

Universidade de Brasília

Faculdade UnB Planaltina

Programa de Pós-Graduação em Ciência de Materiais

Kádila Rocha de Santana Oliveira

**Estabilidade do Grafeno sob Dopagem com
Manganês: Análise de Colisões Dinâmicas e
Topologias de Redes via Inteligência Artificial**

Brasília

2026

Kádila Rocha de Santana Oliveira

**Estabilidade do Grafeno sob Dopagem com Manganês:
Análise de Colisões Dinâmicas e Topologias de Redes via
Inteligência Artificial**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência de Materiais pelo Programa de Pós-Graduação em Ciência de Materiais da Faculdade UnB Planaltina da Universidade de Brasília.

Universidade de Brasília

Faculdade UnB Planaltina

Programa de Pós-Graduação em Ciência de Materiais

Orientador: Marcelo Lopes Pereira Junior

Coorientador: Raphael Matozo Tromer

Brasília

2026

Kádila Rocha de Santana Oliveira

Estabilidade do Grafeno sob Dopagem com Manganês: Análise de Colisões Dinâmicas e Topologias de Redes via Inteligência Artificial/ Kádila Rocha de Santana Oliveira. – Brasília, 2026.

54 p.: il. (algumas color.) ; 30 cm.

Orientador: Marcelo Lopes Pereira Junior

Dissertação (mestrado) – Universidade de Brasília

Faculdade UnB Planaltina

Programa de Pós-Graduação em Ciência de Materiais , 2026.

1. Manganês. 2. Dinâmica. 3. Grafeno. I. Prof. Dr. Marcelo Lopes Pereira Junior II. Universidade de Brasília. III. Faculdade UnB Planaltina. IV. Aprendizado de Máquina e Impressões Digitais de Colisão Dinâmica Revelam Determinantes Estruturais da Estabilidade do Grafeno Dopado com Mn.

Kádila Rocha de Santana Oliveira

Estabilidade do Grafeno sob Dopagem com Manganês: Análise de Colisões Dinâmicas e Topologias de Redes via Inteligência Artificial

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Ciência de Materiais pelo Programa de Pós-Graduação em Ciência de Materiais da Faculdade UnB Planaltina da Universidade de Brasília.

Trabalho aprovado. Brasília, 31 de março de 2026:

Marcelo Lopes Pereira Junior
Orientador – UnB

Raphael Matozo Tromer
Coorientador – UnB

Djardiel da Silva Gomes
Avaliador – UNICAMP

Bruno Bueno Ipaves Nascimento
Avaliador – UNICAMP

Brasília
2026

Dedico este trabalho a Deus e aos meus pais, Graziane e Jonislei.

Agradecimentos

Agradeço a Deus por me proporcionar este momento, aos meus pais pelo apoio desde o início e aos meus professores, em especial ao Prof. Dr. Marcelo Lopes e ao Prof. Dr. Raphael Tromer pelas orientações.

“A vida sem ciência é uma espécie de morte.”
Sócrates (470 a.C.- 399 a.C.)

Resumo

A dopagem do grafeno com metais de transição tem se consolidado como uma estratégia altamente eficaz para modificar e controlar suas propriedades físicas e químicas, permitindo a personalização de características eletrônicas, catalíticas e magnéticas de acordo com a aplicação desejada. Nesse contexto, este estudo dedica-se à análise da estabilidade estrutural do grafeno dopado com manganês (Mn), empregando uma abordagem integrada que combina cálculos semiempíricos do calor de formação (Heat of Formation – HOF) com técnicas de aprendizado de máquina e o uso de descritores interpretáveis, capazes de fornecer insights físicos sobre os fatores que governam essa estabilidade. Para isso, foi construído um banco de dados cuidadosamente selecionado, contemplando uma ampla faixa de concentrações de dopantes e diferentes distâncias de correlação entre os átomos de manganês inseridos na rede do grafeno. Esse conjunto de dados foi explorado por meio de diversos modelos de regressão, incluindo métodos lineares tradicionais, algoritmos baseados em árvores de decisão e modelos avançados de gradient boosting. Entre as abordagens testadas, o algoritmo XGBoost destacou-se por apresentar o melhor desempenho preditivo, evidenciando sua capacidade de capturar relações não lineares complexas entre os descritores estruturais e o calor de formação do sistema. A análise dos resultados obtidos indica de forma clara que três parâmetros estruturais exercem influência dominante sobre o valor do HOF e, consequentemente, sobre a estabilidade do grafeno dopado com Mn. O primeiro deles é a distância entre os átomos de manganês, uma vez que separações Mn–Mn muito curtas tendem a introduzir tensões locais e interações desfavoráveis na rede cristalina. O segundo fator relevante é a concentração do dopante, sendo observado que níveis mais elevados de manganês contribuem para a desestabilização estrutural do material. Por fim, a manutenção da simetria local de seis membros, característica da rede hexagonal do grafeno, mostrou-se um elemento crucial para a estabilização do sistema. Em particular, verificou-se que a preservação da ordem hexagonal atua como um mecanismo estabilizador ao manter a integridade da estrutura intrínseca do grafeno, minimizando distorções na rede e reduzindo o impacto negativo da dopagem. Em contrapartida, a quebra dessa simetria local está associada a um aumento do calor de formação, refletindo uma menor estabilidade termodinâmica.

Palavras-chaves: Manganês; Dinâmica; Grafeno.

Abstract

The doping of graphene with transition metals has emerged as a highly effective strategy for tuning and controlling its physical and chemical properties, enabling the customization of electronic, catalytic, and magnetic functionalities for a wide range of applications. In this context, the present study focuses on the analysis of the structural stability of manganese (Mn)-doped graphene by adopting an integrated approach that combines semi-empirical calculations of the heat of formation (HOF) with machine learning techniques and the use of interpretable descriptors capable of providing physical insight into the factors governing stability. To this end, a carefully curated database was constructed, encompassing a broad range of dopant concentrations and varying correlation distances between manganese atoms incorporated into the graphene lattice. This dataset was explored using several regression models, including conventional linear methods, tree-based algorithms, and advanced gradient-boosting approaches. Among the tested techniques, the XGBoost algorithm stood out by delivering the highest predictive performance, highlighting its ability to capture complex, non-linear relationships between structural descriptors and the heat of formation of the system. The analysis of the obtained results clearly indicates that three structural parameters play a dominant role in determining the HOF and, consequently, the stability of Mn-doped graphene. The first key factor is the separation between manganese atoms, as short Mn–Mn distances tend to introduce local strain and unfavorable interactions within the crystalline network. The second relevant parameter is the dopant concentration, with higher manganese contents contributing to increased structural destabilization. Finally, the preservation of local sixfold symmetry, which is characteristic of graphene's hexagonal lattice, was identified as a critical stabilizing element. In particular, it was observed that the maintenance of hexagonal order acts as a stabilizing mechanism by preserving the intrinsic framework of graphene, minimizing lattice distortions and mitigating the adverse effects of doping. In contrast, the disruption of local sixfold symmetry is associated with higher heat-of-formation values, reflecting reduced thermodynamic stability.

Keywords: Manganese; Dynamics; Graphene

Lista de ilustrações

Figura 1 – Configurações representativas de Mn–Mn no grafeno e distribuição do calor de formação (HOF).	32
Figura 2 – Os coeficientes de Pearson destacam que a característica mais influente é a distância Mn–Mn, que mostra uma forte correlação negativa com HOF, consistente com a desestabilização causada pela substituição de Mn em curto alcance.	35
Figura 3 – Histórico de otimização e importância dos hiperparâmetros para 10%, 20% e 30% da divisão do teste.	37
Figura 4 – Gráficos de paridade para Regressão Linear (topo), Árvore de Decisão (meio) e XGBoost (parte inferior). Esquerda: treino; Direita: teste.	40
Figura 5 – Gráfico beeswarm de SHAP para o modelo XGBoost, classificando as características que mais influenciam o calor de formação previsto do grafeno dopado com Mn.	43

Lista de tabelas

Tabela 1 – Descrição das variáveis e sua correlação com HOF (Calor de Formação) 36

Lista de abreviaturas e siglas

CPU	Central Processing Unit
CUDA	Compute Unified Device Architecture
EISPACK	Eigensystem Package
GPU	Graphics Processing Unit
LAPACK	Linear Algebra Package

Sumário

1	INTRODUÇÃO	13
1.1	Contextualização do grafeno dopado com Mn	13
1.2	Discussão sobre estabilidade em materiais bidimensionais	15
2	OBJETIVOS	18
2.1	Objetivo geral	18
2.2	Objetivos específicos	18
3	MÉTODOS	20
3.1	Construção do modelo e arranjo dos dopantes	20
3.2	Cálculos termodinâmicos e propriedades espectroscópicas	22
3.3	Descritores estruturais e construção do banco de dados	24
3.4	Análise estatística e modelos de aprendizado de máquina	25
3.5	Validação dos modelos e aspectos de implementação computacional	27
4	RESULTADOS	30
5	CONSIDERAÇÕES FINAIS	45
	REFERÊNCIAS	49

1 Introdução

1.1 Contextualização do grafeno dopado com Mn

O grafeno permanece como um dos materiais mais intensamente investigados das últimas décadas, em virtude de suas propriedades eletrônicas, mecânicas e espectroscópicas singulares, além de seu enorme potencial tecnológico. Desde sua descoberta, esse material bidimensional tem despertado grande interesse científico e industrial, impulsionado por aplicações que vão desde dispositivos eletrônicos de próxima geração até sensores químicos de alta sensibilidade e sistemas avançados de armazenamento de energia. A combinação entre elevada mobilidade eletrônica, resistência mecânica excepcional e estrutura cristalina altamente ordenada faz do grafeno uma plataforma versátil para o desenvolvimento de novos materiais funcionais. (1)(16)(17)

Entre as diversas estratégias propostas para modificar e ampliar as propriedades do grafeno, a dopagem controlada destaca-se como uma das mais eficazes. Em particular, a introdução de átomos de metais de transição na rede cristalina do grafeno tem se mostrado uma abordagem promissora para ajustar suas características eletrônicas e magnéticas. Esses dopantes são capazes de alterar profundamente a estrutura eletrônica local, induzir estados magnéticos localizados e, em alguns casos, criar sítios ativos para processos catalíticos. Como consequência, a dopagem com metais de transição abre perspectivas relevantes em áreas como spintrônica, catálise heterogênea e o desenvolvimento de materiais multifuncionais baseados em grafeno. (2) (18)

Dentre os diversos metais de transição possíveis, o manganês (Mn) destaca-se devido à sua configuração eletrônica particular e à sua capacidade de gerar estados magnéticos localizados quando incorporado à rede do grafeno. A incorporação substitucional de átomos de Mn em sítios defeituosos da estrutura do grafeno tem atraído crescente atenção, tanto pelo potencial de engenharia de propriedades quanto pelos desafios experimentais associados à estabilidade desses defeitos. Diversos estudos experimentais relataram a síntese de sistemas Mn-grafeno suportados sobre substratos metálicos, especialmente cobre (Cu), enquanto investigações teóricas demonstraram que a dopagem com Mn pode provocar modificações profundas na estrutura eletrônica do material. (3) (19)(20)

Esses trabalhos teóricos indicam, por exemplo, uma transição para um estado isolante dominado por transferência de carga, bem como o surgimento de momentos

magnéticos localizados na vizinhança do átomo dopante. Além disso, foi demonstrado que o Mn pode ser incorporado a vacâncias ou outros defeitos estruturais do grafeno. No entanto, em muitos desses sistemas suportados, a posição final do átomo de manganês tende a situar-se ligeiramente abaixo do plano do grafeno, em forte interação com o substrato metálico. Embora essa configuração facilite a síntese experimental, ela dificulta a análise dos efeitos intrínsecos do dopante sobre o grafeno isolado, uma vez que a presença do substrato introduz contribuições adicionais à estrutura eletrônica e à estabilidade do sistema. (4)(21)(22)

Apesar dos avanços experimentais e teóricos já alcançados, a literatura ainda carece de estudos sistemáticos que abordem a estabilidade termodinâmica do grafeno dopado com Mn em função de variáveis fundamentais, como a distância de correlação entre impurezas e a concentração do dopante. Trabalhos anteriores concentraram-se, em geral, em propriedades eletrônicas ou estruturais específicas, mas até o momento não há uma investigação abrangente que combine modelagem teórica, análise estatística e caracterização espectral para elucidar de forma integrada a relação entre a distribuição dos dopantes e a estabilidade global do sistema. Essa lacuna limita a compreensão fundamental dos mecanismos que governam a incorporação do manganês no grafeno e o impacto dessa incorporação em suas propriedades funcionais. (5)(23)(24)(25)(26)

Neste trabalho, propomos uma abordagem inédita para investigar a estabilidade do grafeno dopado com manganês. Diferentemente de estudos baseados em sistemas suportados por cobre, modelamos o grafeno na ausência de qualquer substrato, assumindo que os átomos de Mn ocupam posições substitucionais diretamente no plano da rede cristalina. Essa simplificação permite isolar os efeitos intrínsecos do dopante sobre o grafeno, fornecendo uma visão complementar às observações experimentais existentes. Para esse fim, empregamos uma metodologia semiempírica capaz de acessar propriedades relacionadas à estabilidade termodinâmica e a grandezas espectroscópicas, a partir da construção de um banco de dados estruturado, parametrizado pela concentração de dopantes e pela distância de correlação entre os átomos de manganês. (6)(27)(28)

A análise desse banco de dados foi realizada utilizando três modelos amplamente empregados em aprendizado de máquina: regressão linear, árvores de decisão e o algoritmo XGBoost. Esses métodos foram utilizados para explorar de forma sistemática as relações entre parâmetros estruturais e propriedades relevantes do sistema. Para a descrição dos materiais, adotamos o descritor Dynamic Collision Fingerprint (DCF), recentemente proposto para sistemas bidimensionais, que permite capturar características estruturais e dinâmicas essenciais de forma interpretável. Com base nesse arcabouço metodológico, apresentamos discussões originais sobre os mecanismos de estabilidade, os padrões espectrais associados e as tendências termodinâmicas que

caracterizam os sistemas Mn-grafeno. (7)(29)

Os resultados obtidos não apenas fornecem uma compreensão detalhada da influência da distância entre impurezas de Mn e da concentração do dopante sobre a estabilidade do grafeno, como também estabelecem uma base metodológica sólida para a aplicação de técnicas de aprendizado de máquina na predição de propriedades de materiais dopados de baixa dimensionalidade. Dessa forma, este trabalho contribui para o entendimento fundamental da dopagem do grafeno com metais de transição e abre novas perspectivas para o projeto racional de materiais funcionais baseados em grafeno, com propriedades ajustáveis para aplicações tecnológicas avançadas. (8)(30)(31)

1.2 Discussão sobre estabilidade em materiais bidimensionais

A estabilidade termodinâmica de materiais dopados em baixa dimensionalidade representa um desafio central na ciência dos materiais, uma vez que pequenas perturbações estruturais podem levar a mudanças significativas nas propriedades físicas do sistema. Esta questão torna-se particularmente crítica quando consideramos que, ao reduzirmos a dimensionalidade de um material para escalas nanométricas — como em monocamadas de grafeno, dicalcogenetos de metais de transição (TMDs) ou nanofios — a área superficial específica aumenta drasticamente, fazendo com que a energia de superfície e as interações interfaciais passem a dominar o comportamento termodinâmico global. Diferentemente dos materiais tridimensionais (3D), onde a matriz cristalina bulk pode absorver e dissipar tensões locais induzidas por dopantes, em sistemas 2D e 1D a introdução de átomos estranhos à rede cria campos de tensão altamente localizados que podem desencadear instabilidades estruturais, como a formação de vacâncias, reconstruções de superfície ou até a transição para fases amorfas. (22)(32)(33)

Do ponto de vista da mecânica estatística, a dopagem em baixas dimensões frequentemente leva a fenômenos de ordenamento atômico que não possuem análogos em bulk. Estudos teóricos baseados no modelo de Ising e em simulações de Monte Carlo demonstram que a estabilidade de superestruturas induzidas por dopantes depende crucialmente do balanço energético entre as interações de curto e longo alcance. Em sistemas como o grafeno dopado com átomos substitucionais, a competição entre interações intra-sublattice e inter-sublattice pode resultar em comportamentos não monotônicos no parâmetro de ordem atômico durante a relaxação estrutural, indicando que o caminho para o equilíbrio termodinâmico é complexo e dependente da história térmica do material. (22)(42)(43)

A influência do alcance das interações interatômicas emerge como um fator

determinante para a viabilidade dessas estruturas. Enquanto algumas configurações ordenadas de dopantes podem ser estabilizadas exclusivamente por interações de primeiros vizinhos (curto alcance), outras demandam interações de longo alcance para minimizar a energia livre do sistema. Esta distinção é fundamental para a engenharia de materiais: se uma determinada superestrutura funcional (como aquelas que otimizam a condutividade elétrica ou criam centros de cor para aplicações quânticas) depende de interações de longo alcance, sua síntese experimental torna-se mais desafiadora, pois exige um controle rigoroso da periodicidade e da homogeneidade da dopagem em toda a extensão do material. (12) (13)

Além disso, a estabilidade termodinâmica não é governada apenas pela energia configuracional, mas também por fatores entrópicos que ganham relevância em temperaturas finitas. A vibração da rede (fônons) e as flutuações térmicas podem desestabilizar ordenamentos que seriam previstos como estáveis no zero absoluto. Em sistemas bidimensionais, onde a densidade de estados fonônicos é singular, a introdução de dopantes pode suavizar ou agravar instabilidades inerentes à dimensionalidade reduzida. A análise das funções de resposta dielétrica e das constantes elásticas mostra que dopantes com raio iônico muito discrepante do hospedeiro podem induzir modos fônons suaves, que são precursores de transições de fase estruturais ou até do colapso da rede cristalina. (14) (15)

As implicações práticas deste desafio são vastas para a nanotecnologia. Em dispositivos optoeletrônicos, a degradação progressiva de uma monocamada dopada devido à migração de impurezas ou segregação para contornos de grão (guiada pela termodinâmica) limita o tempo de vida útil do dispositivo. Compreender os diagramas de fase destes materiais em função da concentração de dopantes e da temperatura não é apenas uma questão acadêmica, mas uma necessidade para o design racional de materiais com propriedades sintonizáveis e estáveis sob condições operacionais. (22)(44)(45)

No caso do grafeno, cuja rede cristalina é altamente sensível a defeitos e impurezas, a introdução de átomos de metais de transição pode resultar tanto em melhorias funcionais quanto em instabilidades estruturais severas. Dessa forma, compreender como variáveis microscópicas — como a posição relativa entre dopantes e sua densidade na rede — influenciam a energia total do sistema torna-se essencial para o controle racional das propriedades do material. Essa compreensão é particularmente relevante quando se busca conciliar estabilidade estrutural com efeitos emergentes, como magnetismo ou atividade catalítica. (9)(46)(47)

Além disso, a distribuição espacial dos dopantes exerce um papel crucial na determinação das interações eletrônicas e magnéticas no grafeno dopado. Distâncias curtas entre impurezas podem favorecer interações diretas entre os orbitais dos átomos

metálicos, levando a efeitos cooperativos ou competitivos que impactam tanto a estabilidade quanto as propriedades funcionais. Por outro lado, distribuições mais homogêneas tendem a preservar melhor a estrutura hexagonal do grafeno, reduzindo tensões locais e distorções geométricas. Nesse contexto, a distância de correlação entre dopantes emerge como um parâmetro fundamental, capaz de conectar aspectos estruturais, eletrônicos e termodinâmicos em uma única descrição física coerente. (10)(11)(48)

2 Objetivos

2.1 Objetivo geral

Investigar de forma sistemática a estabilidade termodinâmica do grafeno dopado com manganês por meio de modelagem teórica e técnicas de aprendizado de máquina, analisando a influência da concentração do dopante, da distância de correlação entre impurezas e da preservação da simetria estrutural sobre o calor de formação e propriedades associadas, com o intuito de compreender os mecanismos físicos que governam a incorporação do Mn na rede do grafeno.

2.2 Objetivos específicos

- Construir um banco de dados estruturado de sistemas de grafeno dopado com manganês, variando de forma controlada a concentração de dopantes e a distância de correlação entre os átomos de Mn, considerando configurações substitucionais no plano do grafeno.
- Calcular o calor de formação e grandezas espectroscópicas relevantes dos sistemas Mn-grafeno utilizando uma metodologia semiempírica, de modo a avaliar a estabilidade termodinâmica das diferentes configurações estruturais.
- Empregar modelos de aprendizado de máquina, incluindo regressão linear, árvores de decisão e o algoritmo XGBoost, para identificar e quantificar as relações entre parâmetros estruturais e a estabilidade do grafeno dopado com Mn.
- Aplicar o descritor Dynamic Collision Fingerprint (DCF) para representar de forma interpretável as características estruturais e dinâmicas dos sistemas bidimensionais analisados.
- Avaliar o impacto da distância Mn-Mn, da concentração do dopante e do grau de preservação da simetria hexagonal sobre o calor de formação, identificando os principais fatores responsáveis pela estabilização ou desestabilização da rede do grafeno.
- Analisar tendências termodinâmicas e padrões espectrais associados às diferentes configurações de dopagem, correlacionando-os com os mecanismos físicos subjacentes à estabilidade do sistema.

- Estabelecer uma base metodológica que permita a extensão da abordagem proposta a outros materiais bidimensionais dopados, contribuindo para o desenvolvimento de estratégias de projeto racional de materiais funcionais baseados em grafeno.

3 Métodos

3.1 Construção do modelo e arranjo dos dopantes

A modelagem do grafeno dopado com manganês foi realizada a partir de uma supercélula composta por 72 átomos de carbono, escolhida de forma a equilibrar precisão física e viabilidade computacional. Esse tamanho de célula é suficientemente amplo para representar de maneira adequada as interações locais introduzidas pelos átomos dopantes, ao mesmo tempo em que possibilita a geração e análise de um grande número de configurações distintas sem que o custo computacional se torne proibitivo. Dessa forma, o sistema adotado permite investigar efeitos estatísticos associados à distribuição dos dopantes, preservando características essenciais da rede bidimensional do grafeno. (16)(1)

A introdução dos átomos de manganês foi realizada por meio da substituição de átomos de carbono por Mn, sempre respeitando um critério de separação mínima entre as impurezas. Esse procedimento teve como objetivo controlar a correlação espacial entre os dopantes e evitar configurações fisicamente irrelevantes caracterizadas por sobreposição excessiva de defeitos. Especificamente, os átomos de Mn foram posicionados em sítios substitucionais correspondentes a primeiros, segundos, terceiros ou quartos vizinhos mais próximos, permitindo explorar diferentes regimes de interação Mn–Mn dentro da rede cristalina. (17)(1)(36)

Para cada uma dessas restrições de vizinhança, foram geradas 100 configurações estatisticamente independentes, assegurando uma amostragem representativa dos possíveis ambientes locais associados à dopagem. Esse processo foi repetido para três níveis distintos de concentração de manganês — 3%, 6% e 10% — de modo a investigar de forma sistemática o impacto da densidade de dopantes sobre a estabilidade estrutural e as propriedades do sistema. A combinação entre diferentes concentrações e distâncias de correlação resultou em um conjunto de dados diversificado, capaz de capturar tendências globais e efeitos locais relevantes. (18)(1)(46)(26)

Essa estratégia de construção do sistema possibilita uma análise detalhada da influência conjunta da concentração de Mn e de sua distribuição espacial na rede do grafeno. Ao explorar múltiplas configurações para cada regime de dopagem, o método adotado reduz vieses associados a arranjos específicos e fornece uma base sólida para a posterior aplicação de técnicas estatísticas e de aprendizado de máquina. Dessa forma, o conjunto de modelos gerado constitui um ponto de partida robusto para a investigação dos mecanismos que governam a estabilidade termodinâmica e as

propriedades emergentes do grafeno dopado com manganês. (19)(1)

A geração de um grande número de configurações independentes é particularmente importante em sistemas dopados, nos quais pequenas variações na posição dos átomos substitucionais podem resultar em diferenças significativas nas propriedades energéticas e estruturais. Ao adotar uma abordagem estatística, baseada em múltiplas realizações para cada combinação de concentração e distância de correlação, torna-se possível minimizar efeitos de flutuações locais específicas e identificar tendências globais associadas à estabilidade do sistema. Essa estratégia garante que os resultados obtidos não estejam vinculados a configurações isoladas, mas reflitam comportamentos médios representativos do conjunto de sistemas analisados. (9)(27)

A escolha das distâncias de correlação entre os átomos de Mn foi motivada pela necessidade de explorar diferentes regimes de interação entre dopantes. Configurações em que os átomos de manganês ocupam posições de primeiros vizinhos permitem investigar cenários de forte interação Mn–Mn, nos quais distorções estruturais e efeitos eletrônicos cooperativos podem ser intensificados. Por outro lado, ao considerar dopantes separados por distâncias maiores, correspondentes a segundos, terceiros ou quartos vizinhos, torna-se possível avaliar regimes de interação mais fracos, nos quais os átomos de Mn se comportam de maneira mais isolada dentro da rede do grafeno. Essa variação sistemática fornece uma visão abrangente da influência da distribuição espacial dos dopantes sobre a estabilidade termodinâmica do material. (20)(21)(1)

Outro aspecto relevante da construção do sistema diz respeito à preservação da integridade da rede cristalina do grafeno. Embora a introdução de átomos de manganês substitucionais inevitavelmente gere distorções locais, o procedimento adotado busca evitar a formação de defeitos excessivos ou configurações altamente tensionadas que não seriam fisicamente realistas. Dessa forma, a imposição de critérios geométricos e de separação mínima entre impurezas contribui para a geração de modelos estruturalmente consistentes, adequados para a análise das propriedades intrínsecas do grafeno dopado. (9)(50)

A utilização de uma supercélula relativamente grande também desempenha um papel fundamental na redução de efeitos de interação artificial entre imagens periódicas do sistema. Em simulações periódicas, supercélulas pequenas podem introduzir interações espúrias entre dopantes em células adjacentes, comprometendo a interpretação dos resultados. Ao adotar uma supercélula de 72 átomos, garante-se que as interações relevantes ocorram predominantemente dentro da própria célula, permitindo uma descrição mais fiel das interações locais induzidas pela dopagem. (21)(26)(1)

A diversidade estrutural incorporada ao banco de dados construído constitui um elemento central da metodologia proposta. Ao combinar múltiplos níveis de con-

centração, diferentes distâncias de correlação e um número elevado de configurações independentes, o conjunto de dados resultante abrange uma ampla gama de ambientes atômicos possíveis para o manganês no grafeno. Essa diversidade é essencial para a posterior aplicação de técnicas de aprendizado de máquina, uma vez que modelos preditivos robustos dependem fortemente da representatividade e da qualidade do conjunto de treinamento. (22)(30)(22)

Por fim, a abordagem adotada nesta etapa da metodologia estabelece uma base sólida para as análises subsequentes de estabilidade termodinâmica, propriedades espectroscópicas e correlações estruturais. A construção cuidadosa do sistema e a amostragem estatística rigorosa permitem que os efeitos observados sejam atribuídos de forma confiável às variáveis de interesse, como concentração e distribuição dos dopantes. Dessa maneira, o conjunto de modelos gerado serve como alicerce para a investigação detalhada dos mecanismos físicos que governam o comportamento do grafeno dopado com manganês, bem como para a aplicação de descritores estruturais e algoritmos de aprendizado de máquina nas etapas seguintes do estudo. (9)(35)(1)

3.2 Cálculos termodinâmicos e propriedades espectroscópicas

A estabilidade termodinâmica de cada configuração de grafeno dopado com manganês foi avaliada por meio do cálculo do calor de formação, uma grandeza amplamente utilizada para quantificar a favorabilidade energética de sistemas moleculares e periódicos. O calor de formação fornece uma medida direta da energia associada à formação do sistema a partir de seus constituintes elementares, sendo, portanto, um indicador robusto da estabilidade relativa entre diferentes arranjos estruturais. Nesse estudo, o HOF foi adotado como a principal variável resposta para a análise comparativa entre as diversas configurações de dopagem consideradas. (23)(34)(1)

Todos os cálculos termodinâmicos e espectroscópicos foram realizados com o auxílio do pacote computacional MOPAC2016 (62), que implementa métodos semiempíricos amplamente empregados em estudos de sistemas de grande porte. A escolha dessa abordagem metodológica foi motivada pela necessidade de avaliar centenas de configurações estruturais distintas de forma eficiente, mantendo um equilíbrio adequado entre custo computacional e descrição física realista. Os cálculos foram conduzidos considerando uma temperatura de referência de 300 K, permitindo o acesso direto a propriedades termodinâmicas relevantes em condições próximas às ambientais. O método MOPAC usado foi o PM7 (Parameterized Model number 7). Este é o sucessor do PM6 e é o estado da arte dentro do MOPAC2016, pois corrige erros de dispersão e ligações de hidrogênio, sendo ideal para sistemas de carbono como este. (9)(22)(25)

A utilização de métodos semiempíricos possibilita a obtenção de grandezas termodinâmicas fundamentais, como entalpia, entropia e capacidade calorífica, de maneira consistente ao longo de um conjunto extenso de estruturas. Embora esses métodos envolvam aproximações, eles são particularmente adequados para análises comparativas e estudos de tendência, nos quais o foco principal reside na identificação de padrões globais associados a variações estruturais sistemáticas, como a concentração e a distribuição espacial dos dopantes na rede do grafeno. (24)(45)(49)

Para cada configuração analisada, foram extraídos valores de entalpia, que refletem o conteúdo energético total do sistema, bem como a entropia (S), relacionada ao grau de desordem estrutural e vibracional. Adicionalmente, foi calculada a capacidade calorífica a pressão constante (C_p), uma propriedade sensível às contribuições vibracionais do sistema e que pode fornecer informações complementares sobre a dinâmica térmica associada à dopagem com manganês. A análise conjunta dessas grandezas permite uma caracterização termodinâmica abrangente dos sistemas estudados. (25)(1)(48)

Além das propriedades termodinâmicas clássicas, foram também investigadas características espectroscópicas associadas às vibrações da rede cristalina. Para cada estrutura, foram selecionadas as dez frequências vibracionais mais intensas, representando os modos vibracionais com maior contribuição para a resposta espectral do sistema. Esses modos são particularmente relevantes em materiais dopados, uma vez que a introdução de impurezas tende a modificar a dinâmica vibracional local, podendo gerar assinaturas espectrais características associadas à presença do dopante. (26)(1)

Juntamente com as frequências vibracionais, foram extraídos os respectivos momentos de dipolo de transição associados a cada modo selecionado. Esses momentos de dipolo desempenham um papel central na determinação da intensidade das transições vibracionais observadas experimentalmente, especialmente em técnicas espectroscópicas como infravermelho e Raman. Dessa forma, a análise dos momentos de dipolo fornece uma ponte direta entre os resultados computacionais e observáveis experimentais mensuráveis. (9)(48)

É importante destacar que as intensidades das transições vibracionais são proporcionais ao quadrado dos momentos de dipolo de transição. Assim, a inclusão explícita dessas grandezas no conjunto de dados permite uma interpretação mais rica dos resultados espectroscópicos, indo além da simples identificação de frequências. Essa abordagem possibilita correlacionar alterações estruturais induzidas pela dopagem com variações tanto na posição quanto na intensidade das bandas vibracionais, enriquecendo a análise física do sistema. (27)(45)(43)

Por fim, a combinação entre propriedades termodinâmicas e espectroscópicas

fornece uma visão integrada do comportamento do grafeno dopado com manganês. Enquanto o calor de formação e as grandezas termodinâmicas associadas permitem avaliar a estabilidade global do sistema, as informações vibracionais e de dipolo de transição revelam detalhes sutis da estrutura local e da dinâmica atômica. Esse conjunto abrangente de propriedades constitui a base para as análises estatísticas e de aprendizado de máquina apresentadas nas etapas subsequentes do trabalho, permitindo identificar correlações profundas entre estrutura, estabilidade e resposta espectroscópica. (63)

3.3 Descritores estruturais e construção do banco de dados

Além das propriedades termodinâmicas e vibracionais calculadas para cada configuração de grafeno dopado com manganês, foi realizada uma caracterização estrutural detalhada por meio do descritor Dynamic Collision Fingerprint (DCF) (67), recentemente introduzido como uma representação eficaz para sistemas bidimensionais. Esse descritor foi concebido para capturar informações relevantes sobre os ambientes atômicos locais, levando em consideração tanto aspectos geométricos quanto dinâmicos da estrutura, o que o torna particularmente adequado para o estudo de materiais com baixa dimensionalidade. (28)

Diferentemente de abordagens puramente numéricas ou de descritores altamente parametrizados, frequentemente classificados como modelos de “caixa-preta”, o DCF apresenta a vantagem de possuir uma interpretação física clara. Cada componente do descritor está associada a características estruturais bem definidas, permitindo que os resultados obtidos a partir de modelos estatísticos ou de aprendizado de máquina sejam analisados à luz de conceitos físicos tradicionais. Essa interpretabilidade é essencial para investigações cujo objetivo vai além da simples predição de propriedades, buscando também compreender os mecanismos subjacentes que governam o comportamento do sistema. (29)

No contexto do grafeno dopado com manganês, a capacidade do DCF de descrever ambientes atômicos locais é particularmente relevante, uma vez que a presença de dopantes e defeitos introduz heterogeneidades estruturais significativas na rede cristalina. O descritor permite capturar variações na coordenação local, distorções geométricas e alterações na simetria da rede, fatores que desempenham um papel central na estabilidade termodinâmica e nas propriedades eletrônicas e vibracionais do sistema. (28)

A utilização do DCF também se mostra vantajosa para a análise de interações entre dopantes, uma vez que o descritor é sensível à distribuição espacial dos átomos de manganês na rede do grafeno. Dessa forma, torna-se possível investigar de maneira

sistemática como a distância de correlação entre impurezas influencia não apenas a energia total do sistema, mas também suas características estruturais locais. Essa abordagem é especialmente útil em estudos que envolvem múltiplas configurações e diferentes regimes de dopagem, como o presente trabalho. (30)

A partir da combinação dos descritores estruturais obtidos via DCF com as propriedades termodinâmicas e espectroscópicas calculadas, foi construído um banco de dados abrangente e multifacetado. Esse banco de dados descreve de forma consistente sistemas de grafeno dopado com manganês ao longo de uma ampla faixa de concentrações de dopantes e diferentes distâncias de correlação entre impurezas. A integração de informações de naturezas distintas permite uma representação rica e completa do sistema, fundamental para análises estatísticas robustas. (28)

Na etapa de modelagem e análise de dados, o calor de formação foi definido como a variável alvo principal, sendo utilizado como um indicador quantitativo da estabilidade termodinâmica das diferentes configurações estruturais. Essa escolha é justificada pelo papel central do HOF na comparação energética entre sistemas dopados, permitindo identificar arranjos estruturalmente favoráveis e desfavoráveis dentro do conjunto analisado. (31)

Todas as demais grandezas calculadas — incluindo propriedades termodinâmicas adicionais, informações vibracionais e os descritores estruturais derivados do DCF — foram empregadas como variáveis descritivas nos modelos de análise e aprendizado de máquina. Essa abordagem multivariada possibilita explorar correlações complexas entre estrutura, dinâmica e estabilidade, indo além de análises unidimensionais baseadas em poucos parâmetros isolados. (28)

Por fim, a construção desse banco de dados integrado estabelece uma base sólida para as etapas subsequentes do estudo, nas quais técnicas estatísticas e algoritmos de aprendizado de máquina são utilizados para identificar padrões, extrair relações não triviais e interpretar os mecanismos físicos responsáveis pela estabilidade do grafeno dopado com manganês. A estratégia adotada assegura que as conclusões obtidas estejam fundamentadas em uma representação abrangente e fisicamente consistente do sistema, fortalecendo a confiabilidade e a generalidade dos resultados apresentados. (28)

3.4 Análise estatística e modelos de aprendizado de máquina

A etapa de análise dos dados teve início com uma investigação detalhada da distribuição da variável alvo, correspondente ao calor de formação, com o objetivo de identificar possíveis padrões, tendências ou agrupamentos naturais entre as diferentes configurações estruturais. Essa análise preliminar é fundamental para compreender a

complexidade intrínseca do conjunto de dados e avaliar se características estruturais semelhantes resultam em comportamentos termodinâmicos comparáveis. Além disso, essa etapa fornece subsídios importantes para a escolha e interpretação dos modelos de aprendizado de máquina aplicados posteriormente. (32)

A exploração inicial da distribuição dos dados também permite avaliar a presença de assimetrias, regiões de maior densidade de amostras e possíveis valores extremos, aspectos que podem influenciar significativamente o desempenho dos modelos supervisionados. Ao identificar similaridades intrínsecas entre as configurações, torna-se possível antecipar a existência de regimes distintos de estabilidade associados a diferentes padrões de dopagem, reforçando a importância de uma abordagem estatística sistemática para a análise do problema. (28)

Após essa etapa exploratória, o conjunto de dados foi analisado por meio de três algoritmos supervisionados de aprendizado de máquina amplamente utilizados em estudos de ciência dos materiais. Cada um desses modelos foi escolhido com base em suas características específicas, permitindo explorar diferentes níveis de complexidade na relação entre as variáveis descritivas e a estabilidade termodinâmica do sistema. Essa abordagem comparativa possibilita avaliar não apenas o desempenho preditivo, mas também o grau de interpretabilidade de cada método. (28)

A regressão linear foi utilizada como modelo de referência, fornecendo uma linha de base para a identificação de dependências essencialmente lineares entre as variáveis estruturais e o calor de formação. Embora simples, esse modelo permite avaliar a contribuição média dos descritores e serve como ponto de comparação para métodos mais sofisticados. Em seguida, foi aplicado o algoritmo de árvore de decisão, capaz de capturar relações não lineares por meio de regras hierárquicas facilmente interpretáveis, oferecendo uma visão intuitiva das divisões mais relevantes no espaço de características. (33)

O terceiro modelo empregado foi o XGBoost, um método de ensemble baseado em gradient boosting que combina múltiplas árvores de decisão para produzir previsões altamente precisas. Esse algoritmo é amplamente reconhecido por sua robustez frente a dados complexos e por sua capacidade de modelar interações não lineares de alta ordem entre as variáveis. No contexto deste trabalho, o XGBoost mostrou-se particularmente eficaz para capturar a dependência do calor de formação em relação a múltiplos descritores estruturais, termodinâmicos e espectroscópicos. (28)

Para maximizar o desempenho do modelo XGBoost (64), foi realizada uma otimização sistemática de seus hiperparâmetros utilizando o framework Optuna. Essa ferramenta emprega uma estratégia de busca bayesiana, permitindo explorar de forma eficiente o espaço de hiperparâmetros e identificar combinações ótimas com um número reduzido de avaliações. A utilização do Optuna assegura que o modelo

final apresente um equilíbrio adequado entre complexidade, generalização e custo computacional. (34)

Entre os modelos avaliados, o XGBoost apresentou o melhor desempenho em termos de acurácia e consistência das previsões, superando tanto a regressão linear quanto as árvores de decisão individuais. No entanto, apesar de sua elevada capacidade preditiva, esse algoritmo é frequentemente classificado como um modelo de “caixa-preta”, uma vez que suas previsões não são facilmente interpretáveis a partir de seus parâmetros internos. Essa limitação pode dificultar a extração de conclusões físicas diretas a partir dos resultados.(65)

Para contornar essa questão e tornar os resultados do XGBoost interpretáveis do ponto de vista físico, foi empregado o framework SHAP (SHapley Additive exPlanations). Essa abordagem permite decompor cada previsão do modelo em contribuições individuais associadas a cada variável descritiva, com base em conceitos da teoria dos jogos. A análise dos valores SHAP possibilita identificar quais características exercem maior influência sobre o calor de formação, fornecendo insights físicos valiosos sobre os fatores que governam a estabilidade do grafeno dopado com manganês. (28)

3.5 Validação dos modelos e aspectos de implementação computacional

A avaliação da confiabilidade dos modelos de aprendizado de máquina constitui uma etapa essencial para garantir que as previsões obtidas reflitam padrões físicos reais, e não artefatos decorrentes de ajustes inadequados aos dados. Em particular, modelos excessivamente simples podem falhar em capturar relações relevantes entre as variáveis (underfitting), enquanto modelos excessivamente complexos podem se ajustar de forma exagerada ao conjunto de treinamento, comprometendo sua capacidade de generalização (overfitting). Dessa forma, estratégias de validação rigorosas foram adotadas para assegurar a robustez e a consistência dos resultados apresentados neste estudo. (35)

Uma das abordagens empregadas foi a validação do tipo holdout, na qual o conjunto de dados completo é particionado em dois subconjuntos estatisticamente independentes: um destinado ao treinamento dos modelos e outro reservado exclusivamente para teste. Essa estratégia permite avaliar o desempenho preditivo dos modelos em dados não vistos durante o processo de treinamento, fornecendo uma estimativa direta de sua capacidade de generalização. A separação adequada entre treinamento e teste é fundamental para evitar vieses e interpretações excessivamente otimistas dos resultados. (35)

Complementarmente, foi utilizada a validação cruzada, uma técnica amplamente reconhecida por fornecer estimativas mais robustas do desempenho dos modelos. Nesse procedimento, o conjunto de dados é dividido em múltiplas partições, ou folds, e o processo de treinamento e teste é repetido diversas vezes, de modo que cada partição atue alternadamente como conjunto de teste. Essa abordagem reduz a dependência dos resultados em relação a uma única divisão dos dados e permite avaliar a estabilidade do desempenho preditivo ao longo de diferentes subconjuntos. (36)

A combinação entre validação holdout e validação cruzada possibilita uma análise abrangente do comportamento dos modelos, permitindo identificar possíveis discrepâncias entre desempenho em treinamento e teste. Essa estratégia é particularmente importante em conjuntos de dados complexos e de alta dimensionalidade, como aqueles associados a sistemas dopados, nos quais correlações não triviais entre variáveis podem induzir ajustes artificiais se não forem adequadamente controladas. (35)

Do ponto de vista da implementação computacional, todas as etapas do tratamento dos dados, análise estatística e construção dos modelos de aprendizado de máquina foram realizadas utilizando a linguagem Python. A escolha dessa plataforma deve-se à sua ampla adoção na comunidade científica, à disponibilidade de bibliotecas robustas e à facilidade de integração entre diferentes ferramentas de análise de dados e modelagem estatística. (35)

Para a implementação dos modelos de aprendizado de máquina e dos procedimentos de validação, foram utilizadas bibliotecas consolidadas como o scikit-learn, que fornece uma ampla gama de algoritmos e métricas de avaliação. O algoritmo XGBoost foi implementado por meio de sua biblioteca específica, permitindo o acesso a recursos avançados de otimização e regularização. A otimização de hiperparâmetros foi conduzida com o auxílio do Optuna, enquanto a interpretação dos modelos foi realizada por meio do framework SHAP.(35)

O gerenciamento e a manipulação dos dados foram conduzidos com o suporte das bibliotecas numpy e pandas, amplamente utilizadas para operações numéricas e análise de dados estruturados. Essas ferramentas permitiram a organização eficiente do banco de dados, o pré-processamento das variáveis e a integração entre os diferentes módulos do fluxo de trabalho computacional. A utilização de bibliotecas padronizadas também contribui para a reprodutibilidade dos resultados, um aspecto fundamental em estudos computacionais. (35)

Por fim, a integração entre estratégias rigorosas de validação e uma implementação computacional bem estruturada garante que os modelos desenvolvidos apresentem desempenho confiável e interpretações fisicamente consistentes. Essa abordagem fortalece a credibilidade dos resultados obtidos e assegura que as conclusões extraídas

sobre a estabilidade do grafeno dopado com manganês estejam fundamentadas em análises estatísticas sólidas e metodologicamente robustas. (35)

4 Resultados

A Figura 1 (painéis superiores) apresenta quatro configurações representativas de pares Mn–Mn incorporados à rede de grafeno, classificadas de acordo com a menor distância relativa entre os átomos dopantes, correspondendo a primeiros, segundos, terceiros e quartos vizinhos mais próximos. Essas geometrias foram utilizadas como restrições de correlação no processo de geração do conjunto de dados. Para cada classe de separação, sítios de carbono foram selecionados aleatoriamente e substituídos por átomos de manganês, resultando em 100 configurações independentes para cada uma das três concentrações globais consideradas (3%, 6% e 10%), totalizando 1200 estruturas distintas. (37)

Embora pares de primeiros vizinhos sejam pouco prováveis em condições experimentais reais — em função das elevadas penalidades elásticas e magnéticas associadas, além da tendência do Mn de se deslocar para fora do plano do grafeno em sistemas suportados — essas configurações foram mantidas intencionalmente como casos-limite. A inclusão desses cenários extremos é fundamental para ampliar o espaço de hipóteses do modelo, quantificar de forma sistemática o custo energético associado à proximidade Mn–Mn e treinar abordagens de aprendizado de máquina capazes de generalizar adequadamente frente a uma ampla diversidade de ambientes microestruturais. (42)

O histograma apresentado na Figura 1 (painel inferior) reúne os valores de entalpia de formação (HOF) calculados para todas as configurações simuladas. Observa-se que a distribuição é dominada por uma população principal concentrada em baixas energias, exibindo um pico bem definido em torno de 2770 kcal, indicado pela linha tracejada vertical. A distribuição apresenta assimetria positiva e uma cauda alongada, comportamento compatível com o fato de que a maioria das estruturas corresponde a estados moderadamente estáveis, enquanto um subconjunto menor apresenta instabilidades energéticas mais pronunciadas. (38)

Essas configurações de maior energia estão tipicamente associadas a condições estruturalmente desfavoráveis, como separações Mn–Mn excessivamente curtas ou topologias locais pouco compatíveis com a rede de grafeno, o que resulta em um aumento significativo da HOF. Esse padrão reforça a interpretação física de que a estabilidade do sistema é fortemente modulada tanto pela concentração global de dopantes quanto pela correlação espacial entre eles. (37)

De forma particularmente relevante, observa-se o surgimento de uma lacuna estatística acima de aproximadamente 10 000 kcal mol⁻¹, seguida por duas populações

secundárias muito menos numerosas, com picos em energias substancialmente mais elevadas. Sob perspectivas tanto estatísticas quanto físicas, esses casos são não representativos: ocorrem com baixa frequência, encontram-se afastados da massa principal de probabilidade e, com elevada probabilidade, correspondem a artefatos decorrentes da combinação extrema de altas concentrações com correlações de curto alcance. (39)

A presença dessas estruturas exerce influência desproporcional sobre modelos de regressão, aumentando o erro quadrático médio, distorcendo fronteiras preditivas e mascarando relações genuínas entre estrutura e propriedade. Com o objetivo de preservar a robustez preditiva e restringir a análise ao regime fisicamente relevante, foi adotado um critério conservador de filtragem, retendo apenas configurações com HOF inferior a 10 000 kcal mol para as etapas subsequentes. (37)

Esse limiar é sustentado por três argumentos principais: (i) a clara separação estatística observada na distribuição de HOF, (ii) a plausibilidade física das estruturas mantidas após a filtragem e (iii) recomendações consolidadas em aprendizado de máquina aplicado a materiais, que indicam a exclusão de outliers estruturais não compatíveis com condições experimentais realistas. Após essa etapa de curadoria, a distribuição resultante permanece unimodal, com uma cauda moderada, favorecendo tanto procedimentos de validação confiáveis (holdout e validação cruzada k-fold) quanto análises interpretáveis de importância de atributos, como aquelas baseadas em SHAP. (43) (66)

De maneira geral, verifica-se que o aumento da concentração de Mn e a redução da distância entre impurezas tendem a elevar tanto o valor médio quanto a dispersão da HOF. Nesse contexto, a inclusão deliberada de pares de primeiros vizinhos como condições de contorno permite quantificar explicitamente esse gradiente energético, além de fornecer subsídios para o treinamento de modelos capazes de distinguir entre regimes estruturalmente “benignos” — caracterizados por separações maiores e baixas concentrações — e regimes energeticamente penalizados, associados à proximidade excessiva entre dopantes. (37)

A Figura 2 apresenta os coeficientes de correlação de Pearson entre a entalpia de formação, adotada como propriedade alvo, e as trinta características mais relevantes utilizadas para descrever o grafeno dopado com Mn. Diversas tendências fisicamente coerentes emergem dessa análise. Em particular, a correlação mais intensa é observada para a distância entre impurezas de Mn, que exibe uma correlação negativa pronunciada com a HOF, evidenciando que maiores separações entre dopantes estão associadas a estruturas energeticamente mais favoráveis. (44)

Esse comportamento é plenamente consistente do ponto de vista físico. Quando átomos de Mn são posicionados a curtas distâncias entre si, o sistema torna-se progressivamente menos estável. O raio atômico significativamente maior do Mn em

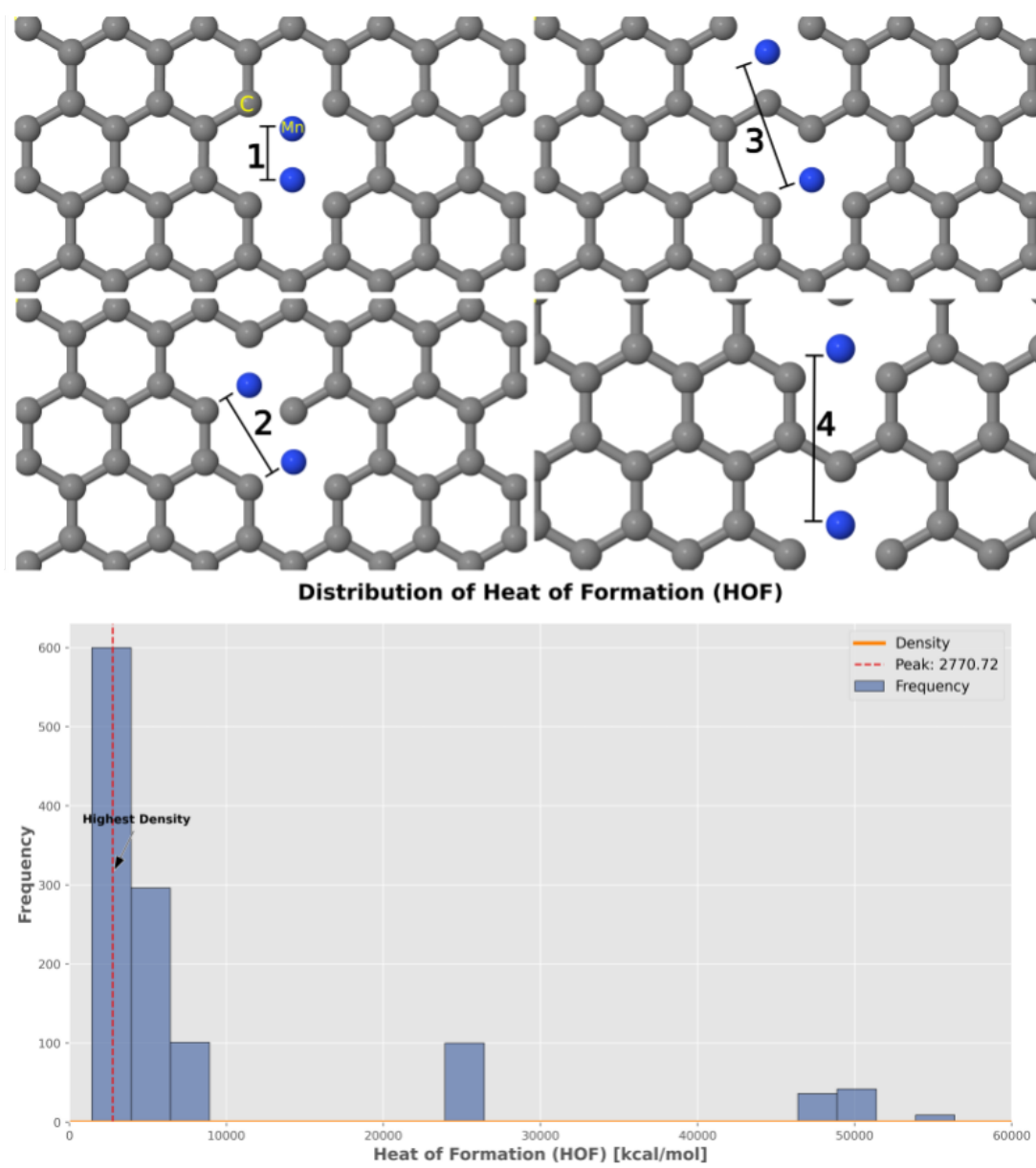


Figura 1 – Configurações representativas de Mn–Mn no grafeno e distribuição do calor de formação (HOF).

comparação ao carbono introduz distorções locais relevantes nas ligações da rede, comprometendo a integridade estrutural do grafeno. Além disso, efeitos eletrônicos adicionais — como o aumento da repulsão d–d e a possibilidade de interações magnéticas desfavoráveis — contribuem de forma cumulativa para a elevação da entalpia de formação. Em contraste, o aumento da separação entre dopantes reduz essas penalidades estruturais e eletrônicas, resultando em valores menores de HOF e, consequentemente, em maior estabilidade do sistema. Essa conclusão está diretamente alinhada com os cenários configuracionais ilustrados na Figura 1, nos quais pares de primeiros vizinhos foram mantidos deliberadamente como casos extremos justamente para quantificar esse efeito desestabilizador. (37)

Um resultado particularmente expressivo emerge ao se analisar a segunda variável mais fortemente correlacionada com a HOF: o parâmetro `6_fold_intensity`, derivado do descritor Dynamic Collision Fingerprint (DCF). Essa grandeza codifica o grau de simetria local de seis vizinhos, refletindo diretamente a ordenação hexagonal intrínseca do grafeno ideal. A correlação negativa observada indica que configurações que preservam maior simetria de seis dobras — isto é, ambientes locais mais próximos da rede hexagonal perfeita — apresentam maior estabilidade termodinâmica. Por outro lado, desvios dessa hexagonalidade levam a um aumento da HOF, sinalizando maior instabilidade estrutural. (45)

Esse resultado evidencia o papel central da ordem cristalina local na estabilização de sistemas bidimensionais dopados e destaca uma das principais vantagens do descritor DCF. Diferentemente de representações vetoriais do tipo “caixa-preta”, o DCF fornece parâmetros fisicamente interpretáveis, capazes de estabelecer conexões diretas entre motivos estruturais específicos e a estabilidade do material. Assim, a intensidade de seis dobras atua como um verdadeiro parâmetro de ordem, ligando de forma transparente simetria local e energia. (37)

A terceira característica mais correlacionada é a concentração de dopantes (`conc`), que apresenta, novamente, uma correlação positiva com a HOF. Esse comportamento é esperado para o grafeno, um material estritamente bidimensional cuja estabilidade máxima decorre da perfeita hibridização sp^2 e da ordenação hexagonal contínua. Cada substituição de um átomo de carbono por Mn rompe essa ordem local e, em concentrações mais elevadas, o efeito cumulativo dessas perturbações conduz a um aumento substancial da instabilidade energética do sistema. (46)

Além desses fatores dominantes, diversos outros descritores estruturais derivados do DCF — como `diffusivity`, `mean_free_path`, `median_path` e `skew_path` — figuram entre as variáveis mais bem correlacionadas com a HOF. A presença recorrente desses parâmetros reforça a noção de que características geométricas e topológicas exercem um papel preponderante na determinação da estabilidade, superando inclusive

variáveis termodinâmicas clássicas em termos de poder discriminativo. (37)

Somente em posições inferiores no ranking surgem grandezas termodinâmicas tradicionais, como entropia (S), capacidade calorífica a pressão constante (C_p) e entalpia. Essas variáveis exibem correlações positivas moderadas, refletindo sua contribuição conjunta para a estabilidade configuracional, porém com menor capacidade de capturar variações estruturais sutis quando comparadas aos descritores baseados em geometria local. As características vibracionais, por sua vez, apresentam correlações mistas — positivas e negativas — em consonância com a resposta complexa e não monotônica dos modos vibracionais à desordem introduzida pela dopagem substitucional. (47)

Em conjunto, os resultados apresentados na Figura 2 fornecem uma visão abrangente da interdependência entre estrutura e estabilidade em grafeno dopado com Mn. A predominância da distância Mn–Mn, da *6_fold_intensity* e da concentração de dopantes confirma a intuição física já sugerida na Figura 1: a proximidade entre impurezas e a perda da ordem hexagonal local constituem os principais mecanismos de desestabilização. O fato de descritores estruturais do DCF superarem sistematicamente variáveis termodinâmicas clássicas reforça o valor de descritores interpretáveis em ciência de materiais, permitindo a construção de modelos que são não apenas preditivos, mas também fisicamente transparentes. (37)

Essa dominância estrutural, aliada à curadoria estatística do conjunto de dados ($HOF < 10\,000$ kcal mol), assegura que as análises subsequentes de aprendizado de máquina estejam fundamentadas em características fisicamente relevantes e interpretáveis, estabelecendo uma ponte sólida entre mecanismos atômicos microscópicos e previsões orientadas por dados. (48)

A Figura 3 apresenta os resultados do processo de otimização de hiperparâmetros conduzido com o Optuna para três cenários distintos, correspondentes a divisões de teste de 10%, 20% e 30%. Em todos os casos, observa-se um padrão consistente caracterizado por uma convergência rápida seguida de estabilização em torno de um platô de desempenho. As curvas do histórico de otimização revelam que as primeiras iterações exibem valores elevados da função objetivo, refletindo a exploração inicial de regiões subótimas do espaço de hiperparâmetros. (37)

Entretanto, após apenas algumas dezenas de iterações, ocorre uma queda acentuada nos valores da função objetivo, com os melhores resultados sendo alcançados antes mesmo das cem primeiras tentativas. A partir desse ponto, as melhorias tornam-se progressivamente raras e de magnitude reduzida, indicando que o espaço de busca foi explorado de forma eficiente e que o orçamento de 1000 tentativas foi mais do que suficiente para garantir a convergência robusta do modelo. (49)

Esse comportamento, reproduzido de forma praticamente idêntica nos três cená-

rios de divisão dos dados, constitui uma evidência robusta de que os hiperparâmetros ótimos identificados não são artefatos associados a uma partição específica do conjunto de dados. Pelo contrário, eles refletem padrões estruturais consistentes na forma como o modelo interage com o problema subjacente, indicando uma convergência genuína para uma região estável do espaço de hiperparâmetros. (37)

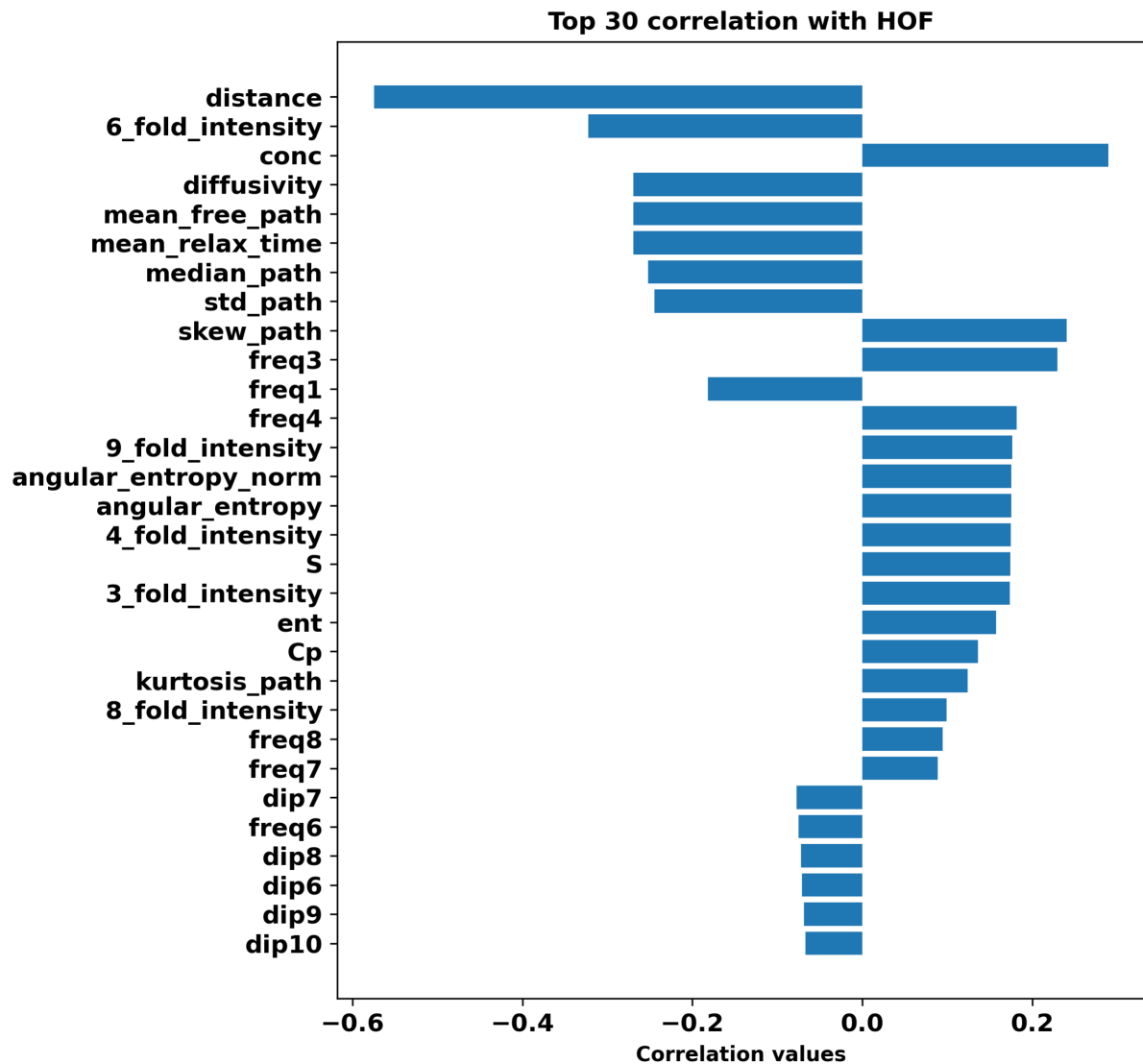


Figura 2 – Os coeficientes de Pearson destacam que a característica mais influente é a distância Mn–Mn, que mostra uma forte correlação negativa com HOF, consistente com a desestabilização causada pela substituição de Mn em curto alcance.

A análise de importância dos hiperparâmetros (Figura 3) reforça essa interpretação ao revelar que apenas um número reduzido de parâmetros domina o desempenho do XGBoost. Em todos os experimentos, o `learning_rate` se destacou como o fator mais determinante, apresentando importâncias relativas que variam aproximadamente entre 0,42 e 0,67. Em seguida, aparecem `colsample_bytree` e `reg_lambda`, que ocasionalmente

Tabela 1 – Descrição das variáveis e sua correlação com HOF (Calor de Formação)

Variável	Descrição
distance	Distância física entre átomos ou centros de massa das moléculas.
6_fold_intensity	Intensidade da simetria local de ordem 6 (ex: arranjos hexagonais).
conc	Concentração das espécies ou densidade de partículas no sistema.
diffusivity	Coeficiente de difusividade, que mede a taxa de transporte de massa.
mean_free_path	Caminho livre médio; distância média percorrida entre colisões.
mean_relax_time	Tempo médio de relaxação do sistema após uma perturbação.
median_path	Mediana do comprimento dos caminhos percorridos pelas partículas.
std_path	Desvio padrão da distribuição dos comprimentos dos caminhos.
skew_path	Assimetria (skewness) da distribuição dos comprimentos dos caminhos.
freq1, 3, 4, 6, 7, 8	Componentes de frequência específicos obtidos de análise espectral.
9_fold_intensity	Intensidade da simetria local de ordem 9.
angular_entropy	Entropia angular (normalizada ou absoluta); mede a desordem na orientação.
4_fold_intensity	Intensidade da simetria local de ordem 4 (ex: arranjos quadrados).
S / ent	Entropia termodinâmica, representando o grau de desordem do sistema.
3_fold_intensity	Intensidade da simetria local de ordem 3.
Cp	Capacidade calorífica específica a pressão constante.
kurtosis_path	Curtose da distribuição dos caminhos; mede o "achatamento" da curva.
8_fold_intensity	Intensidade da simetria local de ordem 8.
dip 6, 7, 8, 9, 10	Mínimos locais (vales) identificados na distribuição ou no espectro.

trocaram de posição dependendo da fração de dados reservada para teste. Esse resultado indica que a capacidade do modelo de se ajustar adequadamente ao sinal dos dados depende fortemente da taxa de aprendizado, atuando de forma sinérgica com o amostramento estocástico de atributos em cada árvore e com o controle da regularização L2. (50)

Em contraste, parâmetros clássicos do XGBoost, como `max_depth`, `min_child_weight` e até mesmo o número total de estimadores, apresentaram impacto praticamente desprezível dentro do espaço de busca explorado. Essa observação é coerente com um regime caracterizado por taxas de aprendizado baixas, no qual o modelo privilegia ajustes incrementais por meio de árvores relativamente rasas, tornando o refinamento da complexidade estrutural menos relevante para o desempenho final. (50)

Outro aspecto digno de destaque é a notável consistência na hierarquia de importância dos hiperparâmetros ao longo das diferentes divisões do conjunto de dados. Embora pequenas variações sejam observadas — como um papel ligeiramente mais pronunciado de `reg_lambda` no cenário com 10% de dados de teste ou uma dominância ainda maior do `learning_rate` no caso de 20% — a estrutura global permanece inalterada. O desempenho do modelo é explicado quase exclusivamente por três hiperparâmetros, enquanto os demais exercem influência marginal. Essa estabilidade entre partições distintas fortalece a interpretação de que o modelo é bem condicionado e que as configurações ótimas convergem para um regime robusto, pouco sensível a

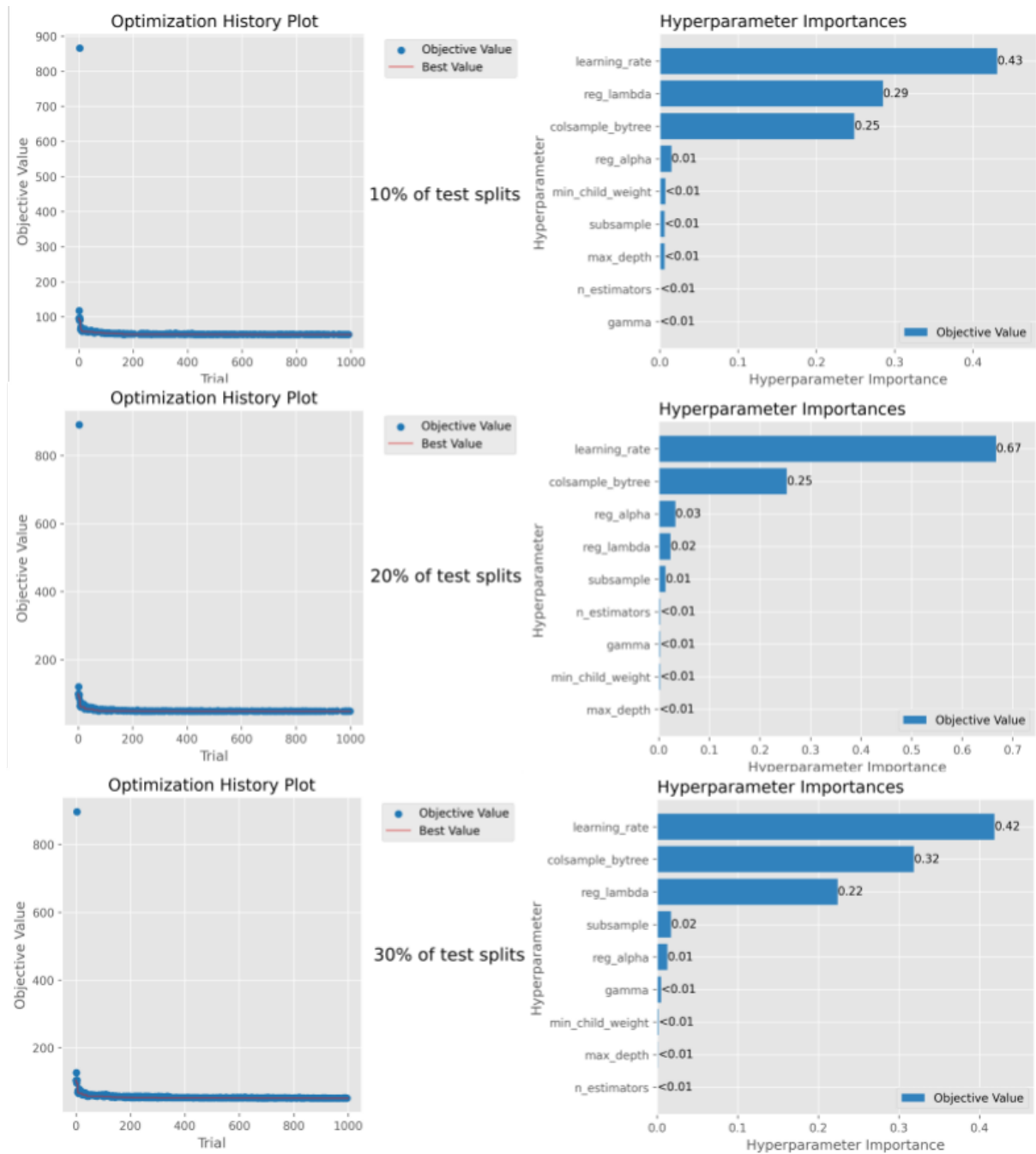


Figura 3 – Histórico de otimização e importância dos hiperparâmetros para 10%, 20% e 30% da divisão do teste.

flutuações amostrais. (51)

No que diz respeito à escolha da divisão mais adequada em termos de generalização, os resultados exigem uma análise cuidadosa. Com apenas 10% dos dados reservados para teste, o modelo alcança o menor erro absoluto médio ($MAE = 41,6$) e um coeficiente de determinação elevado ($R^2 = 0,9991$). No entanto, a pequena fração de validação levanta preocupações quanto a um possível viés otimista, uma vez que o conjunto de teste pode não representar adequadamente a heterogeneidade dos dados. Ao aumentar a fração de teste para 20%, observa-se um incremento no $MAE = 46,9$ e uma leve redução no R^2 (0,9987), sugerindo maior variância nas previsões, mas também uma avaliação mais realista do desempenho do modelo. (50)

No cenário com 30% dos dados destinados ao teste, o desempenho se estabiliza novamente ($MAE = 43,4$; $R^2 = 0,9990$), alcançando um equilíbrio entre precisão preditiva e robustez estatística. Considerando conjuntamente a magnitude do erro e a confiabilidade da avaliação, essa divisão emerge como a configuração mais balanceada, pois fornece dados suficientes tanto para o treinamento quanto para a validação, evitando estimativas excessivamente otimistas associadas a conjuntos de teste muito reduzidos. (52)

Em conjunto, os gráficos de histórico de otimização e as figuras de importância dos hiperparâmetros — apresentados sequencialmente para divisões de teste de 10%, 20% e 30% — evidenciam tanto a eficiência do processo de busca quanto a estabilidade estrutural das configurações finais do modelo. Os resultados mostram claramente que os ganhos de desempenho são obtidos rapidamente, que melhorias após as primeiras centenas de tentativas são marginais e que apenas um subconjunto restrito de hiperparâmetros — principalmente `learning_rate`, `colsample_bytree` e `reg_lambda` — governa a capacidade preditiva do modelo. Esses achados fornecem uma visão abrangente da dinâmica de otimização, dos determinantes estruturais do desempenho e da robustez do XGBoost sob diferentes particionamentos dos dados. (50)

A Figura 4 revela uma hierarquia marcante entre capacidade do modelo e generalização, com implicações diretas para o desenvolvimento de modelos substitutos preditivos em química e ciência dos materiais. O modelo linear, apesar de sua transparência interpretativa, oferece uma representação claramente inadequada da paisagem energética subjacente. Os valores de R^2 , tanto em treinamento quanto em teste, permanecem próximos de 0,63, enquanto os erros absolutos médios ultrapassam 1000. Esse subajuste sistemático demonstra que a relação entre descritores e propriedade-alvo é fortemente não linear, tornando os regressores lineares incapazes de capturar sequer tendências de primeira ordem. (53)

O regressor baseado em árvore de decisão, por sua vez, ilustra o extremo oposto. O modelo ajusta-se perfeitamente aos dados de treinamento ($R^2 = 1,0$; MAE

= 0) e apresenta um desempenho aparentemente excelente no conjunto de teste ($R^2 = 0,998$). No entanto, os gráficos de paridade revelam a natureza em “degraus” das previsões — aproximações por partes constantes que interpolam apenas regiões bem representadas do espaço de atributos. O risco associado a esse comportamento é sutil, porém significativo: métricas de desempenho intra-distribuição sugerem precisão quase perfeita, enquanto a falta de suavidade e a elevada variância intrínseca de árvores individuais comprometem seriamente a confiabilidade sob extrapolações, mesmo que moderadas. Esse resultado destaca um ponto metodológico fundamental: valores excepcionalmente altos de acurácia em conjuntos de dados quimicamente redundantes podem mascarar vieses indutivos inadequados e fraca capacidade de generalização. A apresentação de modelos baseados em árvores sem uma análise crítica aprofundada pode, portanto, induzir a uma falsa sensação de robustez. (54)

O modelo XGBoost representa uma síntese equilibrada entre os extremos observados nos modelos linear e de árvore de decisão. Ao combinar conjuntos de árvores rasas com técnicas de shrinkage, subamostragem e regularização explícita, o gradient boosting mantém um erro não nulo no conjunto de treinamento ($MAE = 24$), ao mesmo tempo em que alcança erros de teste praticamente indistinguíveis daqueles obtidos pelo regressor de árvore única ($MAE = 43$, $R^2 = 0,999$). Um aspecto fundamental é revelado pelos gráficos de paridade: ao contrário das previsões discretizadas em patamares, típicas de árvores isoladas, o XGBoost exibe uma aderência suave e contínua à linha 1:1, refletindo uma representação mais fiel das dependências físicas contínuas codificadas nos descritores. (50)

Em termos práticos, esse comportamento implica erros relativos inferiores a 1% ao longo de toda a faixa de entalpia de formação analisada. Tal nível de precisão sugere que ensembles impulsionados por boosting podem atuar como substitutos quantitativamente confiáveis para cálculos de primeiros princípios computacionalmente dispendiosos. Diferentemente de uma única árvore de decisão, esses modelos exibem maior resiliência ao ruído nos atributos e maior estabilidade ao interpolar entre diferentes famílias químicas, características essenciais para aplicações preditivas em ciência dos materiais. (55)

Considerados em conjunto, esses resultados evidenciam a importância de uma avaliação criteriosa entre diferentes classes de modelos. Eles demonstram que, embora modelos lineares sejam insuficientes para descrever observáveis complexos de natureza quântica, modelos excessivamente flexíveis podem superajustar os dados e produzir métricas artificialmente otimistas. Apenas o ensemble impulsionado por boosting atinge o equilíbrio delicado entre controle de variância e poder expressivo necessário para a construção de substitutos de alta fidelidade e boa capacidade de generalização. (50)

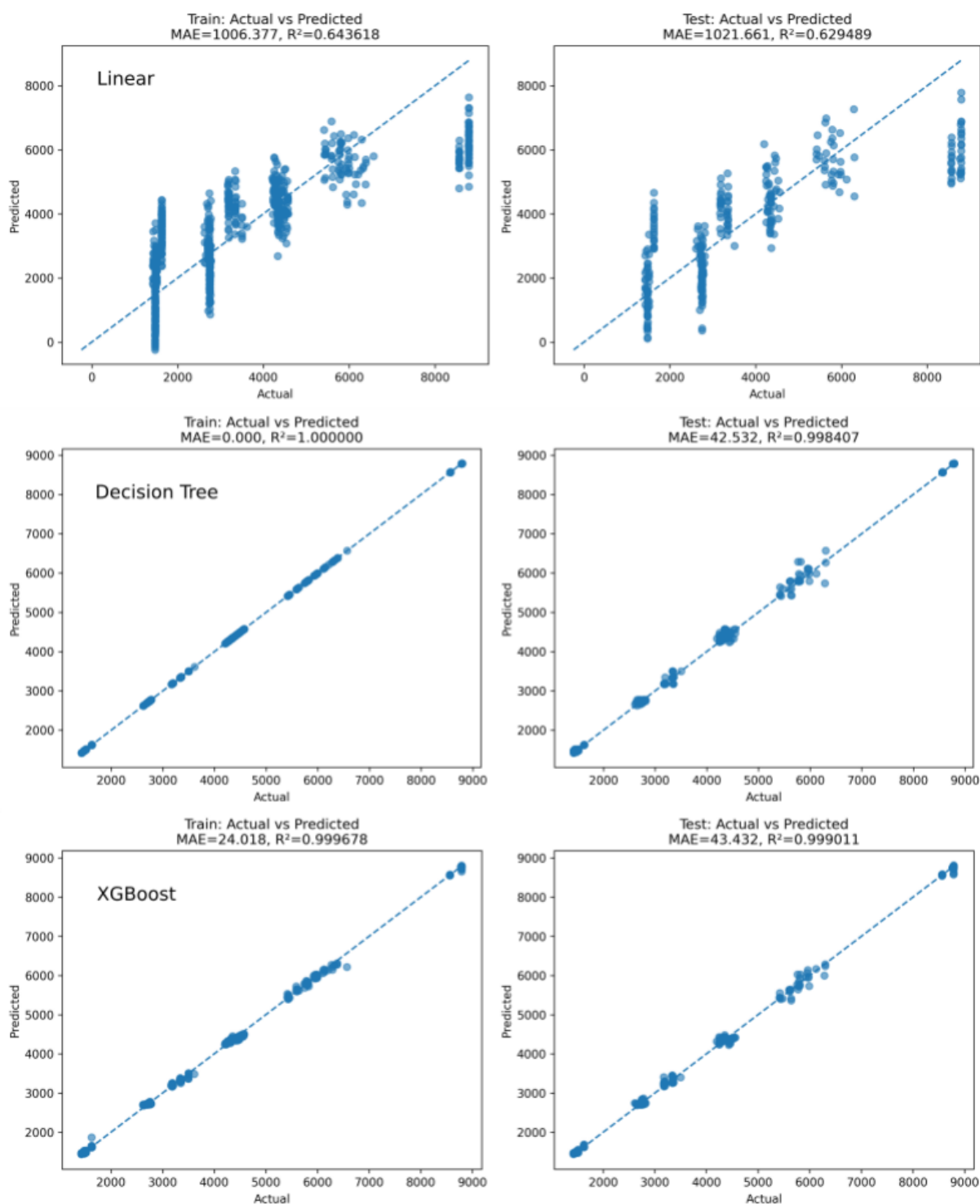


Figura 4 – Gráficos de paridade para Regressão Linear (topo), Árvore de Decisão (meio) e XGBoost (parte inferior). Esquerda: treino; Direita: teste.

Do ponto de vista mais amplo, a mensagem central é clara: a seleção robusta de modelos em materials informatics deve ir além da simples comparação de métricas de erro. Aspectos como suavidade funcional, viés indutivo e plausibilidade física da extrapolação precisam ser considerados de forma explícita. Esse nível de escrutínio assegura que previsões baseadas em aprendizado de máquina sejam não apenas precisas, mas também cientificamente consistentes e operacionalmente viáveis em fluxos de trabalho como descoberta acelerada de materiais e triagem em larga escala. (56)

O desempenho superior do XGBoost motiva uma investigação mais profunda das relações estrutura–propriedade aprendidas pelo modelo. Para esse fim, empregamos valores SHAP (SHapley Additive exPlanations) para decompor previsões individuais em contribuições aditivas associadas a cada descritor. A Figura 5 apresenta o gráfico beeswarm de SHAP, no qual os atributos são ranqueados de acordo com sua influência na previsão da entalpia de formação. (50)

De forma particularmente expressiva, as duas variáveis centrais deste estudo — a distância de correlação entre dopantes e a concentração de Mn — emergem como os fatores dominantes que governam a estabilidade termodinâmica. Distâncias curtas entre átomos de Mn contribuem sistematicamente para a desestabilização do sistema, manifestando-se como valores SHAP positivos, enquanto separações maiores favorecem entalpias de formação mais negativas. Esse comportamento está em total concordância com a expectativa física de que o agrupamento de impurezas induz tensões locais e frustração eletrônica. De modo análogo, o aumento da concentração de dopantes exerce um efeito desestabilizador, refletindo o custo energético associado a densidades elevadas de defeitos. (57)

Além desses determinantes estruturais primários, descritores termodinâmicos como a capacidade calorífica (C_p) e a entropia (S) também exibem contribuições relevantes, indicando que fatores vibracionais e configuracionais modulam a estabilidade de maneira não trivial. A presença de descritores espectroscópicos — em particular as frequências vibracionais mais intensas e grandezas relacionadas a momentos de dipolo — entre os atributos mais influentes reforça a conexão íntima entre a dinâmica da rede cristalina e a estabilidade termodinâmica. Nesse contexto, o modelo não apenas reproduz dependências físicas esperadas, mas também revela como assinaturas espectroscópicas específicas codificam tendências de estabilidade. (50)

A análise de interpretabilidade confirma, portanto, que o poder preditivo do XGBoost está firmemente ancorado em descritores fisicamente significativos. Longe de se comportar como uma “caixa-preta”, o modelo racionaliza a estabilidade em termos de separação estrutural e concentração de dopantes, integrando simultaneamente contribuições provenientes do espectro vibracional. Essa sinergia entre precisão e

interpretabilidade reforça a confiança no arcabouço de aprendizado de máquina como um substituto robusto para a exploração do panorama termodinâmico do grafeno dopado com Mn, abrindo caminho para estratégias de projeto orientadas por dados que conciliem otimização de estabilidade e propriedades espectroscópicas direcionadas. (50)

Uma avaliação rigorosa da capacidade de generalização é essencial quando modelos de aprendizado de máquina são aplicados à ciência dos materiais, especialmente em cenários nos quais substitutos preditivos podem influenciar decisões experimentais. Com esse objetivo, comparamos o desempenho por hold-out e estatísticas de validação cruzada para os três regressores candidatos. (50)

O modelo linear estabelece uma referência inicial: seus erros em treinamento e teste são praticamente idênticos ($MAE = 1000$, $R^2 = 0,63$), com uma diferença mínima entre os conjuntos e comportamento estável na validação cruzada. A ausência de sobreajuste, contudo, não decorre de robustez, mas sim de subajuste severo. O modelo é incapaz de capturar as dependências fortemente não lineares que caracterizam a estabilidade do grafeno dopado com Mn, confirmando sua inadequação para a descrição desse sistema complexo. (50)

Esse resultado evidencia o risco de interpretar pequenas discrepâncias entre erros de treinamento e teste como um indicativo de adequação do modelo quando, na realidade, a fidelidade global é baixa. A ausência de overfitting aparente pode mascarar um problema mais fundamental de incapacidade representacional, levando a conclusões equivocadas sobre a qualidade preditiva do modelo. (50)

Em contraste marcante, o regressor baseado em árvore de decisão alcança uma interpolação perfeita do conjunto de treinamento ($MAE = 0$, $R^2 = 1,0$) e apresenta desempenho excepcionalmente elevado no conjunto de teste ($R^2 = 0,998$). Embora heurísticas clássicas de sobreajuste não sejam imediatamente acionadas, o erro de treinamento praticamente nulo, aliado à forma funcional em degraus característica desse modelo, revela um regime de variância extremamente elevada. As métricas obtidas por validação cruzada permanecem altas, mas isso não elimina a limitação central: uma única árvore não possui o viés indutivo suave necessário para extrapolar de forma confiável além das configurações amostradas. Na prática, tais modelos são frágeis, fornecendo estimativas de acurácia artificialmente otimistas que se deterioram rapidamente diante de mudanças na distribuição dos dados. (58)

O XGBoost, por sua vez, emerge como uma solução equilibrada. O erro de treinamento permanece finito ($MAE = 24$), sinalizando a atuação efetiva de mecanismos de regularização, enquanto o desempenho em teste é praticamente indistinguível daquele obtido pela árvore de decisão ($MAE = 43$, $R^2 = 0,999$). A diferença moderada entre erros de treinamento e teste — da ordem de 45% do erro de teste — indica um

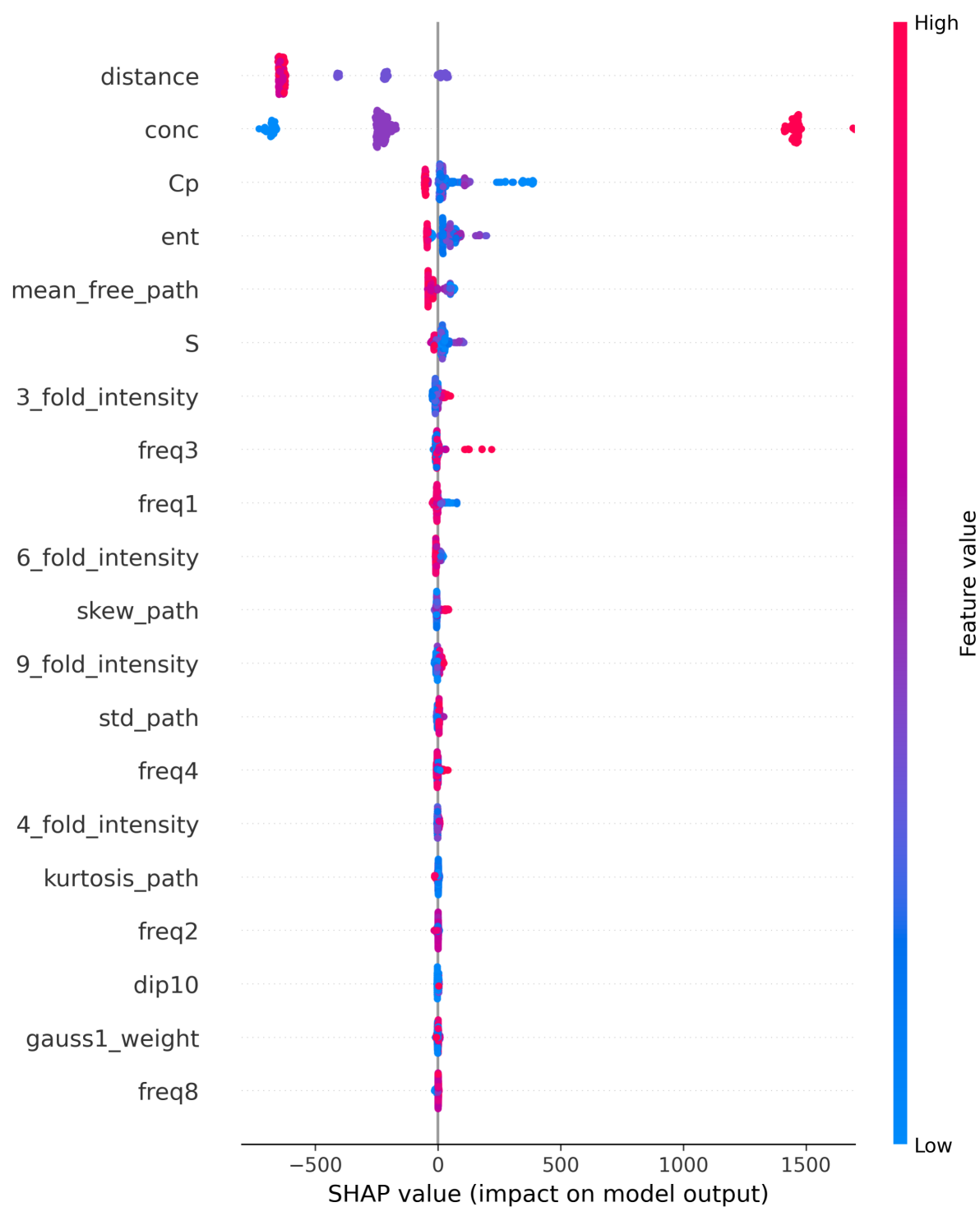


Figura 5 – Gráfico beeswarm de SHAP para o modelo XGBoost, classificando as características que mais influenciam o calor de formação previsto do grafeno dopado com Mn.

nível controlado de variância, em oposição a um sobreajuste patológico. De forma crucial, a validação cruzada confirma tanto a magnitude quanto a estabilidade do desempenho, apresentando desvios padrão reduzidos entre os diferentes folds. Assim, o ensemble combina o poder expressivo das árvores de decisão com maior robustez frente a ruído e sobreparametrização. (58)

Em termos gerais, esses diagnósticos demonstram que apenas o ensemble impulsionado por boosting atinge o equilíbrio necessário para aplicações científicas de alto impacto. Modelos lineares sofrem de subajuste e obscurecem a física relevante; árvores de decisão isoladas superajustam os dados e induzem a uma falsa sensação de confiabilidade; o XGBoost, por outro lado, oferece simultaneamente alta acurácia e capacidade de generalização. (58)

Mais do que comparar métricas isoladas, essa análise ressalta um ponto metodológico mais amplo: a verificação da ausência de sobreajuste exige a consideração conjunta da magnitude absoluta dos erros, das diferenças entre treinamento e teste e da estabilidade sob validação cruzada. A dependência exclusiva de um único critério pode levar a interpretações equivocadas. No contexto do grafeno dopado com Mn, esse processo criterioso de benchmarking valida o XGBoost como um substituto preditivo confiável, capaz de extrair tendências termodinâmicas não triviais sem sucumbir a correlações espúrias ou artefatos estatísticos. (58)

5 Considerações Finais

Nesta pesquisa foi estabelecida uma abordagem integrada para o estudo da estabilidade do grafeno dopado com manganês, combinando cálculos termodinâmicos semiempíricos, descritores estruturais fisicamente interpretáveis e algoritmos avançados de aprendizado de máquina. Essa estratégia permitiu investigar de maneira sistemática como diferentes configurações de dopagem influenciam o comportamento energético do material. Ao variar de forma controlada tanto a concentração de dopantes quanto a distância de correlação entre átomos de Mn, foi possível mapear um amplo espaço configuracional, contemplando cenários representativos das possíveis distribuições de impurezas na rede cristalina. O resultado desse processo foi a construção de um banco de dados cuidadosamente estruturado, capaz de capturar a diversidade estrutural relevante do sistema. Essa base de dados não apenas sustentou análises estatísticas consistentes, como também permitiu identificar padrões físicos robustos que relacionam diretamente a organização atômica com a estabilidade termodinâmica do grafeno dopado. (40)

A confiabilidade das previsões obtidas foi avaliada por meio de um protocolo rigoroso de validação estatística, que incluiu tanto a divisão tradicional de dados em conjuntos de treinamento e teste (*holdout*) quanto procedimentos de validação cruzada. Esses testes demonstraram que modelos baseados em *ensembles*, particularmente o algoritmo XGBoost, apresentam desempenho preditivo extremamente elevado para o problema investigado. Além da alta precisão, outro aspecto relevante é a capacidade de interpretação física dos resultados. A utilização de ferramentas de explicabilidade baseadas em valores SHAP permitiu decompor as previsões do modelo em contribuições individuais dos descritores, revelando quais variáveis exercem maior influência sobre a estabilidade do sistema. Dessa forma, o modelo não apenas produz estimativas precisas de entalpia de formação, mas também oferece uma leitura física clara dos mecanismos estruturais e termodinâmicos que governam essas previsões. (59)

A análise detalhada dos resultados revelou três mecanismos fundamentais que controlam a estabilidade do grafeno dopado com manganês. O primeiro e mais significativo deles está relacionado à separação espacial entre os átomos de Mn. Observou-se que distâncias curtas entre dopantes geram um aumento expressivo da instabilidade estrutural da rede de carbono. Esse efeito está associado principalmente às distorções locais provocadas pela diferença de tamanho atômico entre Mn e C, bem como às perturbações introduzidas no sistema de ligações eletrônicas do grafeno. Quando os dopantes se encontram muito próximos, ocorre uma intensificação dessas perturbações

estruturais e eletrônicas, elevando a entalpia de formação do sistema. Em contrapartida, o aumento da distância entre impurezas reduz tais tensões locais, favorecendo configurações energeticamente mais estáveis. (41)

O segundo fator determinante identificado está relacionado à manutenção da simetria hexagonal característica do grafeno ideal. Esse aspecto estrutural foi capturado de maneira eficaz pelo descritor Dynamic Collision Fingerprint, que quantifica o grau de ordenamento local em torno de cada sítio atômico. Os resultados mostram que configurações que preservam ambientes com simetria de seis vizinhos apresentam entalpias de formação significativamente menores. Em outras palavras, quanto mais próximo o ambiente local estiver da rede hexagonal perfeita, maior será a estabilidade termodinâmica do sistema. Esse comportamento evidencia que a estabilidade em materiais bidimensionais está profundamente ligada à preservação de sua ordem cristalina original. Pequenas perturbações estruturais que comprometem essa simetria tendem a aumentar o custo energético do sistema. (40)

O terceiro mecanismo identificado está relacionado ao efeito da concentração de dopantes. À medida que o número de átomos de Mn substituindo átomos de carbono aumenta, observa-se uma tendência clara de crescimento da instabilidade energética. Esse comportamento pode ser compreendido como resultado do acúmulo de defeitos estruturais introduzidos na rede sp^2 do grafeno. Cada substituição altera o ambiente eletrônico local e modifica a distribuição de ligações químicas no material. Em concentrações mais elevadas, essas perturbações passam a interagir entre si, amplificando o impacto global sobre a estrutura do sistema. Como consequência, a rede cristalina perde gradualmente sua regularidade, elevando a entalpia de formação e reduzindo a estabilidade global do material. (60)

Embora este trabalho não tenha considerado explicitamente os efeitos magnéticos associados à presença de manganês, os mecanismos estruturais identificados possuem implicações diretas sobre as propriedades magnéticas do sistema. A distribuição espacial dos dopantes, o grau de ordem cristalina local e a densidade de defeitos são fatores que influenciam fortemente a formação de momentos magnéticos localizados e suas interações coletivas. Estudos experimentais e teóricos anteriores já indicaram que esses elementos desempenham papel crucial na determinação das propriedades magnéticas de grafeno dopado com metais de transição. Dessa forma, os resultados apresentados aqui fornecem uma base estrutural e termodinâmica sólida que pode orientar investigações futuras voltadas especificamente para o magnetismo nesses sistemas. (40)

De maneira mais ampla, o presente estudo também demonstra o potencial de abordagens integradas baseadas em **materials informatics** para a compreensão e o projeto de novos materiais. A combinação entre descritores estruturais interpretáveis,

modelos de aprendizado de máquina robustos e técnicas de inteligência artificial explicável permite construir modelos preditivos que são simultaneamente precisos e cientificamente transparentes. Essa característica é particularmente importante em ciência dos materiais, onde a interpretação física dos resultados é tão relevante quanto a capacidade de previsão. A metodologia desenvolvida neste trabalho pode, portanto, ser facilmente adaptada para o estudo de outros sistemas dopados ou de materiais bidimensionais complexos. (61)

Os resultados obtidos apontam para novas perspectivas no desenvolvimento racional de materiais baseados em grafeno dopado. A compreensão detalhada de como fatores estruturais e termodinâmicos influenciam a estabilidade do sistema abre caminho para estratégias de engenharia de materiais orientadas por dados. Esse tipo de abordagem pode acelerar significativamente a descoberta de novas configurações estáveis com propriedades eletrônicas, magnéticas ou catalíticas específicas. Assim, ao integrar modelagem atomística, análise estatística e aprendizado de máquina interpretável, este trabalho contribui para fortalecer o papel do design orientado por dados no avanço da ciência e da engenharia de materiais. (40)

Dando continuidade às implicações científicas deste estudo, é importante destacar que a integração entre modelagem atomística e aprendizado de máquina permite superar limitações tradicionais associadas a métodos puramente baseados em cálculos de primeiros princípios. Embora abordagens *ab initio* forneçam descrições altamente precisas das propriedades eletrônicas e estruturais dos materiais, seu custo computacional torna-se rapidamente proibitivo quando se deseja explorar grandes espaços configuracionais. Nesse contexto, modelos preditivos treinados a partir de bases de dados bem estruturadas oferecem uma alternativa poderosa, capaz de reproduzir tendências termodinâmicas complexas com custos computacionais significativamente menores. A construção de substitutos computacionais confiáveis permite acelerar investigações exploratórias e direcionar esforços experimentais ou teóricos para regiões do espaço de materiais com maior probabilidade de apresentar propriedades desejáveis. (40)

Outro aspecto relevante evidenciado neste trabalho é o papel central da interpretabilidade no desenvolvimento de modelos baseados em aprendizado de máquina aplicados à ciência dos materiais. Diferentemente de abordagens puramente preditivas, nas quais o modelo atua como uma “caixa-preta”, a incorporação de ferramentas de explicabilidade permite compreender como as variáveis estruturais e termodinâmicas contribuem para as previsões finais. No presente estudo, a análise baseada em valores SHAP mostrou-se particularmente eficaz para revelar as relações entre descritores e estabilidade energética. Essa transparência é fundamental para garantir que os modelos não apenas reproduzam padrões estatísticos presentes nos dados, mas também

capturem relações fisicamente plausíveis entre estrutura e propriedade. (40)

Referências

- 1 K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov, *Electric Field Effect in Atomically Thin Carbon Films*, *Science* **306**, 666–669 (2004).
- 2 S. Wu, Q. He, C. Tan, Y. Wang, and H. Zhang, *Graphene-Based Electrochemical Sensors*, *Small* **9**, 1160–1172 (2013).
- 3 A. Ali, F. Liang, J. Zhu, and P. K. Shen, *The role of graphene in rechargeable lithium batteries*, *Nano Materials Science* (2022).
- 4 J. H. Lee, S. H. Kwon, S. Kwon, M. Cho, K. H. Kim, T. H. Han, and S. G. Lee, *Tunable Electronic Properties of Nitrogen and Sulfur Doped Graphene*, *Nanomaterials* **9**, 268 (2019).
- 5 J. E. Santos, N. M. R. Peres, J. M. B. Lopes dos Santos, and A. H. Castro Neto, *Electronic doping of graphene by deposited transition metal atoms*, *Phys. Rev. B* **84**, 085430 (2011).
- 6 R. Langer, P. Błoński, C. Hofer, P. Lazar, K. Mustonen, J. C. Meyer, T. Susi, and M. Otyepka, *Tailoring Electronic and Magnetic Properties of Graphene by Phosphorus Doping*, *ACS Appl. Mater. Interfaces* **12**, 34074–34085 (2020).
- 7 A. Roy, D. S. Buchanan, D. J. Holmgren, and D. M. Ginsberg, *Localized magnetic moments on chromium and manganese dopant atoms*, *Phys. Rev. B* **31**, 3003–3014 (1985).
- 8 B. Sanyal, O. Bengone, and S. Mirbt, *Electronic structure and magnetism of Mn-doped GaN*, *Phys. Rev. B* **68**, 205210 (2003).
- 9 S. Zhou and H. Schmidt, *Mn-doped Ge and Si: A Review of the Experimental Status*, *Materials* **3**, 5054–5082 (2010).
- 10 J. Marquina, J. Martín, J. Luengo, F. Vera, L. Roa, G. E. Delgado, F. Rodríguez, C. Renero-Lecuna, R. Valiente, and J. González, *Structural refinement and spectroscopy of Mn-doped ZnO*, *Rev. Mex. Fís.* **63**, 32–39 (2017).
- 11 M. Y. Anagaw and S. M. Hailemariam, *Effect of Mn substitutional doping on graphene*, *J. Appl. Mater. Comput. Phys.* **7** (2020).
- 12 M. Wu, C. Cao, and J. Z. Jiang, *Electronic structure of substitutionally Mn-doped graphene*, *New J. Phys.* **12** (2010).

- 13 R. Villarreal *et al.*, *Achieving High Substitutional Incorporation in Mn-Doped Graphene*, ACS Nano **18**, 17815–17825 (2024).
- 14 P.-C. Lin *et al.*, *Doping Graphene with Substitutional Mn*, ACS Nano **15**, 5449–5458 (2021).
- 15 B. Saunders, L. Hörmann, and R. J. Maurer, *Comprehensive Structure Exploration and Thermodynamics of Heteroatom Doped Graphene*, Preprint (2025).
- 16 C. R. Dean, A. F. Young, I. Meric, C. Lee, L. Wang, S. Sorgenfrei, K. Watanabe, T. Taniguchi, P. Kim, K. L. Shepard, and J. Hone. Boron nitride substrates for high-quality graphene electronics. Nat. Nanotechnol. **5**, 722 (2010)
- 17 K. K. Kim, A. Hsu, X. T. Jia, S. M. Kim, Y. S. Shi, M. Hofmann, D. Nezich, J. F. R. Nieva, M. Dresselhaus, T. Palacios. Synthesis of Monolayer Hexagonal Boron Nitride on Cu Foil Using Chemical Vapor Deposition. Nano Letters. **12**, 161 (2012).
- 18 A. Pakdel, C. Y. Zhi, Y. Bando, and D. Golberg. Production and processing of graphene and 2d crystals. Mater. Today **15**, 256 (2012).
- 19 Y. Lin and J. W. Connell. Advances in 2D boron nitride nanostructures: nanosheets, nanoribbons, nanomeshes, and hybrids with graphene. Nanoscale **4**, 6908 (2012).
- 20 J. M. Xue, J. Sanchez-Yamagishi, D. Bulmash, P. Jacquod, A. Deshpande, K. Watanabe, T. Taniguchi, P. Jarillo-Herrero and B. J. Leroy. Scanning tunnelling microscopy and spectroscopy of ultra-flat graphene on hexagonal boron nitride. Nat. Mater. **10**, 282 (2011).
- 21 A. S. Mayorov, R. V. Gorbachev, S. V. Morozov, L. Britnell, R. Jalil, L. A. Ponomarenko, P. Blake, K. S. Novoselov, K. Watanabe, T. Taniguchi and A. K. Geim. Micrometer-Scale Ballistic Transport in Encapsulated Graphene at Room Temperature. Nano Lett. **11**, 2396 (2011).
- 22 G. H. Lee, Y. J. Yu, C. Lee, C. Dean, K. L. Shepard, P. Kim and J. Hone. Electron tunneling through atomically flat and ultrathin hexagonal boron nitride. Appl. Phys. Lett. **99**, 243114 (2011).
- 23 L. Britnell, R. V. Gorbachev, R. Jalil, B. D. Belle, F. Schedin, A. Mishchenko, T. Georgiou, M. I. Katsnelson, L. Eaves, S. V. Morozov, N. M. R. Peres, J. Leist, A. K. Geim, K. S. Novoselov and L. A. Ponomarenko. Field-Effect Tunneling Transistor Based on Vertical Graphene Heterostructures. Science **335**, 947 (2012)
- 24 R. V. Gorbachev, I. Riaz, R. R. Nair, R. Jalil, L. Britnell, B. D. Belle, E. W. Hill, K. S. Novoselov, K. Watanabe, T. Taniguchi, A. K. Geim and P. Blake. Hunting for monolayer boron nitride: optical and Raman signatures. Small. **7**, 465 (2011).

- 25 Y. Lin, T. V. Williams and J. W. Connell. Soluble, Exfoliated Hexagonal Boron Nitride Nanosheets. *J. Phys. Chem. Lett.* 1, 277 (2010)
- 26 L. Song, L. J. Ci, H. Lu, P. B. Sorokin, C. H. Jin, J. Ni, A. G. Kvashnin, D. G. Kvashnin, J. Lou, B. I. Yakobson and P. M. Ajayan. Large Scale Growth and Characterization of Atomic Hexagonal Boron Nitride Layers. *Nano Lett.* 10, 3209 (2010).
- 27 Y. M. Shi, C. Hamsen, X. T. Jia, K. K. Kim, A. Reina, M. Hofmann, A. L. Hsu, K. Zhang, H. N. Li, Z. Y. Juang, M. S. Dresselhaus, L. J. Li and J. Kong. Synthesis of Few-Layer Hexagonal Boron Nitride Thin Film by Chemical Vapor Deposition. *Nano Lett.* 10, 4134 (2010).
- 28 J. N. Coleman, M. Lotya, A. Oneill, S. D. Bergin, P. J. King, U. Khan, K. Young, A. Gaucher, S. De, R. J. Smith, I. V. Shvets, S. K. Arora, G. Stanton, H. Y. Kim, K. Lee, G. T. Kim, G. S. Duesberg, T. Hallam, J. J. Boland, J. J. Wang, J. F. Donegan, J. C. Grunlan, G. Moriarty, A. Shmeliov, R. J. Nicholls, J. M. Perkins, E. M. Grieveson, K. Theuwissen, D. W. McComb, P. D. Nellist and V. Nicolosi. Two-Dimensional Nanosheets Produced by Liquid Exfoliation of Layered Materials. *Science.* 331, 568 (2011).
- 29 K. Watanabe, T. Taniguchi and H. Kanda. Direct-bandgap properties and evidence for ultraviolet lasing of hexagonal boron nitride single crystal. *Nat. Mater.* 3, 404 (2004)
- 30 J. Yu, L. Qin, Y. F. Hao, S. Kuang, X. D. Bai, Y. M. Chong, W. J. Zhang and E. Wang. Vertically Aligned Boron Nitride Nanosheets: Chemical Vapor Synthesis, Ultraviolet Light Emission, and Superhydrophobicity. *ACS Nano.* 4, 414 (2010).
- 31 C. P. Gonçalves e J. R. Mohallem. Point group symmetries of the molecular orbitals of HD⁺ beyond the Born-Oppenheimer approximation. *Chem. Phys. Lett.* 367, 533 (2003).
- 32 C. P. Gonçalves e J. R. Mohallem. Ab initio isotope simulated dynamics in the adiabatic approximation. *Chem. Phys. Lett.* 320, 118 (2000).
- 33 CASTRO, N.A.H., GUINEA, F. PERES, N.M.R. NOVOSELOV, K.S., GEIM A.K. *Rev. Mod. Phys* 81, 109, 2009.
- 34 CHEN, G. *Carbon*. [S.l]: [s.n.]Vol 41, 2003. p. 579-625.
- 35 CORREA D. R. N. A síntese do Grafeno: O premeio Nobel de Física de 2010 – 2010
- 36 DARTORA, C.A.; CABRERA G.G., *Phys. Lett. A* 377, 907, 2013.
- 37 DAS SARMA, S. Das Sarma, S. Adam, E. H. Hwang and E. Rossi, *Rev. Mod. Phys.* 83, 407, 2011

- 38 EUSTAQUIO, J.; DAVIS, J. Grafeno – Aplicações Tecnológicas. 2017.
- 39 FRAZIER, Rachel M., DALY, Daniel T., SWATLOSKI, Richard P., HATHCOCK, Kevin W. Recent Progress in Graphene-Related Nanotechnologies. *Recent Patents on Nanotechnology*, 3, p.164-176, 2009.
- 40 KASTNELSON, M.I., NOVOSELOV, K.S., *Solid State Communications* 143, 3-13, 2007.
- 41 NASCIMENTO, J. P. - Esfoliação do Grafite Natural em N-metilpirrolidona e Clorofórmio para Obtenção de Grafenos, 2013.
- 42 ALLEN, M. J.; TUNG, V. C.; KANER, R. B. Honeycomb carbon: a review of graphene. *Chemical reviews*, ACS Publications, v. 110, n. 1, p. 132–145, 2010.
- 43 ALONSO, M.; FINN, E. J. *Fundamental university physics. vol. 3: Quantum and statistical physics*. Reading, 1968.
- 44 ALTLAND, A.; SIMONS, B. D. *Condensed matter field theory*. New York: Cambridge university press, 2010.
- 45 ASHCROFT, N. W.; MERMIN, N. D. et al. *Física do estado sólido*. São Paulo: holt, rinehart and winston, new york, London, são paulo, 2011.
- 46 BJORKEN, J.; DRELL, S.; KAHN, P. B. Relativistic quantum fields. *American Journal of Physics*, v. 34, n. 4, p. 367–367, 1966.
- 47 BULBUL, B.; SEZER, M.; GREINER, W. *Relativistic Quantum Mechanics—Wave Equations*. New Yourk: Springer, Berlin, Germany,, 2000.
- 48 CARUSO, F.; OGURI, V. *Física moderna*. Rio de Janeiro: Grupo Gen-LTC, 2016.
- 49 DARTORA, C.; JIMENEZ, M. J. S.; ZANELLA, F. Os fundamentos da física dos férmions de dirac sem massa em (1+ 2)-d e o grafeno. *Revista Brasileira de Ensino de Física*, SciELO Brasil, v. 37, p. 3301–1, 2015
- 50 DIRAC, P. A. M. The quantum theory of the electron. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, The Royal Society London, v. 117, n. 778, p. 610–624, 1928.
- 51 FELTRE, R. *Química Geral*, vol. 1. São Paulo: Editora Moderna, São Paulo, 2004.
- 52 FUCHS, J.-N.; GOERBIG, M. O. Introduction to the physical properties of graphene. *Lecture notes*, v. 10, p. 11–12, 2008.

- 53 GARG, B.; BISHT, T.; LING, Y.-C. Graphene-based nanomaterials as heterogeneous acid catalysts: a comprehensive perspective. *Molecules*, MDPI, v. 19, n. 9, p. 14582–14614, 2014
- 54 GHANY, N. A. A.; ELSHERIF, S. A.; HANDAL, H. T. Revolution of graphene for different applications: State-of-the-art. *Surfaces and Interfaces*, Elsevier, v. 9, p. 93–106, 2017.
- 55 GRIFFITHS, D. J.; SCHROETER, D. F. Introduction to quantum mechanics. New York: Cambridge University Press, 2018.
- 56 JR, W. P. On the hydrogen spectrum from the standpoint of the new quantum mechanics. In: *Dynamical Groups and Spectrum Generating Algebras: (In 2 Volumes)*. www.worldscientific.com: World Scientific, 1988. p. 369–399.
- 57 NOVOSELOV, K. S.; GEIM, A. K.; MOROZOV, S. V.; JIANG, D.-e.; ZHANG, Y.; DUBONOS, S. V.; GRIGORIEVA, I. V.; FIRSOV, A. A. Electric field effect in atomically thin carbon films. *science*, American Association for the Advancement of Science, v. 306, n. 5696, p. 666–669, 2004
- 58 NOVOSELOV, K. S.; JIANG, D.; SCHEDIN, F.; BOOTH, T.; KHOTKEVICH, V.; MOROZOV, S.; GEIM, A. K. Two-dimensional atomic crystals. *Proceedings of the National Academy of Sciences*, National Acad Sciences, v. 102, n. 30, p. 10451–10453, 2005
- 59 RAY, S. Applications of graphene and graphene-oxide based nanomaterials. Waltham: William Andrew, 2015.
- 60 SEMENOFF, G. W. Condensed-matter simulation of a three-dimensional anomaly. *Physical Review Letters*, APS, v. 53, n. 26, p. 2449, 1984.
- 61 TORRES, L. E. F.; ROCHE, S.; CHARLIER, J.-C. Introduction to graphene-based nanomaterials: from electronic structure to quantum transport. New York: Cambridge University Press, 2014.
- 62 STEWART, J. J. P. MOPAC: a semiempirical molecular orbital program. *Journal of Computer-Aided Molecular Design*, v. 4, n. 1, p. 1-103, 1990.
- 63 PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825-2830, 2011.
- 64 CHEN, T.; GUESTRIN, C. XGBoost: A scalable tree boosting system. In: *PROCEEDINGS OF THE 22ND ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, p. 785-794, 2016.

- 65 AKIBA, T. et al. Optuna: A next-generation hyperparameter optimization framework. In: PROCEEDINGS OF THE 25TH ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING, p. 2623-2631, 2019.
- 66 LUNDBERG, S. M.; LEE, S. I. A unified approach to interpreting model predictions. In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS (NEURIPS), p. 4765-4774, 2017.
- 67 RINIKER, S.; LANDRUM, G. A. Dynamic fingerprints: proposed design and application to conformationally flexible molecules. *Journal of Cheminformatics*, v. 5, n. 1, p. 43, 2013.