplo, onde o erro de previsão na resistência à compressão e flexão foi reduzido a 3.64% e 11.24%, respectivamente, diminuindo drasticamente o custo e o tempo de experimentação empírica (8).

O impacto na Engenharia Civil e Estrutural é marcado pela transição para o design eco-eficiente. A otimização multiobjetivo via NSGA-II permitiu identificar Soluções de Pareto-Ótimo, que são receitas que equilibram a máxima resistência com a mínima pegada de CO₂, como misturas que alcançam 65 MPa de resistência com emissões de 30 kg CO₂/m³ (7). O ML também foi essencial na modelagem da cinética de endurecimento, com o modelo híbrido GA-GPR atingindo R² de 0.9885 e validando a significância de variáveis de processo como a Temperatura de Cura para otimizar o ciclo de engenharia (23). Além disso, o GPR demonstrou ser a melhor ferramenta para otimizar o CO₂ ao validar a significância do teor de Superplastificante (15).

Na Engenharia Química Ambiental, o ML validou o design funcional dos geopolímeros como adsorventes e agentes de imobilização. O modelo Extra Trees (ET), com R² de 0.99, identificou que o Tempo de Cura e a relação n(Na₂O/SiO₂) são os fatores mais críticos para maximizar a capacidade de adsorção (9). Na imobilização de resíduos perigosos, o Gradient Boosting (GB) (R² de 0.9214) otimizou a formulação ao correlacionar o sucesso com a densidade da matriz, descobrindo que o baixo teor de Fe₂O₃ e a alta relação n(Al₂O₃)/n(SiO₂) maximizam a imobilização (11). Em aplicações avançadas, o Deep Learning validou o design estrutural do GP/ASN (compósito fotocatalítico) e otimizou as condições operacionais para maximizar a degradação dinâmica de antibióticos (22). Modelos ML também foram usados como ferramentas de caracterização para otimizar a cinética de adsorção e prever o tempo de vida do filtro (Break Point e HUNB) ao quantificar a microestrutura (Pixels Segmentados) (10).

Em Engenharia de Processos e Fundamental, a precisão do ML atacou desafios práticos de scale-up e teoria. O modelo PSO-FNN (R² 0.979) otimizou a composição para garantir a bombeabilidade e trabalhabilidade (controlando

o Yield Stress), uma etapa crucial para o transporte industrial de pastas concentradas (12). No nível mais fundamental, o ML (com MAE de 0.1973) atingiu precisão atômica para otimizar a capacidade de simulação do mecanismo químico de síntese (ligações cruzadas e conteúdo de água) do geopolímero, ligando a teoria atômica ao desempenho macroscópico (24). Em suma, a conclusão chave dos estudos é que o Machine Learning transcendeu a mera previsão de desempenho, tornando-se uma ferramenta prescritiva que fornece a base científica para o design de materiais mais sustentáveis e funcionais.

Principais Contribuições

Na área de Alta Precisão e Sustentabilidade Estrutural, a contribuição central reside na capacidade de modelos híbridos e de Ensemble de preverem o desempenho com extrema acurácia. Por exemplo, a Otimização Multiobjetivo (via NSGA-II) tornou possível o design de receitas eco-eficientes, o que valida a viabilidade técnica de misturas de alta resistência com pegada de carbono drasticamente reduzida (7). Essa precisão se estende ao planejamento de longo prazo, onde o modelo GA-GPR se provou essencial para a previsão precisa da Resistência à Compressão Não Confinada (UCS) de longo prazo, um requisito vital para o planejamento e a segurança de minas, modelando com sucesso a cinética de endurecimento do material (23). O GPR também foi validado como a melhor ferramenta para otimizar o CO₂ em função de variáveis como a Temperatura de Cura e o teor de Superplastificante, reforçando o design orientado a dados de baixo carbono (15).

Em Otimização Funcional e Prescritiva, o ML agiu como um decodificador das relações causa-efeito em aplicações ambientais e de processo. O Extra Trees (ET), com R² de 0.99, validou que o Tempo de Cura e a relação n(Na₂O/SiO₂) são os fatores mais críticos para maximizar a capacidade de adsorção de metais pesados, fornecendo um modelo preditivo robusto para o design funcional de sistemas de tratamento de água (9). Na área de imobilização, o Gradient Boosting (GB) otimizou a formulação ao correlacionar a imobilização máxima com o baixo teor de Fe₂O₃ e a alta relação n(Al₂O₃)/n(SiO₂), permitindo o controle de qualidade e a otimização do design de matrizes de imobilização de resíduos (11). O ML também modelou e otimizou o uso de aditivos, como o agente poroso H₂O₂, para maximizar a capacidade de adsorção e o tempo de vida do filtro (HUNB) (10), e validou o design funcional de geopolímeros modificados para otimizar a degradação dinâmica de antibióticos (22).

As contribuições para Apoio a Processos, Escalas e Acessibilidade garantiram a transição da pesquisa para a indústria e a teoria fundamental. Para o scale-up, o modelo PSO-FNN (R² de 0.979) resolveu um gargalo industrial ao

otimizar a composição para garantir a bombeabilidade e trabalhabilidade da pasta, controlando a tensão de escoamento, que é crucial para a aplicação em larga escala (12). Para aumentar a acessibilidade e o uso prático, os estudos não apenas geraram modelos, mas também ofereceram o fornecimento de um software GUI (Interface Gráfica) para otimização rápida do design por engenheiros e pesquisadores (9) (11). No campo da pesquisa fundamental, o ML foi usado como uma ferramenta de caracterização para quantificar a microestrutura e a cinética de difusão (10) e, mais notavelmente, permitiu a aceleração da simulação de mecanismos de síntese em escala atômica (milhões de átomos), atingindo precisão quântica para otimizar a capacidade de simulação do mecanismo químico de síntese do geopolímero (24).

Limitações

Apesar do avanço na precisão preditiva, as aplicações de Machine Learning na otimização de geopolímeros enfrentam um conjunto de limitações que convergem principalmente para a qualidade, quantidade e complexidade dos dados de treinamento.

Um desafio recorrente em toda a literatura é a insuficiência de padronização e a baixa quantidade de dados experimentais para o modelo (como o modelo Scheffé) (9) (8), um problema que se torna mais agudo devido ao fato de os datasets serem frequentemente pequenos para a complexidade do problema ou para a alta dimensionalidade do espaço de design das misturas (7)(12). Essa limitação é central, pois torna os modelos vulneráveis ao overfitting e prejudica a capacidade de generalização para condições de síntese variadas. No campo das aplicações ambientais, por exemplo, a complexidade do processo de imobilização e o pequeno dataset tornam a otimização da matriz cimentícia um desafio constante (11).

Relacionado à qualidade dos modelos, a complexa relação entre propriedades e variáveis de input é uma limitação citada, desafiando a abstração dos modelos não-lineares, especialmente na modelagem de microestrutura e propriedades mecânicas de misturas complexas (23). Em algumas aplicações estruturais, os modelos enfrentam a dificuldade em prever com modelos não lineares e a falta de dados globais, limitando a transferibilidade das equações preditivas a contextos geográficos e de matérias-primas específicos (15).

A complexidade dos processos de Engenharia Química e Ambiental também impõe restrições específicas. Na remediação, a deficiência de dados padronizados e a complexidade do processo de adsorção impedem a criação de modelos universais para diferentes tipos de geopolímeros (9). Para sistemas de filtração e colunas, a

modelagem não-linear do breakthrough é uma tarefa complexa, assim como a otimização de sistemas de adsorção em coluna (10). No design de novos materiais compostos funcionais, a base de dados experimental relativamente pequena e a complexidade do sistema fotocatalítico na degradação de antibióticos limitam o potencial de otimização de variáveis operacionais (22).

Em nível microestrutural e de simulação, as barreiras são de escala e custo. A aplicação do ML na caracterização microestrutural enfrenta o desafio da acurácia da segmentação de poros de diferentes escalas em imagens, afetando a precisão na quantificação de propriedades de transporte (10). Finalmente, a pesquisa fundamental, que busca simular a química dos geopolímeros, é limitada pelo alto custo computacional para o treinamento e pela transferibilidade dos potenciais interatômicos de Machine Learning a novas composições, exigindo um esforço computacional significativo para garantir a precisão atômica (24). Em suma, o principal trade-off do campo é a alta dimensionalidade do espaço de design de geopolímeros versus a insuficiência de datasets abrangentes e padronizados em toda a literatura.

Perspectivas Futuras

As perspectivas futuras na otimização de geopolímeros via Machine Learning apontam para uma consolidação do campo em três direções principais: robustez e interpretabilidade dos modelos, aceleração do design prescritivo, e avanço na simulação em escala atômica.

O tema mais recorrente é a necessidade de fortalecer a base de dados e a robustez dos modelos. A literatura enfatiza a urgência em expandir o dataset experimental (9) (11)(7) (15), pois a baixa quantidade e a não-padronização dos dados atuais limitam a capacidade de generalização dos modelos. Em paralelo à coleta de dados, aprimorar a capacidade preditiva exige o aprimoramento da acurácia dos modelos de ML e a exploração de algoritmos mais avançados (9)(11)(7)(22). Essa robustez está ligada à necessidade de integrar o conhecimento do domínio químico diretamente nos modelos, resultando na criação de algoritmos que sejam interpretáveis e confiáveis (9), superando a limitação da "caixa-preta".

A segunda grande perspectiva é a transição para o design prescritivo e a acessibilidade industrial. O foco é capacitar o engenheiro a usar o ML para tomar decisões ótimas de forma rápida. Paralelamente, espera-se o avanço na calibração inteligente de parâmetros microscópicos para simulações de elementos discretos (26), permitindo integrar diferentes escalas de modelagem. Para isso, há uma clara necessidade de integrar a Otimização Multiobjetivo (MOO) nos frameworks de ML (15), permitindo que as novas misturas sejam otimizadas

simultaneamente para múltiplas propriedades (como resistência e custo) antes mesmo da experimentação laboratorial. Crucialmente, diversos estudos indicam a necessidade de desenvolver software com Interface Gráfica (GUI) (9) (11) para que os modelos de ML, uma vez validados, possam ser utilizados por pesquisadores e engenheiros na prática, facilitando o design rápido e a otimização de variáveis do processo.

Por fim, o futuro do campo reside na Inovação de Materiais Sustentáveis e na Fundamentação Teórica. A perspectiva geral dos estudos é a de continuar a explorar alternativas sustentáveis e a reutilização de múltiplos resíduos industriais para a síntese de materiais de preenchimento, compostos e adsorventes (9)(11) (7)(10)(23). Em uma escala fundamental, o futuro da modelagem aponta para a expansão da aplicabilidade de potenciais interatômicos de Machine Learning (como o CeNEP) para outros sistemas de geopolímeros e o aumento da transferibilidade desses potenciais, permitindo que simulações atômicas de alta fidelidade se tornem ferramentas rotineiras para o entendimento dos mecanismos de síntese (24).

Conclusões

A presente Revisão Integrativa confirma o sucesso da implementação de estratégias de Machine Learning (ML) como a solução metodológica para superar o gargalo da complexidade estequiométrica e cinética da síntese de geopolímeros, cumprindo o objetivo de transpor o design empírico para o design orientado por dados. A topologia bibliométrica e a análise do corpus validam que o ML é o catalisador que viabiliza a economia circular, permitindo que os geopolímeros se consolidem não apenas como alternativa ao Cimento Portland (OPC) para mitigar a emissão de CO₂, mas também como materiais funcionais de alto desempenho.

A principal conclusão dos resultados reside na transição da precisão preditiva para a prescrição de materiais. Modelos de Ensemble Learning (como XGBoost e Extra Trees) e arquiteturas de Deep Learning demonstraram consistentemente um R² superior a 0.96, alcançando a acurácia necessária para a validação industrial (7)(9). Este desempenho permitiu o desenvolvimento de ferramentas prescritivas: o acoplamento de ML com algoritmos de Otimização Multiobjetivo (NSGA-II) estabeleceu o design eco-eficiente, gerando o Conjunto Pareto-Ótimo que fornece ao engenheiro químico um mapa de soluções que equilibra a máxima Resistência à Compressão com a mínima Pegada de CO₂ (7)(15). Além disso, a precisão do ML no modelo GA-GPR validou a cinética de endurecimento de longo prazo, essencial para o planejamento e a segurança de

minas, e otimizou a reologia crucial para o scale-up e a bombeabilidade industrial (23) (12).

O valor agregado do ML foi formalizado pelos Algoritmos de Interpretabilidade. O uso do SHAP foi crucial para dismantelar o modelo "caixa-preta," permitindo que a pesquisa avançasse para a descoberta científica ao invés de apenas prever. O SHAP confirmou que a relação Si/Al é o fator dominante na resistência estrutural, enquanto variáveis ambientais, como o Tempo de Cura e a dosagem do adsorvente, são as mais críticas para a otimização funcional (7)(9). Esta capacidade de desvendar as relações causais se estende ao nível fundamental, onde o ML (Potenciais Interatômicos) se tornou uma ferramenta para a aceleração de simulações em escala atômica com precisão quântica, decodificando o mecanismo químico do gel geopolimérico (24).

Apesar dessas contribuições robustas, o campo é fundamentalmente limitado pela escassez de datasets abrangentes e padronizados, que impede a plena transferibilidade e generalização dos modelos (9)(7)(15). Portanto, as perspectivas futuras convergem para a superação deste trade-off central: a agenda de pesquisa exige um esforço colaborativo para a expansão e padronização dos datasets, o desenvolvimento de modelos que incorporem o conhecimento químico do domínio e a criação de Interfaces Gráficas (GUI) e softwares (9)(11) que levem essas ferramentas preditivas e prescritivas diretamente à mão do engenheiro químico, consolidando o ML como um instrumento indispensável na próxima geração de materiais sustentáveis.

Referências

1. J. Davidovits, Geopolymer Chemistry and Applications, 3rd ed.; Institut Géopolymère, Saint-Quentin, **2013**.
2. J. L. Provis, Cem. Concr. Res. **2018**, 114, 40-48.
3. D. Wu; et al., Particuology **2025**, 104, 313-331.
4. B. Xu; et al., J. Environ. Chem. Eng. **2025**, 13, 119847.
5. L. Wei; et al., J. Environ. Chem. Eng. **2025**, 13, 116238.
6. Z. Feng; et al., Miner. Eng. **2025**, 230, 109439.
7. M. Aliyari; et al., Powder Technol. **2026**, 469, 121900.
8. C. Jiang; et al., Process Saf. Environ. Prot. **2025**, 199, 107242.
9. Y. Han; et al., J. Environ. Chem. Eng. **2025**, 13, 115978.
10. T. Sithole; et al., J. Water Process Eng. **2024**, 61, 105305.
11. L. Guo; et al., J. Hazard. Mater. **2024**, 467, 133682.
12. M. Kang; et al., Powder Technol. **2025**, 466, 121484.
13. A. Sepehr; et al., J. Food Eng. **2025**, 403, 112712.
14. S. Zhou; et al., Process Saf. Environ. Prot. **2025**, 193, 589-613.
15. Y. S. Wudil; et al., J. Environ. Chem. Eng. **2024**, 12, 111835.
16. R. Whittemore; K. Knafl, J. Adv. Nurs. **2005**, 52, 546-553.
17. M. J. Page; et al., BMJ **2021**, 372, n71.
18. Elsevier, ScienceDirect. Disponível em: <https://www.sciencedirect.com>. Acesso em: nov. 2025.
19. N. J. Van Eck; L. Waltman, Scientometrics **2010**, 84, 523-538.
20. H. Jin; et al., Sep. Purif. Technol. **2025**, 356, 129872.
21. V. Proust; et al., J. Water Process Eng. **2024**, 66, 106066.
22. H. Jin; et al., J. Environ. Chem. Eng. **2025**, 13, 116319.
23. Q. Li; et al., Powder Technol. **2024**, 434, 119337.
24. Z. Ma; et al., J. Clean. Prod. **2024**, 434, 139978.
25. Y. Wang; et al., Int. J. Heat Mass Transfer **2025**, 255, 127733.
26. X. Sun; et al., Powder Technol. **2025**, 469, 121769.
27. M. Shamsi; et al., J. Environ. Chem. Eng. **2025**, 13, 116118.
28. A. Marszałek; E. Puszczalo, Desalin. Water Treat. **2023**, 311, 100-110.
29. U. A. Qureshi; et al., J. Water Process Eng. **2020**, 38, 101380.
30. J. Liu; et al., Powder Technol. **2024**, 440, 119778.
31. L. Li; et al., J. CO2 Util. **2024**, 83, 102826.
32. M. Shirzad; M. Karimi, Desalin. Water Treat. **2020**, 182, 194-207.
33. P. U. Okoye; et al., Appl. Therm. Eng. **2022**, 214, 118901.
34. G. Gu; et al., Chem. Eng. J. **2026**, 527, 171770.
35. M. S. Netto; et al., J. Environ. Chem. Eng. **2021**, 9, 104960.
36. A. Mishra; et al., Miner. Eng. **2025**, 233, 109678.
37. P. Shrivastava; et al., Desalin. Water Treat. **2023**, 283, 222-236.
38. T. Huang; et al., J. Fuel Chem. Technol. **2025**, 53, 1191-1202.
39. S. B. Prasanna; et al., Sep. Purif. Technol. **2024**, 330, 125439.
40. M. M. H. Rocky; et al., Sep. Purif. Technol. **2026**, 383, 135987.

41. Y. Ali; et al., *Desalin. Water Treat.* **2024**, 319, 100458.
42. J. D. L. Chiquito; et al., *Desalin. Water Treat.* **2020**, 184, 1-14.
43. N. Dai; et al., *J. Environ. Chem. Eng.* **2024**, 12, 114925.
44. N. Chavana; et al., *Tribol. Int.* **2024**, 196, 109733.
45. E. G. Al-Sakkari; et al., *J. Environ. Chem. Eng.* **2023**, 11, 110732.
46. H. Khalili; et al., *Desalin. Water Treat.* **2021**, 221, 406-427.
47. R. Kayabaşı; et al., *Powder Technol.* **2026**, 469, 121729.
48. H. Zhang; et al., *Process Saf. Environ. Prot.* **2025**, 202, 107649.
49. C. Li; et al., *Powder Technol.* **2026**, 469, 121842.
50. L. Sun; et al., *Desalin. Water Treat.* **2020**, 201, 121-138.
51. G. Ye; et al., *Process Saf. Environ. Prot.* **2025**, 198, 107138.
52. M. Aksu; et al., *Desalin. Water Treat.* **2022**, 266, 226-235.
53. W. Wang; et al., *Powder Technol.* **2026**, 470, 122013.
54. F. Pan; et al., *Chem. Eng. J.* **2025**, 526, 170639.
55. P. Zhao; et al., *Powder Technol.* **2025**, 457, 120905.
56. I. Mohan; et al., *J. Environ. Chem. Eng.* **2023**, 11, 111039.
57. M. Uçurum; et al., *Powder Technol.* **2025**, 451, 120463.
58. B. Zhou; et al., *Chem. Eng. J.* **2024**, 479, 147503.
59. S. Guo; et al., *Carbon Capture Sci. Technol.* **2022**, 4, 100059.
60. D. Naghipour; et al., *Desalin. Water Treat.* **2020**, 190, 237-246.
61. N. Golestani; et al., *Desalin. Water Treat.* **2022**, 280, 312-329.
62. M. K. Tufail; et al., *Sep. Purif. Technol.* **2025**, 354, 128739.
63. X. Luo; et al., *Powder Technol.* **2023**, 426, 118635.
64. H. Sun; et al., *J. Water Process Eng.* **2025**, 69, 106687.
65. K. Ali; et al., *Desalin. Water Treat.* **2023**, 302, 231-252.
66. L. Su; et al., *J. Environ. Chem. Eng.* **2025**, 13, 116695.
- 67.
68. J. Huang; B. Wang, *Chem. Eng. J.* **2025**, 525, 169347.
69. H. Lyu; et al., *Appl. Catal. B* **2024**, 340, 123223.
70. A. Mullaimalar; R. Jeyalakshmi, *J. Water Process Eng.* **2025**, 69, 106620.
71. C. E. Enyoh; et al., *J. Environ. Chem. Eng.* **2024**, 12, 111909.
72. K. Velusamy; et al., *J. Taiwan Inst. Chem. Eng.* **2025**, 177, 105732.
73. A. Talidi; et al., *Desalin. Water Treat.* **2022**, 253, 24-38.
74. X. Shi; et al., *Carbon Capture Sci. Technol.* **2025**, 15, 100419.
75. X. Zhang; et al., *J. Environ. Chem. Eng.* **2025**, 13, 118201.
76. J. Zhu; et al., *J. Loss Prev. Process Ind.* **2025**, 98, 105727.
77. Z. Chen; et al., *J. Water Process Eng.* **2024**, 68, 106524.
78. M. Shamsi; M. Zakerinejad, *Process Saf. Environ. Prot.* **2025**, 200, 107378.
79. T. Lv; et al., *J. Environ. Chem. Eng.* **2025**, 13, 118441.
80. B. Zhou; et al., *Chem. Eng. J.* **2023**, 471, 144636.
81. J. Liu; et al., *J. Hazard. Mater.* **2024**, 463, 132856.
82. S. Ijaz; et al., *Desalin. Water Treat.* **2025**, 324, 101550.
83. B. Nemati; et al., *Chem. Eng. J.* **2025**, 524, 169414.
84. X. Bai; et al., *Process Saf. Environ. Prot.* **2025**, 194, 422-436.
85. Z. Xu; et al., *Powder Technol.* **2025**, 455, 120753.
86. M. Tadić; et al., *Desalin. Water Treat.* **2023**, 294, 139-148.
87. R. Vaithianathan; et al., *Desalin. Water Treat.* **2023**, 306, 75-88.
88. M. Yaqub; S. H. Lee, *Desalin. Water Treat.* **2020**, 198, 151-162.
89. C. Huang; et al., *Sep. Purif. Technol.* **2025**, 356, 129803.
90. I. Ghosh; et al., *Process Saf. Environ. Prot.* **2021**, 149, 345-361.
91. M. Danish; et al., *Surf. Interfaces* **2025**, 75, 107798.
92. S. Ma; et al., *J. Ind. Eng. Chem.* **2025**, 141, 305-318.
93. N. A. Khan; et al., *Biochem. Eng. J.* **2026**, 226, 110005.
94. I. B. Hvidsten; et al., *Fuel Process. Technol.* **2025**, 275, 108265.
95. W. Shao; et al., *J. Water Process Eng.* **2025**, 73, 107650.
96. V. S. Mohite; et al., *Desalin. Water Treat.* **2025**, 324, 101545.