



UNIVERSIDADE FEDERAL DE SANTA CATARINA
CENTRO TECNOLÓGICO
DEPARTAMENTO ENGENHARIA MECÂNICA MATERIAIS
CURSO DE GRADUAÇÃO EM ENGENHARIA DE MATERIAIS

Eduardo Marese Pereira

Metodologias de machine learning aplicadas à descoberta de perovskitas
magnetocalóricas

Florianópolis - SC

2024

Eduardo Marese Pereira

Metodologias de machine learning aplicadas à descoberta de perovskitas
magnetocalóricas

Trabalho de Conclusão de Curso submetido ao
curso de Engenharia de Materiais do Centro
Tecnológico da Universidade Federal de Santa
Catarina como requisito parcial para a obtenção
do título de Engenheiro de Materiais

Orientador: Prof.Dr.Sergio Yesid Gómez González
Coorientador: Me. Luis Eduardo Anaya Castro

Florianópolis - SC
2024

Marese Pereira, Eduardo

Metodologias de aprendizado de máquina aplicadas à descoberta de perovskitas magnetocalóricas / Eduardo Marese Pereira ; orientador, Sergio Yesid Gómez González, coorientador, Luis Eduardo Anaya Castro, 2024.

61 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico, Graduação em Engenharia de Materiais, Florianópolis, 2024.

Inclui referências.

1. Engenharia de Materiais. 2. Machine learning. 3. Perovskitas. 4. Efeito magnetocalórico. I. Yesid Gómez González, Sergio . II. Anaya Castro, Luis Eduardo. III. Universidade Federal de Santa Catarina. Graduação em Engenharia de Materiais. IV. Título.

Eduardo Marese Pereira

Metodologias de machine learning aplicadas à descoberta de perovskitas
magnetocalóricas

Este Trabalho de Conclusão de Curso foi julgado adequado para obtenção do título
de Engenheiro de Materiais e aprovado em sua forma final pelo Curso de
Engenharia de Materiais da Universidade Federal de Santa Catarina.

11 de julho de 2024.

Coordenação do Curso

Banca examinadora

Prof. Sergio Yesid Gómez González, Dr.

Orientador

Prof. Dr. Jaime A. Lozano Cadena

Instituição Universidade Federal de Santa Catarina

Me. Guilherme Fidelis Peixer

Instituição Universidade Federal de Santa Catarina

Florianópolis - SC
2024

AGRADECIMENTOS

Gostaria de expressar minha profunda gratidão à minha família, pelo amor incondicional, suporte emocional e por sempre acreditarem em mim. Sem vocês, nada disso seria possível.

Ao meu orientador, Professor Sergio Yesid Gómez González, agradeço pela paciência e pelos valiosos ensinamentos que contribuíram imensamente para a realização deste trabalho. Suas orientações foram essenciais para que eu pudesse superar os desafios e alcançar os objetivos propostos.

Gostaria também de agradecer ao meu coorientador, MSc. Luis Eduardo Castro Anaya, agradeço pelo suporte adicional, pelas sugestões construtivas e pelo comprometimento com meu desenvolvimento acadêmico. Sua expertise e dedicação foram fundamentais para a qualidade e a profundidade deste trabalho.

Agradeço também a todos os professores do curso de engenharia de materiais, que ao longo destes anos compartilharam seus conhecimentos e experiências, contribuindo significativamente para minha formação acadêmica e profissional.

Um agradecimento especial à Universidade Federal de Santa Catarina, por fornecer os recursos necessários para a realização desta pesquisa e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela oportunidade de bolsa de projeto, que foi fundamental para o desenvolvimento deste trabalho.

RESUMO

A descoberta de novos materiais magnetocalóricos é fundamental para aprimorar a eficiência de dispositivos de refrigeração baseados nesse princípio. Neste estudo, foram criados dois bancos de dados diferentes a partir de dados publicados (com tamanho de cristalito e sem tamanho de cristalito), e quatro modelos de machine learning foram combinados com diferentes metodologias para obtenção e seleção de descritores, visando prever as principais propriedades magnetocalóricas de perovskitas (T_c , ME, RCP). A melhor combinação de modelos e conjuntos de características foi empregada para explorar o espaço químico de manganitas de lantânio, praseodímio e neodímio. Os resultados apontaram novas composições de manganitas de lantânio que podem competir com gadolínio em termos de RCP, mas ainda são inferiores em termos de entropia magnética máxima. Além disso, foi constatado que diferentes manganitas superam outras dependendo da faixa de temperatura. A robustez dos modelos e a importância e efeito do tamanho de cristalito também são discutidos. Os resultados e diretrizes desenvolvidos neste estudo podem ser estendidos a outras propriedades. Além de obter várias percepções para estudos futuros sobre esta família de materiais, pode-se inferir que este trabalho pode ajudar na implementação de machine learning aplicado em aplicações além do escopo dos materiais magnetocalóricos, como catalisadores e materiais para células solares.

Palavras-chave: Efeito Magnetocalórico, Perovskitas, Aprendizado de Máquina. Mudança Máxima de Entropia Magnética, Potência Relativa de Resfriamento.

ABSTRACT

The discovery of new magnetocaloric materials is fundamental for enhancing the efficiency of refrigeration devices based on this principle. In this study, two different databases were created from published data (with crystallite size and without crystallite size), and four Machine Learning models were coupled with different methodologies for obtaining and selecting descriptors, aiming to predict the main magnetocaloric properties of perovskites (T_c , ME, RCP). The best combination of models and feature sets was employed to explore the chemical space of lanthanum, praseodymium, and neodymium manganites. The results pointed to new compositions of lanthanum manganites that can compete with gadolinium in terms of RCP but are still inferior in terms of maximum magnetic entropy. Additionally, it was found that different manganites outperform others depending on the temperature range. The robustness of the models and the importance and effect of crystallite size are also discussed. The results and guidelines developed in this study can be extended to other properties. Besides retrieving several insights for future studies on this family of materials, It can also be inferred that this work may help in applied machine learning implementation in applications beyond the scope of magnetocaloric materials, such as catalysts and solar cell materials.

Keywords: Magnetocaloric Effect, Perovskites, Machine Learning, Maximum Magnetic Entropy Change, Relative Cooling Power.

LISTA DE FIGURAS

Figura 1. Esquema representando o método deste trabalho: Iniciando à esquerda com os conjuntos de dados utilizados sem o tamanho do cristalito (Dataset A) e com o tamanho do cristalito (Dataset B).....	26
Figura 2. Histogramas das três propriedades magnetocalóricas indicando o total de dados e o subconjunto de dados com tamanho de cristalito.....	29
Figura 3. Efeito do tamanho do cristalito na Tc. O coeficiente de correlação de Pearson (PC) é exibido para todos os pontos de dados.....	31
Figura 4. Efeito do tamanho do cristalito na ME. O coeficiente de correlação de Pearson (PC) é exibido para todos os pontos de dados.....	32
Figura 5. Efeito do tamanho do cristalito no RCP. O coeficiente de correlação de Pearson (PC) é exibido para todos os pontos de dados.....	33
Figura 6. Mapas de calor indicando o R^2 na previsão da Tc, utilizando diferentes combinações de modelos de machine learning e conjuntos de características.....	38
Figura 7. Gráficos de paridade da melhor combinação de modelos de machine learning e conjunto de características para Tc experimental versus previsto de perovskitas magnetocalóricas para dados de treinamento e teste.....	40
Figura 8. Mapas de calor indicando o R^2 na previsão da ME, utilizando diferentes combinações de modelos de machine learning e conjuntos de características.....	41
Figura 9. Gráficos de paridade da melhor combinação de modelos de machine learning e conjunto de características para ME experimental versus previsto de perovskitas magnetocalóricas para dados de treinamento e teste.....	43
Figura 10. Mapas de calor indicando o R^2 na previsão do RCP, utilizando diferentes combinações de modelos de machine learning e conjuntos de características.....	44
Figura 11. Gráficos de paridade da melhor combinação de modelos de machine learning e conjunto de características para RCP experimental versus previsto de perovskitas magnetocalóricas para dados de treinamento e teste.....	45
Figura 12. Contorno de um gráfico de dispersão da Tc prevista versus RCP	

previsto a 2 T para o espaço químico de perovskita.....	47
Figura 13. Contorno de um gráfico de dispersão da T_c prevista versus ME previsto a 2 T para o espaço químico de perovskita.....	49
Figura 14. Previsões de TC como função do tamanho do cristalito para o modelo KRR em combinação com o conjunto de características RFE10.....	50
Figura 15. Previsões de ME como função do tamanho do cristalito para o modelo KRR em combinação com o conjunto de características RFE10. Para um campo magnético de 2 T.....	51
Figura 16. Previsões de RCP como função do tamanho do cristalito para o modelo KRR em combinação com o conjunto de características RFE10. Para um campo magnético de 2 T.....	52

LISTA DE ABREVIATURAS E SIGLAS

DC	Correlação de distância
DRX	Difração de Raio-X
DT	Árvore de decisão
EMC	Efeito magnetocalórico
Gd	Gadolínio
Geo	Média geométrica
Har	Média harmônica
KNN	k-nearest-neighbors
KRR	Kernel ridge regression
La	Lantânio
MAE	Erro médio absoluto
Max	Agregação máxima
ME	Mudança máxima na entropia magnética
Mean	Média ponderada
ML	Machine learning/aprendizado de máquina
Mn	Manganes
Min	Agregação mínima
MinDelta	Diferença mínima entre sítios
Nd	Neodímio
PC	Coeficiente de correlação de Pearson
Pr	Praseodímio
R ²	Coeficiente de determinação
RCP	Relative cooling power
RF	Random forest
RFE	Eliminação recursiva de características
RMSE	Erro quadrado médio
RNA	Rede neural artificial
SC	Coeficiente de correlação de Spearman
Sr	Estrôncio
Sum	Soma ponderada

SVM	Supporting vector machine
TC	Temperatura de Curie
ToIF	Fator de tolerância
XGB	Extreme gradient boosting

SUMÁRIO

1. INTRODUÇÃO.....	13
2. FUNDAMENTAÇÃO TEÓRICA.....	16
2.1 Efeito Magnetocalórico / Refrigeração Magnética.....	16
2.2 Machine Learning.....	17
2.2.1 Regressão de Kernel Ridge (KRR).....	19
2.2.2 Random Forest (RF).....	19
2.2.3 XGBoost (Extreme Gradient Boosting).....	20
2.2.4 Rede Neural Artificial (RNA).....	20
2.3 Perovskitas.....	21
2.4 Tamanho do cristalito.....	22
3. MÉTODOS.....	23
3.1 Base de dados.....	24
3.2 Obtenção e seleção de características.....	24
3.3 Modelos de Machine Learning.....	27
3.4 Predição de novos compostos.....	28
4. RESULTADOS.....	29
4.1 Descrição dos dados.....	29
4.2 Comparação de Seleção de Características.....	33
4.3 Modelos de Machine Learning.....	37
4.4 Temperatura de Curie.....	38
4.5 Mudança Máxima na Entropia Magnética.....	41
4.6 Relative Cooling Power.....	44
4.7 Análise dos Modelos.....	46
4.8 Efeito do tamanho do Cristalito.....	49
5. CONCLUSÃO.....	53
6. REFERÊNCIAS.....	54

1. INTRODUÇÃO

Nos últimos anos, houve um aumento considerável no interesse por materiais destinados à refrigeração magnética, o que tem motivado uma investigação aprofundada sobre o efeito magnetocalórico (EMC) (Kitanovski, 2020). Este fenômeno, identificado em materiais magnéticos específicos, desencadeia mudanças de temperatura quando expostos a campos magnéticos externos (Szymczak & Szymczak, 2008). Além de seu caráter científico intrínseco, o estudo do EMC também é significativamente relevante por fornecer uma abordagem promissora para conciliar os avanços da engenharia com a sustentabilidade ambiental. Ao oferecer uma alternativa mais ecológica e sustentável às tecnologias convencionais de resfriamento, essa linha de pesquisa tem o potencial de mitigar consideravelmente os impactos ambientais associados à refrigeração em diversos domínios de aplicação. Embora o efeito magnetocalórico tenha sido objeto de estudo há mais de cem anos, ainda não foi implementada uma aplicação prática em grande escala do fenômeno. Essa dificuldade está amplamente ligada às limitações dos materiais magnetocalóricos, que abrangem tanto o custo quanto a durabilidade. Muitos dos materiais que exibem um forte efeito magnetocalórico são caros de produzir, o que pode tornar a tecnologia impraticável em escala comercial. Além disso, a durabilidade desses materiais pode ser inadequada para suportar os ciclos de operação necessários em sistemas de refrigeração de longa duração. A busca contínua por novos materiais e avanços tecnológicos é crucial para superar essas limitações e abrir caminho para a adoção generalizada do efeito magnetocalórico como uma tecnologia de refrigeração viável e eficaz em larga escala.

Neste contexto, as perovskitas com fórmula geral ABO_3 , têm despertado grande interesse na busca por novos materiais magnetocalóricos devido à sua excepcional estrutura cristalina e propriedades magnéticas customizáveis. As diversas composições e propriedades customizáveis das perovskitas oferecem uma variedade de opções para ajustar respostas magnetocalóricas, possibilitando o design de aparelhos de resfriamento eficientes adaptáveis a várias faixas de temperatura e ambientes químicos do ponto de vista do material. As perovskitas também possuem o potencial de reduzir custos e facilitar características de fabricação; por exemplo, as manganitas possuem as vantagens mencionadas e são

mais econômicas e fáceis de preparar em comparação aos materiais magnetocalóricos clássicos, como o gadolínio e ligas intermetálicas (SMITH *et al.*, 2012; TOZRI *et al.*, 2015)

A exploração na área de descoberta de novos materiais é comumente realizada por meio de experimentos utilizando uma abordagem de tentativa e erro (AYAS *et al.*, 2023); no entanto, recentemente, estudos computacionais empregando simulações moleculares de Monte Carlo e Teoria do Funcional da Densidade (MAHJOUB *et al.*, 2015; TAHIRI *et al.*, 2021) também têm sido realizados. O cenário experimental é limitado devido à sua natureza demorada e exigente em recursos. No cenário de simulação computacional, o tempo permanece como uma restrição, uma vez que essas simulações são intensivas computacionalmente, e a precisão dos cálculos depende da capacidade do hamiltoniano de reproduzir a realidade física atômica (ELYACOUB *et al.*, 2018; LOUDAINI *et al.*, 2019). Uma alternativa a esses dois cenários é o uso de metodologias baseadas em dados, como abordagens de machine learning.

As metodologias de machine learning contam com uma quantidade limitada, porém significativa, de dados obtidos experimentalmente ou simulados, juntamente com a extração de características dos dados coletados para permitir a aplicação de algoritmos, garantindo que os atributos mais relevantes sejam considerados durante o desenvolvimento do modelo, para treinamento e otimização dos modelos de machine learning (GAO *et al.*, 2022). Uma vez treinados, esses modelos podem explorar de forma eficiente um vasto espaço químico, exigindo recursos mínimos. Notavelmente, esses cenários podem ser complementados, como no caso de experimentação ativa, utilizando refinamento iterativo de modelos de machine learning (BALACHANDRAN *et al.*, 2018).

Metodologias orientadas por dados têm sido utilizadas para explorar famílias de materiais, como por exemplo, os compostos intermetálicos de Heusler (LONG *et al.*, 2021; NELSON & SANVITO, 2019; SINGH *et al.*, 2023). Notável neste campo é o trabalho de Ucar *et al.* (2022), que estudou a entropia magnética máxima em várias famílias de compostos, encontrando como o efeito magnetocalórico é explicado pelas mudanças nas características de família para família, assim como a precisão alcançada, sugerindo a necessidade de estudar cada família separadamente. Antes do trabalho de Ucar *et al.*, apenas duas investigações de ML focaram em perovskitas: Zhang e Xu (2020) investigaram a T_c de aproximadamente

100 manganitas de lantânio empregando os parâmetros da estrutura cristalina como entradas e um modelo de Regressão do Processo de Gaussiana (GPR). Eles obtiveram um R^2 de 0,7, o que é notável considerando o conjunto de características usado, mas não muito significativo dada a quantidade de dados e a ausência de um conjunto de testes. O outro trabalho nesse sentido é de Zhai *et al.* (2018), que avaliaram 3 métodos de ML ao longo de 47 perovskitas para prever a T_c . Apesar de terem uma excelente seleção de parâmetros estruturais e composicionais, eles obtiveram um coeficiente de determinação moderado de baixa significância devido à ausência de uma etapa de teste, sendo esse valor igual a 0,85.

Como previamente apresentado, a quantidade de dados disponíveis ou utilizados limitou o desenvolvimento de modelos robustos. Além disso, a única propriedade relacionada ao EMC nesses estudos foi a T_c . Felizmente, essa situação mudou, pois o número disponível de dados experimentais aumentou a ponto de haver um apelo por um novo estudo de machine learning nesta família de materiais (SALAZAR-MUNOZ *et al.*, 2022). Além disso, as revisões realizadas (AYAS *et al.*, 2023), (SALAZAR-MUNOZ *et al.*, 2022) destacaram a importância do tamanho de cristalito na descrição do EMC em perovskitas. Portanto, surge a necessidade de avaliar esse parâmetro como uma característica morfológica que não havia sido avaliada anteriormente.

Em resposta a essas necessidades, este trabalho é dedicado ao estudo crítico de modelos de machine learning, procedimentos de coleta de dados e seleção de características para prever propriedades magnetocalóricas (temperatura de Curie, mudança máxima na entropia magnética e potência de resfriamento relativa) em perovskitas. Quatro modelos de machine learning foram avaliados em diferentes subconjuntos de características composicionais. Um descritor morfológico também foi empregado, e seu efeito nas propriedades mencionadas foi avaliado e discutido. Por fim, modelos de machine learning mais precisos foram usados para explorar o subespaço químico, e várias perspectivas são delineadas para estudos futuros nesta família de materiais.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 Efeito Magnetocalórico / Refrigeração Magnética

O efeito magnetocalórico (EMC) é um fenômeno caracterizado pela mudança de temperatura em um material magnético quando submetido a um campo magnético (NÓBREGA, 2006). Nos sólidos, esse fenômeno é intrínseco a todos os materiais magnéticos e ocorre a partir de uma variação da entropia causada pelo acoplamento de um sistema de spins magnéticos com o campo magnético (PECHARSKY & GSCHNEIDNER, 1999). O EMC foi primeiramente observado no níquel por Weiss e Piccard em 1917, e é importante na aplicação da tecnologia de refrigeração magnética, visto que os materiais magnéticos não são poluentes e demonstram maior eficiência energética. Além disso, esses materiais podem ser utilizados na medicina para o tratamento de alguns tipos de câncer (ANDRADE, 2015). Um dos materiais mais utilizados e suscetíveis a este efeito é o gadolínio, um metal reativo que uma vez exposto ao ar úmido reage com o oxigênio formando óxidos (GdO e Gd_2O_3) que se desprendem facilmente da superfície do metal e são desfavoráveis para a durabilidade e eficiência do produto fabricado. Este efeito pode ser definido pela equação 1:

(1)

$$\Delta T_{ad} = - \int_{H_0}^{H_1} \left(\frac{T}{C(T, H)} \right)_H \left(\frac{\partial M(T, H)}{\partial T} \right)_H dH$$

, onde T = temperatura, H = campo magnético aplicado, C = capacidade térmica e M = magnetização. O efeito magnetocalórico contempla propriedades como a Temperatura de Curie (T_c), a Mudança Máxima na Entropia Magnética (ME) e a Potência Relativa de Resfriamento (RCP). A T_c corresponde ao valor de temperatura em que um material ferromagnético começa a perder suas propriedades magnéticas, fazendo com que o momento magnético intrínseco do material mude para um estado desordenado, tornando-se paramagnético (BELO *et al.*, 2012). É em torno desta temperatura que um material atinge o efeito magnetocalórico máximo, motivando portanto a busca por materiais com baixo custo de produção e T_c próxima da temperatura ambiente para a aplicação em refrigeração magnética (ANDRADE, 2015). A entropia magnética, por sua vez,

representa o estado de desordem de um sistema físico dentro de um sistema magnetocalórico, sendo a variação da entropia magnética um parâmetro muito utilizado para avaliar a performance do efeito magnetocalórico de materiais (GSCHNEIDNER JR. *et al.*, 2005.). A propriedade RPC, por fim, é um parâmetro que afere a quantidade de calor transferido de um reservatório de maior temperatura para um reservatório de menor temperatura em um ciclo termodinâmico, sendo usado para verificar se um determinado material é capaz de servir em aplicações de refrigeração magnética. Essa propriedade corresponde ao produto da variação da entropia magnética máxima e a faixa de temperatura na qual a mudança de entropia é a metade de ME, conforme a equação 2 (NANTO & YU, 2016).

$$RCP = |ME| * \delta T_{FWHM} \quad (2)$$

2.2 Machine Learning

Machine learning ou aprendizado de máquina é um campo da inteligência artificial que se concentra em construir algoritmos e modelos capazes de aprender a partir de dados e realizar previsões ou tomar decisões com base nesse aprendizado.

Sua origem se deu na década de 1950 através de Alan Turing, desenvolvedor de testes para que máquinas desempenham raciocínio (SUTA *et al.*, 2020). Dois anos depois, o engenheiro Arthur Samuel desenvolveu o primeiro sistema capaz de aprender. Foi possível então, a criação de sistemas muito mais avançados, capazes de processar dados mais complexos e de forma mais rápida.

Dentro desse campo, existem dois tipos principais de aprendizado supervisionado: regressão e classificação. Além disso, existe o aprendizado não supervisionado, que também desempenha um papel importante em machine learning.

No aprendizado supervisionado, os algoritmos aprendem a partir de um conjunto de dados rotulados, ou seja, onde as entradas estão associadas a saídas conhecidas. Na tarefa de regressão, o objetivo é prever um valor contínuo ou numérico com base nas variáveis de entrada fornecidas. Os modelos de regressão

incluem técnicas como regressão linear e regressão logística. Já na classificação, o objetivo é categorizar os dados em classes distintas com base nas variáveis de entrada (GÉRON, 2019).

Os algoritmos de classificação mais conhecidos incluem árvores de decisão, k-vizinhos mais próximos (k-NN) e máquinas de vetores de suporte (SVM). Por outro lado, o aprendizado não supervisionado ocorre quando os algoritmos trabalham com dados não rotulados, ou seja, sem saídas conhecidas associadas às entradas. Nesse caso, o objetivo é identificar padrões ou estruturas subjacentes nos dados.

Uma prática comum a ser realizada sobre os modelos de machine learning é o feature engineering ou engenharia de recursos, que corresponde ao processo de utilizar o conhecimento sobre os dados para criar novas características (ou features) que possam melhorar o desempenho de modelos de machine learning. Este processo envolve a transformação, extração e seleção de características relevantes dos dados brutos, visando facilitar a tarefa de aprendizado dos algoritmos. A importância do feature engineering reside no fato de que a qualidade e a relevância das features criadas podem ter um impacto significativo na precisão e na capacidade de generalização dos modelos. Em essência, boas features são capazes de transformar dados complexos em informações mais úteis e interpretáveis, permitindo que os algoritmos capturem melhor os padrões subjacentes, assim, melhorando a performance preditiva das aplicações de machine learning.

Dois problemas recorrentes em modelos de machine learning são o overfitting e underfitting, que impactam significativamente a capacidade de um modelo de generalizar bem para novos dados. Overfitting ocorre quando um modelo se ajusta excessivamente aos dados de treinamento, capturando padrões específicos desse conjunto, resultando em uma alta precisão no treinamento, porém baixa precisão nos dados de teste. Isso geralmente ocorre quando o modelo é muito complexo, como nos casos de redes neurais profundas com muitos parâmetros e uma grande quantidade de camadas (YING, 2019). Para mitigar este problema, técnicas como validação cruzada, regularização (como L1 e L2), e métodos de ensemble (como bagging e boosting) são comumente utilizadas. Por outro lado, underfitting acontece quando o modelo é demasiado simples para capturar o comportamento dos dados, resultando em uma baixa performance tanto no conjunto de treinamento quanto no conjunto de teste (GÉRON, 2017). Modelos com

underfitting não conseguem capturar as relações importantes nos dados devido a uma complexidade insuficiente. Dessa forma, para evitar o underfitting pode-se aumentar a complexidade do modelo, adicionar mais features relevantes, ou ajustar os hiperparâmetros.

2.2.1 Regressão de Kernel Ridge (KRR)

É uma técnica de regressão poderosa que estende a regressão ridge tradicional ao aproveitar o truque do kernel, permitindo lidar com relações não lineares entre os preditores e a variável alvo (MURPHY, 2012). A KRR opera minimizando a soma regularizada dos erros quadráticos entre os valores reais e previstos, que é aumentada por um termo de penalidade que controla a complexidade do modelo e ajuda a prevenir o overfitting. Diferente dos métodos de regressão linear, a KRR emprega uma função de kernel, que é uma função matemática que possibilita mapear os dados de entrada para um espaço de dimensionalidade superior, para mapear implicitamente as características de entrada em um espaço de dimensões superiores, onde um modelo linear é ajustado para capturar padrões presentes nos dados. Existem diferentes tipos de funções de kernel, incluindo o kernel linear, o kernel polinomial e o kernel gaussiano (HE *et al.*, 2014). Essa abordagem baseada em kernel permite que a KRR modele relações complexas de forma flexível, enquanto mantém a eficiência computacional.

2.2.2 Random Forest (RF)

Esta técnica de machine learning em conjunto combina várias árvores de decisão para aumentar a precisão preditiva e reduzir o overfitting (BRUNTON & KUTZ, 2022). Ela começa com a criação de várias amostras de bootstrap, onde pontos de dados são selecionados aleatoriamente do conjunto de dados original com reposição, formando subconjuntos com potenciais sobreposições. Diversas árvores de decisão são construídas usando essas amostras, considerando um subconjunto aleatório de características a cada divisão para introduzir variabilidade e descorrelacionar as árvores, assim mitigando o overfitting. No final, as previsões de árvores individuais são ou médias para tarefas de regressão ou submetidas à

votação majoritária para tarefas de classificação, proporcionando previsões mais estáveis e precisas enquanto minimizam o impacto de outliers e ruídos.

2.2.3 XGBoost (Extreme Gradient Boosting)

Assim como o modelo anterior, o Extreme Gradient Boosting é um algoritmo de machine learning em conjunto que cria um modelo preditivo adicionando sequencialmente árvores de decisão, cada uma com o objetivo de corrigir os erros cometidos pelas árvores anteriores (CHEN & HE, 2023). Essas árvores de decisão são construídas usando um conjunto de dados com pesos ajustados com base nos erros das árvores anteriores. O algoritmo também incorpora técnicas de regularização para controlar o overfitting. O XGBoost melhora seu desempenho otimizando uma função de perda e aplicando uma abordagem baseada em descida de gradiente. Este método permite que o XGBoost se ajuste de maneira eficiente a relações complexas dentro dos dados, tornando-o um algoritmo altamente competitivo para várias tarefas de machine learning, como classificação e regressão.

2.2.4 Rede Neural Artificial (RNA)

A Rede Neural Artificial é um modelo de machine learning que imita o funcionamento do cérebro humano. Consiste em neurônios artificiais dispostos em diferentes camadas, que processam informações e aprendem padrões com base nos dados fornecidos (MURPHY, 2012). Dependendo da complexidade do sistema pode haver um maior número de camadas ocultas dentro da rede neural. Dentro destas camadas existem funções e parâmetros que auxiliam no seu funcionamento, além dos componentes chamados neurônios, que se comunicam através de sinapses, transmitindo informações e conhecimentos. A função de ativação é o elemento responsável por decidir se o neurônio deve ser ativado ou não, também por introduzir não-linearidades no modelo, permitindo que a rede aprenda representações mais complexas. Desta forma, ela permite checar se a informação que está chegando até o neurônio deve ser ignorada ou não. Nas camadas de saída, as funções de ativação podem ter finalidades específicas, dependendo do

problema que a rede neural está tentando resolver. Ela ajuda a decidir se um neurônio vai ser ativado ou não, além de verificar se os pesos definidos são ideais para a rede aprender dado o problema.

2.3 Perovskitas

Perovskita é uma estrutura mineral originalmente descoberta por Gustav Rose em 1839, sendo ela o CaTiO_3 (NAVROTSKY & WEIDNER.,1989). Atualmente, o termo se refere a uma classe de materiais que podem assumir composições diversas e cuja fórmula geral é AB_nX_3 , em que A e B são cátions (sendo A um cátion grande, como os metais alcalinos, alcalinos terrosos e as terras raras, e B um cátion menor, como um metal de transição) e X é um ânion, sendo ele geralmente o íon O (SOARES, 2005).

As perovskitas também podem ser classificadas como duplas (ou complexas), as quais apresentam uma fórmula mais elaborada, $\text{A}_2\text{BB}'\text{X}_6$, envolvendo dois tipos diferentes de cátions 'B' e 'B'', o que permite ajustes ainda mais precisos nas propriedades do material através da substituição de cátions (VASALA & KARPPINEN, 2015).

Em geral, as perovskitas são caracterizadas pela sua ferroeletricidade, alta termoeletricidade e interação de propriedades magnéticas e de transporte (HERNANDEZ, 2018), apresentando uma variedade de aplicações, como em sistemas fotovoltaicos, painéis solares e sistemas magnetocalóricos. Grande parte das propriedades, no entanto, são individuais e atribuídas a fatores como o tamanho dos átomos, a natureza dos óxidos metálicos presentes no sistema e a subclasse de perovskitas a que o material pertence (BELICH *et al.*, 2020). Uma importante subclasse de perovskitas é a do tipo manganita, que apresenta fórmula química: $\text{Ln}_{1-x}\text{A}_x\text{MO}_3$ (Ln = metais de terras-raras, A = metais alcalinos ou bismuto, M = óxidos de metais de transição), como LaMnO_3 , PrMnO_3 , e NdMnO_3 , cujo desempenho é superior em aplicações magnetocalóricas.

Além disso, as manganitas apresentam uma rica variedade de fenômenos físicos, como magnetorresistência colossal, supercondutividade em certas condições, e comportamento de ferroeletricidade em alguns casos. Essas

propriedades tornam esses materiais interessantes para a produção de sensores, memórias magnéticas, e outros dispositivos eletrônicos avançados.

Pesquisas recentes têm se concentrado em explorar a modificação química dessas manganitas, alterando as proporções dos elementos Ln, A, e M, para otimizar suas características para aplicações específicas. Isso inclui melhorar sua capacidade de transferência de calor em aplicações de resfriamento ou aumentar a eficiência de suas propriedades magnéticas para sensores e outras tecnologias emergentes (KOUBAA *et al.*, 2009). Assim, as manganitas continuam a ser um campo de pesquisa em rápida evolução, com grande potencial para avanços tecnológicos significativos.

2.4 Tamanho do cristalito

No contexto de materiais cristalinos, o tamanho de cristalito é uma propriedade fundamental que influencia significativamente o comportamento dos materiais em diversos aspectos, tais como propriedades mecânicas, elétricas, ópticas e magnéticas.

O cristalito, por definição, é uma região dentro de um material cristalino que possui uma orientação uniforme da estrutura cristalina, e pode ter tamanhos que variam de poucos nanômetros a milímetros (HASSANZADEH-TABRIZI, 2023). No entanto, ele pode ser distinto de um grão, que é uma área mais ampla com possíveis subestruturas internas.

Existem diferentes métodos para medir o tamanho de cristalito, sendo a técnica mais comum a difração de raios-X (DRX). A análise da largura da linha de difração permite estimar o tamanho do cristalito utilizando a fórmula de Scherrer (3), que relaciona a largura à meia altura do pico com o tamanho médio do cristalito (MUNIZ *et al.*, 2016). Outros métodos, como a microscopia eletrônica de transmissão, também fornecem uma visão direta do tamanho e da morfologia dos cristalitos, além de informações complementares sobre a estrutura interna.

$$D = \frac{K\lambda}{\beta \cos \theta} \quad (3)$$

O tamanho de cristalito tem um impacto direto nas propriedades dos materiais. Por exemplo, materiais com tamanho de cristalito menor frequentemente exibem maior resistência mecânica, pois a presença de mais interfaces de grãos pode dificultar o movimento de discordâncias. Por outro lado, o aumento do tamanho do cristalito pode afetar as propriedades térmicas e elétricas, como a condução de calor e eletricidade.

O controle do tamanho de cristalito é uma estratégia amplamente utilizada para otimizar materiais em aplicações específicas, como a produção de semicondutores (RIVNAY *et al.*, 2011), catalisadores (MARKOVIC *et al.*, 1997) e materiais magnetocalóricos (ZHANG *et al.*, 2022).

3. MÉTODOS

Neste trabalho, é apresentada uma abordagem sistemática para coletar dados sobre composições de perovskita e suas três propriedades magnetocalóricas: T_c, ME e RCP. Também são discutidos em detalhes os métodos de pré-processamento de dados e técnicas de engenharia de características utilizadas. Além disso, descrevemos os modelos de machine learning empregados e seus parâmetros, ressaltando sua eficiência na previsão das propriedades magnetocalóricas de interesse. A figura 1 ilustra as principais variações no conjunto de dados, nas características e no modelo de machine learning utilizados.

3.1 Base de dados

Os dados foram coletados por meio de extração manual das três propriedades magnetocalóricas de interesse encontradas em gráficos ou tabelas em 146 referências¹. Uma base de dados foi gerada para cada propriedade. O conjunto de dados contém a fórmula química das perovskitas, o valor da propriedade e o tamanho de cristalito (D). O campo magnético aplicado também foi registrado no caso da ME e RCP. Após verificar os valores e a validade das fórmulas químicas,

¹ Link disponível para o github com o acesso aos artigos usados na base de dados: <https://github.com/luis-chemeng/MCML>

uma base de dados separada foi criada para cada propriedade magnetocalórica. Em cada conjunto de dados, os registros foram filtrados para considerar apenas as perovskitas simples com uma composição estequiométrica de oxigênio.

Os conjuntos de dados resultantes continham valores ausentes na coluna de tamanho de cristalito; foram exploradas duas alternativas, pois modelos de machine learning não podem ser treinados com características que contenham valores vazios. Na primeira alternativa, o tamanho de cristalito foi desconsiderado, preservando o número total de amostras do conjunto de dados, o que levou ao denominado "Dataset A". A segunda estratégia consistiu em usar apenas os dados com informações sobre o tamanho dos cristalitos, priorizando a informação fornecida pela propriedade morfológica em detrimento da quantidade de dados, o que resultou no denominado "Dataset B", conforme a figura 1.

Como consequência de não considerar o tamanho de cristalito no conjunto de dados A, algumas entradas desse conjunto continham múltiplos valores da propriedade de interesse para a mesma composição química. Essa condição aumentou o erro nos modelos de machine learning e diminuiu seu desempenho na generalização dos dados. Para resolver esse problema, foram eliminadas as entradas duplicadas e mantivemos apenas o registro correspondente à mediana dos valores da propriedade.

3.2 Obtenção e seleção de características

O módulo Python XenonPy (YAMADA *et al.*, 2019) foi utilizado para obter uma tabela contendo 58 propriedades atômicas para cada elemento da tabela periódica. Este módulo também oferece sete fórmulas combinatórias para derivar descritores de composição a partir das propriedades atômicas dos elementos constituintes do composto, incluindo, soma ponderada, média ponderada, variância ponderada, média geométrica, média harmônica, agregação máxima e agregação mínima. Isso resultou em um vetor com 406 características.

Além disso, foram adicionadas duas novas fórmulas combinatórias. A primeira (MinDelta) é a diferença mínima entre sítios, que calcula o menor valor entre a diferença da propriedade atômica do sítio A ou B com a propriedade atômica do oxigênio, o que gerou 58 descritores adicionais. Essa combinação foi inspirada

pela alta dependência entre esse descritor e a energia de bandgap das perovskitas, conforme relatado por Körbel *et al.* (2016) e posteriormente utilizado por Gladkikh *et al.* (2020). A segunda (TolF) é um descritor semelhante ao fator de tolerância, com uma fórmula análoga ao fator de tolerância de Goldschmidt, o qual se refere à formação e estabilidade das perovskitas e definido pela fórmula 4:

$$t = \frac{(rA + rX)}{\sqrt{2}(rB + rX)} \quad (4)$$

onde rA , rB , rX são os raios iônicos dos cátions A, B e do ânion X, respectivamente. No entanto, ao invés de usar os raios iônicos dos sítios A, B e oxigênio, foram utilizadas todas as 58 propriedades atômicas disponíveis no XenonPy. Como algumas das propriedades atômicas são zero para o sítio B e o oxigênio, o cálculo dessa fórmula foi impossível para 5 das propriedades atômicas, resultando em 53 descritores.

Essa fórmula combinatória foi utilizada com base no trabalho de Franco *et al.*, no qual foi demonstrado que o fator de tolerância possui uma correlação moderada com a T_c (FRANCO *et al.*, 2018). Isso resulta em um vetor com 517 características, designadas como descritores de composição. O tamanho de cristalito foi empregado como um descritor morfológico, sem nenhum refinamento adicional.

Foram implementados quatro esquemas de seleção de características sobre os descritores composicionais: Primeiro, (i) as características foram agrupadas de acordo com o tipo de combinação usado, resultando em 9 grupos, cada um composto por 58 características, exceto pelo grupo de características do tipo fator de tolerância, que possuía 53 características. (ii) consistiu em classificar as 517 características com seu coeficiente de correlação em relação à propriedade alvo; os 10 melhores desempenhos foram selecionados. Três tipos de cálculo de coeficiente de correlação foram aplicados: correlação de Pearson, correlação de Spearman e correlação de distância. A correlação de Pearson, a correlação de Spearman e a correlação de distância são três métodos estatísticos utilizados para medir a associação entre variáveis. A correlação de Pearson avalia a relação linear entre duas variáveis contínuas, assumindo que os dados seguem uma distribuição normal

(PEARSON, 1896). A correlação de Spearman, por outro lado, é uma medida não paramétrica que calcula a relação monotônica entre variáveis, baseando-se nos rankings dos dados em vez dos valores brutos, tornando-a adequada para dados que não seguem uma distribuição normal ou que possuem outliers (SEDGWICK, 2014). Já a correlação de distância é uma técnica mais recente que pode detectar relações não lineares entre variáveis, ao utilizar distâncias entre pontos em vez de suas coordenadas diretas, oferecendo uma maior flexibilidade em detectar associações complexas entre os dados (SZEKELY *et al.*, 2008); esses conjuntos de características foram denominados PC, SC e DC, respectivamente. Para garantir que as informações mais únicas fossem capturadas, foi utilizada uma estratégia de decorrelação, ou seja, as características com alta correlação com uma característica de classificação superior a 90% foram descartadas a fim de evitar redundâncias. (iii) A seleção de características baseada no regressor de árvore de decisão (DT) foi empregada. Para isso, um modelo de regressão de árvore de decisão foi treinado. A relevância das características foi classificada de acordo com o método de eliminação recursiva de características (RFE), que é uma técnica de seleção de características que funciona pela eliminação iterativa das características menos relevantes ou insignificantes (aquelas com o menor coeficiente de Gini) até que um número desejado de características seja alcançado, neste caso, 10. Este conjunto de características foi chamado de 'DT+RFE'. Por fim, o esquema (iv) avaliou um conjunto que inclui todos os descritores composicionais sem nenhuma seleção adicional de características, denominado 'All'.

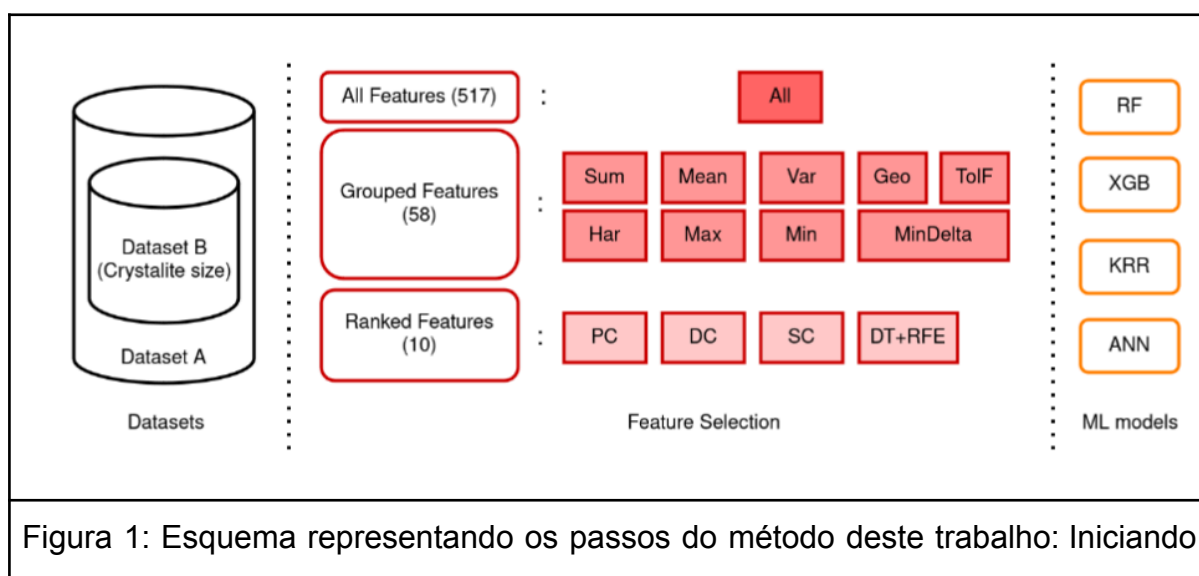


Figura 1: Esquema representando os passos do método deste trabalho: Iniciando

à esquerda com os conjuntos de dados utilizados sem o tamanho do cristalito (Dataset A) e com o tamanho do cristalito (Dataset B). Em seguida, a Seleção de Características, que inclui a classificação e nomenclatura dos conjuntos de características. O lado direito mostra os modelos de machine learning empregados.

3.3 Modelos de Machine Learning

Os modelos de machine learning foram aplicados para modelar as propriedades magnetocalóricas usando cada um dos conjuntos de características previamente definidos. Os dados foram divididos para cada propriedade, respeitando um equilíbrio de 80% para um conjunto de treinamento e 20% para um conjunto de teste. O conjunto de treinamento foi dividido em 5 subconjuntos e, utilizando um processo de treinamento de cinco rodadas, 4 dos 5 subconjuntos foram usados para treinamento e o subconjunto restante foi usado para validação. Esse procedimento, conhecido como validação cruzada (5-fold cross-validation), foi aplicado a uma grade de hiperparâmetros para identificar a melhor combinação de hiperparâmetros que não resultasse em sobreajuste (overfitting) para cada modelo, isto é, aquele que apresentasse o menor erro quadrático médio nas validações. Essa etapa foi realizada através do método GridSearchCV do Scikit-learn (PEDREGOSA *et al.*, 2011).

O desempenho dos modelos otimizados foi avaliado no conjunto de teste utilizando três métricas-chave: o coeficiente de determinação (R^2), que mede a proporção da variância explicada, variando de 0 a 1, erro médio quadrático (RMSE), que mede a raiz quadrada do erro quadrático médio e erro médio absoluto (MAE), que mede os erros absolutos médios.

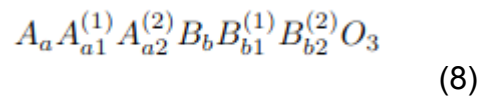
$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

3.4 Predição de novos compostos

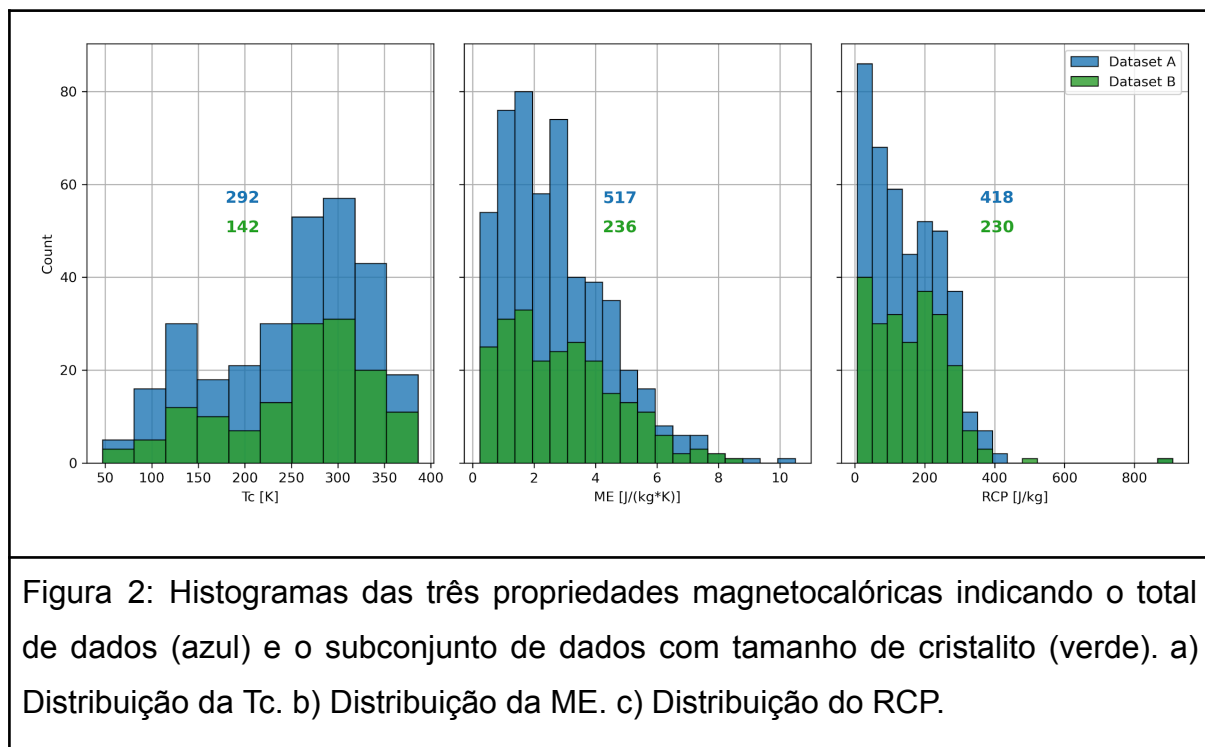
A fórmula sistemática abaixo gerou um espaço químico com potenciais perovskitas (Fórmula 8). As letras A e B correspondem aos sítios A e B da perovskita simples; os sobrescritos 1 e 2 indicam os elementos de dopagem do sítio principal (aqueles sem sobrescritos). O elemento no sítio principal A foi limitado a La, Sr e Nd como os 3 componentes mais comuns e abundantes em nossa base de dados, e a concentração variou entre 0,6 e 1. No caso do sítio principal B, apenas Mn foi considerado, uma vez que a base de dados é quase exclusivamente composta de manganitas. A concentração de Mn variou de 0,8 a 1. Os sítios de dopagem A e B foram selecionados a partir dos elementos em nossa base de dados nos sítios A ou B, respectivamente. A concentração de dopagem nos sítios foi ajustada incrementalmente em intervalos de 0,05. Isso resultou em um conjunto de pouco mais de 1,2 milhão de composições.



4. RESULTADOS

4.1 Descrição dos dados

A base de dados contém registros nas seguintes quantidades: 292 para a T_c , 517 para a ME e 418 para o RCP. A distribuição de cada propriedade é ilustrada na Figura 2. Uma segunda distribuição, que considera apenas os registros que informam o tamanho dos cristalitos, também é exibida. Conforme observado, a distribuição é similar, embora a quantidade de dados seja aproximadamente a metade, com percentuais de 48,6%, 45,6% e 55,0% para T_c , ME e RCP, respectivamente. É notável que mais da metade dos estudos analisados não incluía esse parâmetro, dado que a maioria possui caracterização e análise por Difração de Raios-X (XRD), porém a determinação do tamanho dos cristalitos foi omitida.



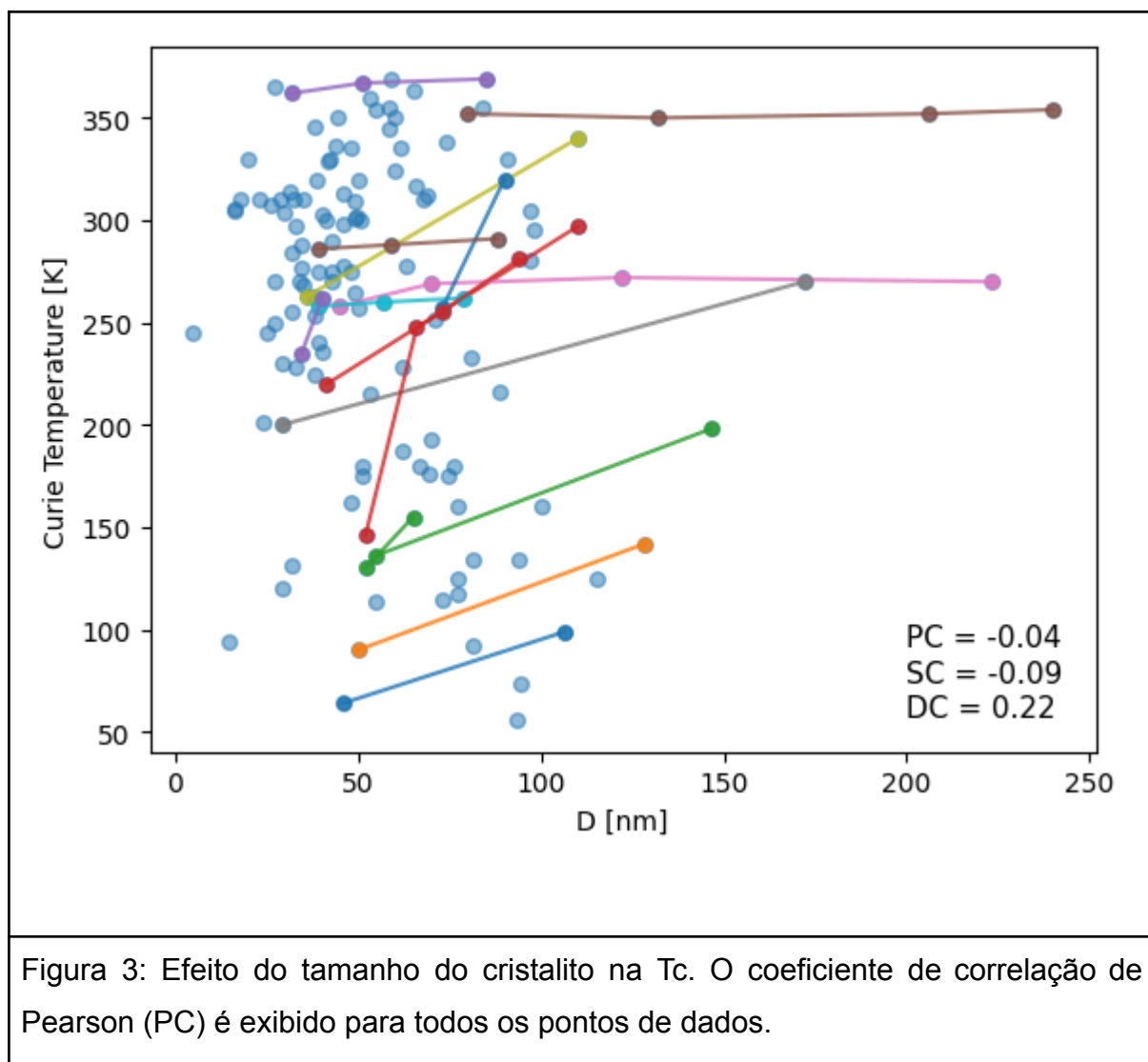
Anteriormente, estudos sobre composições específicas relataram o impacto deste parâmetro na magnitude das propriedades magnetocalóricas (TLILI *et al.*, 2017; MAHJOUB *et al.*, 2015).

Para evidenciar o efeito do tamanho dos cristalitos em todos os dados coletados, os registros com tamanho de cristalito reportado para cada uma das

propriedades foram plotados, conforme pode ser visto nas Figuras 3-5. Os valores destacados em diferentes cores são de estudos nos quais as condições de síntese foram variadas (como o método de síntese, temperatura de sinterização, tempo de sinterização ou pressão aplicada). Assim, perovskitas com a mesma composição química, mas diferentes morfologias e com diferentes tamanhos de cristalitos foram obtidas. Para a T_c , o efeito do tamanho dos cristalitos é proporcional. No entanto, a magnitude desta proporcionalidade varia entre as diferentes perovskitas.

No entanto, a magnitude dessa proporcionalidade varia entre as diferentes perovskitas. Para algumas das composições, o efeito é significativo, enquanto para outras, é quase nulo conforme o tamanho do cristalito aumenta, ou seja, apresenta um comportamento assintótico. Para a ME, o efeito aparenta ser proporcional e mais intenso em relação à T_c . No entanto, é possível observar dois estudos nos quais esse efeito é inverso. Algo semelhante ocorre com o RCP, onde, embora o efeito também pareça positivo, há alguns estudos com um efeito inversamente proporcional; isso não é surpreendente, pois o RCP é diretamente proporcional com ME.

Quanto à tendência geral e ao uso de coeficientes de correlação, embora uma tendência discernível seja observável nos gráficos de dispersão, o coeficiente de correlação não confirma essa relação, especialmente no caso da TC, onde os coeficientes de Pearson e Spearman indicam uma tendência negativa. Para ME e RCP, os coeficientes são baixos. No entanto, o efeito é de fato palpável. Isso destaca a necessidade de coeficientes de correlação para discernir tendências significativas neste problema de informática de materiais, já que eles consideram apenas o impacto de uma variável por vez.



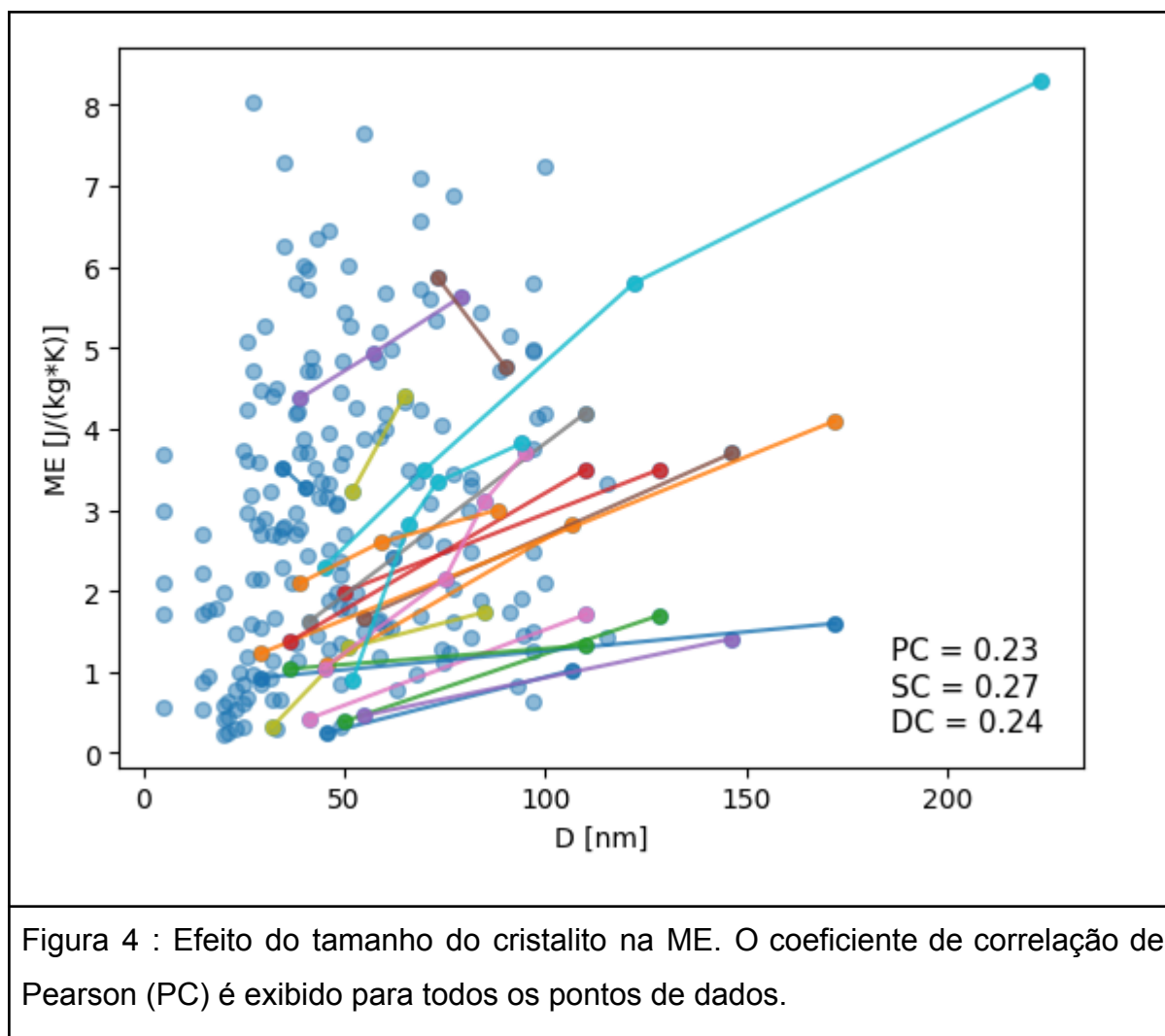


Figura 4 : Efeito do tamanho do cristalito na ME. O coeficiente de correlação de Pearson (PC) é exibido para todos os pontos de dados.

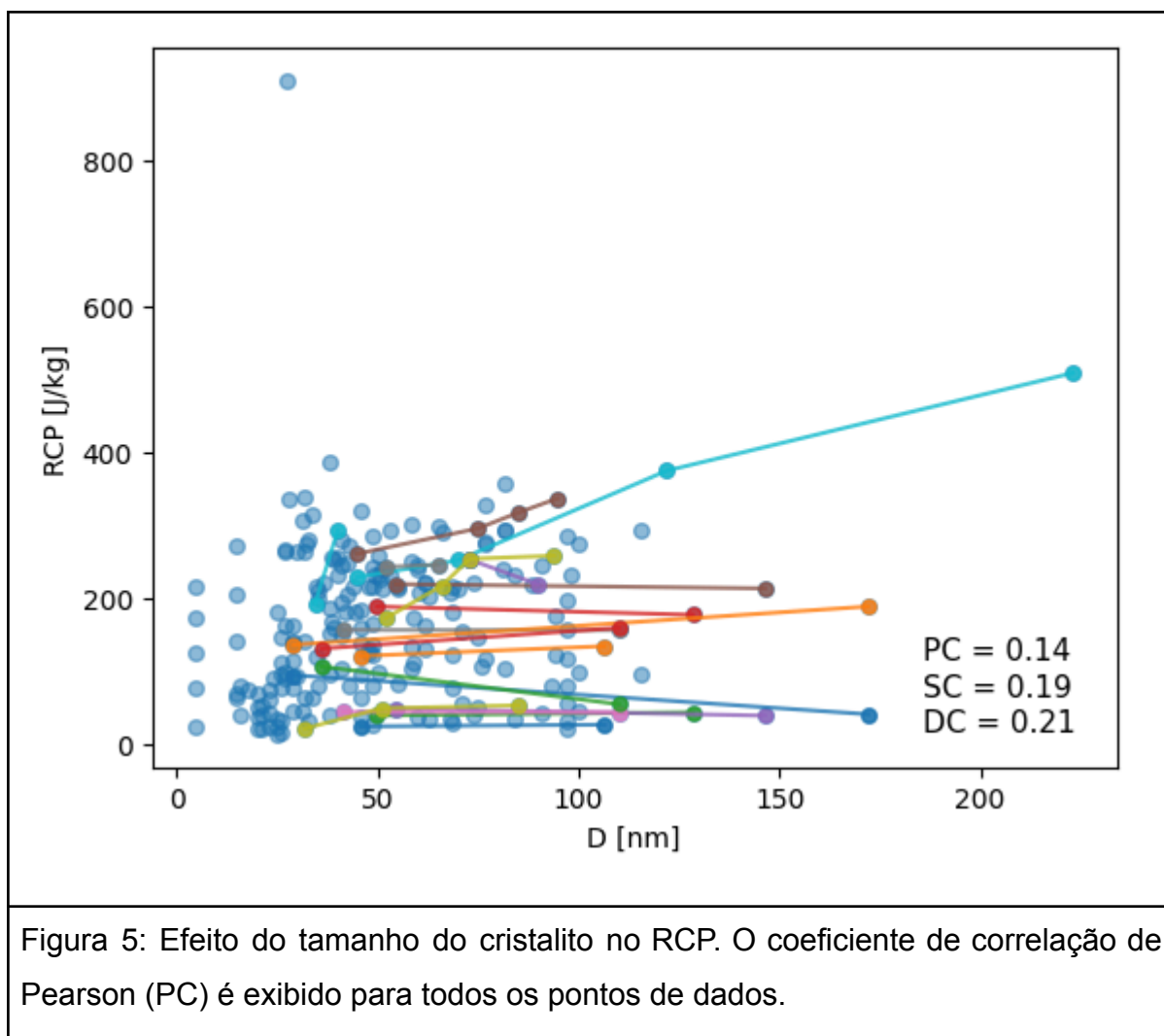


Figura 5: Efeito do tamanho do cristalito no RCP. O coeficiente de correlação de Pearson (PC) é exibido para todos os pontos de dados.

4.2 Comparação de Seleção de Características

Gladkikh *et al.*(2020) relataram como selecionar características para entrada em um algoritmo de machine learning e como mapear essas características para a variável-alvo. Eles realizaram uma análise de correlação com os descritores e a propriedade de interesse usando vários coeficientes de correlação, como os coeficientes de Pearson e Spearman, Correlação de Distância, entre outros. Conclui-se que os graus de correlação entre uma característica e a propriedade em questão devem ser considerados como algo não definitivo, pois podem variar dependendo do coeficiente de correlação utilizado. Neste trabalho, inicialmente foram coletadas todas as características resultantes da fusão das propriedades atômicas com os 9 diferentes tipos de combinações. Posteriormente, buscou-se identificar as características mais influentes para cada uma das três propriedades

magnetocalóricas usando coeficientes de correlação. Ao examinar as relações entre essas características e as propriedades magnetocalóricas, foi possível simplificar o processo de seleção de características e identificar os descritores mais significativos, facilitando a subsequente aplicação de técnicas de machine learning para a previsão precisa de propriedades e o design de materiais.

As Tabelas 1-3 apresentam as 10 melhores características para a TC, ME e RCP, conforme indicado pelos respectivos coeficientes de correlação. Vale mencionar que as características dos conjuntos classificados com base em coeficiente de correlação não possuem alta intercorrelação e não são diretamente comparáveis devido à estratégia de decorrelação adotada. Apesar desses fatores, algumas características específicas se sobressaem nos grupos classificados. Geralmente, as características ordenadas com base nos coeficientes de correlação tendem a ocupar posições muito semelhantes nos rankings. No entanto, ao analisar as características classificadas pelo método DT+RFE, observam-se diferenças em comparação com as classificações baseadas nos coeficientes PC, SC e DC. Isso destaca a relevância de empregar múltiplos métodos na seleção de características. No caso particular da Tc, as características relacionadas ao número de elétrons não preenchidos ou elétrons de valência nas últimas camadas atômicas surgem como a característica mais significativa nos quatro conjuntos. De fato, Liu et al., que estudaram compostos intermetálicos de terras raras, descobriram que a contagem de elétrons 4f influencia significativamente a magnitude da Tc. Esta propriedade atômica também foi uma característica significativa na análise da temperatura crítica de materiais supercondutores (CASTRO *et al.*, 2020). No caso da ME, o Parâmetro de Rede foi a propriedade que mais se destacou. Por outro lado, para o RCP, nenhuma propriedade se destacou significativamente. Isso indica uma interação mais complexa de fatores que influenciam o RCP, destacando a importância de uma análise abrangente para entender seu comportamento. A descrição de cada feature utilizada nos modelos encontra-se no apêndice A.

Tabela 1: 10 melhores preditores de acordo com os métodos de seleção de características (DC, SC, PC, DT+RFE) para Tc

Ranking	PC_decorr	SC_decorr	DC_decorr	DT+RFE
1	Sum	TolF_electron_affinity	TolF_electron_affinity	Mean en_pauling

	num_f_unfilled			
2	Mean num_f_unfilled	MinDelta_num_d_unfilled	MinDelta_num_d_unfilled	Mean num_f_unfilled
3	MinDelta_num_d_unfilled	TolF_evaporation_heat	Mean num_d_unfilled	Var en_allen
4	Mean num_d_unfilled	Min en_allen	Sum num_d_unfilled	Geo electron_affinity
5	Sum num_d_unfilled	Max num_d_unfilled	Sum num_f_unfilled	Geo lattice_constant
6	Var num_f_unfilled	TolF_heat_of_formation	Mean num_f_unfilled	Har num_valance
7	MinDelta_electron_affinity	Har fusion_enthalpy	TolF_num_d_unfilled	TolF_electron_affinity
8	Max num_d_unfilled	Sum num_f_unfilled	MinDelta_electron_affinity	MinDelta_bulk_modulus
9	TolF_electron_affinity	Mean num_f_unfilled	Min en_allen	MinDelta_hhi_r
10	Min en_allen	Mean num_d_unfilled	TolF_evaporation_heat	MinDelta_num_valance

Tabela 2: 10 melhores preditores de acordo com os métodos de seleção de características (DC, SC, PC, DT+RFE) para ME

Ranking	PC_decorr	SC_decorr	DC_decorr	DT+RFE
1	Min lattice_constant	MinDelta_heat_capacity_molar	TolF_lattice_constant	Sum electron_affinity
2	Max bulk_modulus	TolF_lattice_constant	MinDelta_hhi_r	Sum num_f_unfilled
3	MinDelta_thermal_conductivity	Max bulk_modulus	MinDelta_heat_capacity_molar	Var melting_point
4	MinDelta_hhi_r	MinDelta_thermal_conductivity	MinDelta_atomic_radius_rahm	Var molar_volume

5	TolF_lattice_constant	Min lattice_constant	MinDelta_hhi_p	Geo electron_negativity
6	Sum electron_affinity	Mean melting_point	MinDelta_covalent_radii_s_cordero	Geo num_valance
7	MinDelta_hhi_p	Max melting_point	Mean melting_point	Har bulk_modulus
8	Max melting_point	Min thermal_conductivity	Sum electron_affinity	Har lattice_constant
9	Mean electron_affinity	TolF_heat_of_formation	Min lattice_constant	Har sound_velocity
10	Mean bulk_modulus	Min atomic_radius	MinDelta_covalent_radii_slater	TolF_vdw_radius_alvarez

Tabela 3: 10 melhores preditores de acordo com os métodos de seleção de características (DC, SC, PC, DT+RFE) para RCP

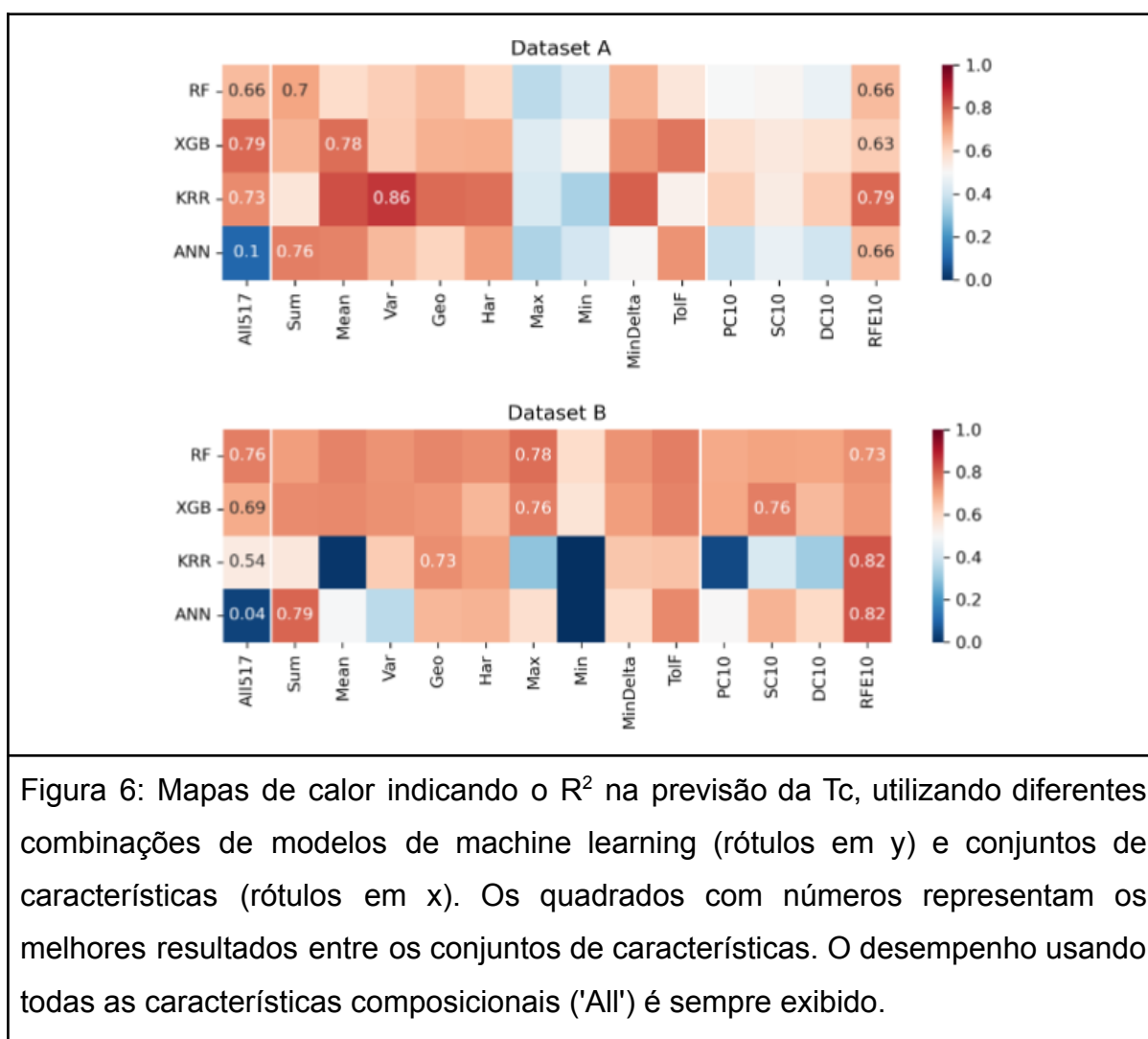
Ranking	PC_decorr	SC_decorr	DC_decorr	DT+RFE
1	Sum electron_affinity	Min thermal_conductivity	MinDelta_vdw_radius_alvarez	Sum heat_capacity_molar
2	Mean electron_affinity	Var atomic_number	MinDelta_vdw_radius	Sum thermal_conductivity
3	MinDelta_vdw_radius_alvarez	Sum electron_affinity	MinDelta_atomic_radius_rahm	Mean en_allen
4	Mean evaporation_heat	Har heat_capacity_molar	MinDelta_num_unfilled	Mean hhi_r
5	Mean heat_capacity_molar	Geo heat_capacity_molar	Mean thermal_conductivity	Mean heat_capacity_molar
6	MinDelta_atomic_radius_rahm	Max electron_affinity	Sum thermal_conductivity	Var atomic_radius_rahm

7	Geo heat_capacity_molar	Geo electron_affinity	MinDelta_c6_gb	Var covalent_radius_cordero
8	Har heat_capacity_molar	Sum heat_capacity_molar	MinDelta_covalent_radiu s_pyykko	Har hhi_p
9	Geo electron_affinity	Mean heat_capacity_molar	MinDelta_atomic_radius	TolF_heat_capacity_mass
10	Sum lattice_constant	Min covalent_radius_slater	Sum electron_affinity	TolF_num_valance

4.3 Modelos de Machine Learning

O desempenho da combinação do conjunto de dados utilizado, do modelo de machine learning otimizado e do conjunto de características para as propriedades previstas (T_c , ME e RCP) é apresentado nas Figuras 6, 8, 10 respectivamente, utilizando a métrica R^2 aplicada ao conjunto de teste, respectivamente. Os mapas de calor mostram a porcentagem da variação nos dados para o modelo de machine learning utilizando todos os descritores composicionais, o melhor desempenho entre os conjuntos das características agrupadas (grouped features) e o melhor desempenho entre os conjuntos das características classificadas (ranked features). A escala varia de 0 (azul escuro) a 1 (vermelho escuro), sendo 0,5 representado pela cor branca. Assim, os tons de vermelho correspondem a coeficientes de determinação superiores a 0,5, e os tons de azul, a erros abaixo desse limiar. Embora as tendências para cada propriedade prevista sejam diferentes e serão discutidas mais adiante, observou-se uma tendência geral: o uso de todos os 517 descritores não apresenta um desempenho significativamente superior ao uso de conjuntos de características agrupadas e classificadas. Em alguns modelos específicos, o desempenho foi muito inferior. Isso pode ser explicado pelo fato de que um número excessivamente grande de entradas introduz muito ruído ao modelo e reduz sua capacidade de identificar padrões.

4.4 Temperatura de Curie



O mapa de calor para a combinação do modelo de ML e conjunto de características revela alguns aspectos interessantes: (i) Para o Conjunto de dados A, os modelos treinados com o agrupamento máximo e mínimo das propriedades atômicas (descritores Máx e Min, respectivamente) apresentam desempenho inferior em comparação aos outros descritores, em todos os 4 modelos de machine learning testados. Isso sugere que os descritores obtidos por essas combinações precisam de melhores informações para prever a T_c . Essa observação ainda é válida no Conjunto de dados B para o conjunto de características Min. Enquanto isso, o conjunto de características Max apresenta um desempenho notavelmente bom no Conjunto de dados B com os modelos XGB e RF. (ii) Ao observar os conjuntos das características classificadas (ranked features) para ambos os conjuntos de dados, a

metodologia de seleção RFE foi dominante. Alcançando até a melhor pontuação R^2 entre todos os conjuntos de características para os modelos RNA e KRR no conjunto de dados B.

Como o objetivo do uso dos modelos de machine learning neste trabalho é prever as propriedades magnetocalóricas, a Figura 7 mostra a melhor combinação do modelo de machine learning e conjunto de características para a previsão da T_c , juntamente com os gráficos de paridade e as métricas de desempenho para os Conjuntos de dados A e B. A Figura 7 também apresenta histogramas de resíduos, indicando a diferença entre os dados coletados e os dados previstos pelos modelos. Para o Conjunto de dados A, ambos os histogramas de treino e teste exibem uma moda simétrica, indicando consistência entre os dados de treinamento e teste. No entanto, uma distinção notável surge para o Conjunto de dados B: os histogramas de treino e teste mostram modas diferentes, notavelmente inclinadas para valores negativos de resíduos. Essa mudança sugere uma subestimação sistemática dos valores experimentais no conjunto de teste, destacando possíveis discrepâncias na precisão da previsão. O melhor modelo foi a KRR, usando o conjunto de dados “A” e “Variância Ponderada”, resultando em um RMSE de 22,6 K, MAE de 16,3 K e R^2 de 0,86. Essas métricas são comparáveis às obtidas por Zhai *et al.* (2018): RMSE de 28,6 K, MAE de 21,21 K e R^2 de 0,85. No entanto, as métricas de Zhai *et al.* (2018) foram calculadas sobre o conjunto de validação e não no conjunto de teste. Portanto, a generalidade do modelo expresso por Zhai *et al.* não é assegurada. Os resultados obtidos também são comparáveis com os de Court *et al.* (2021) que empregaram o módulo Xenon.py para gerar descritores para prever a T_c em ligas de Heusler. Eles obtiveram um RMSE de 60 K e um R^2 de 0,71 usando um modelo XGB. Isso sugere que melhores resultados na previsão da T_c podem ser obtidos usando menos características.

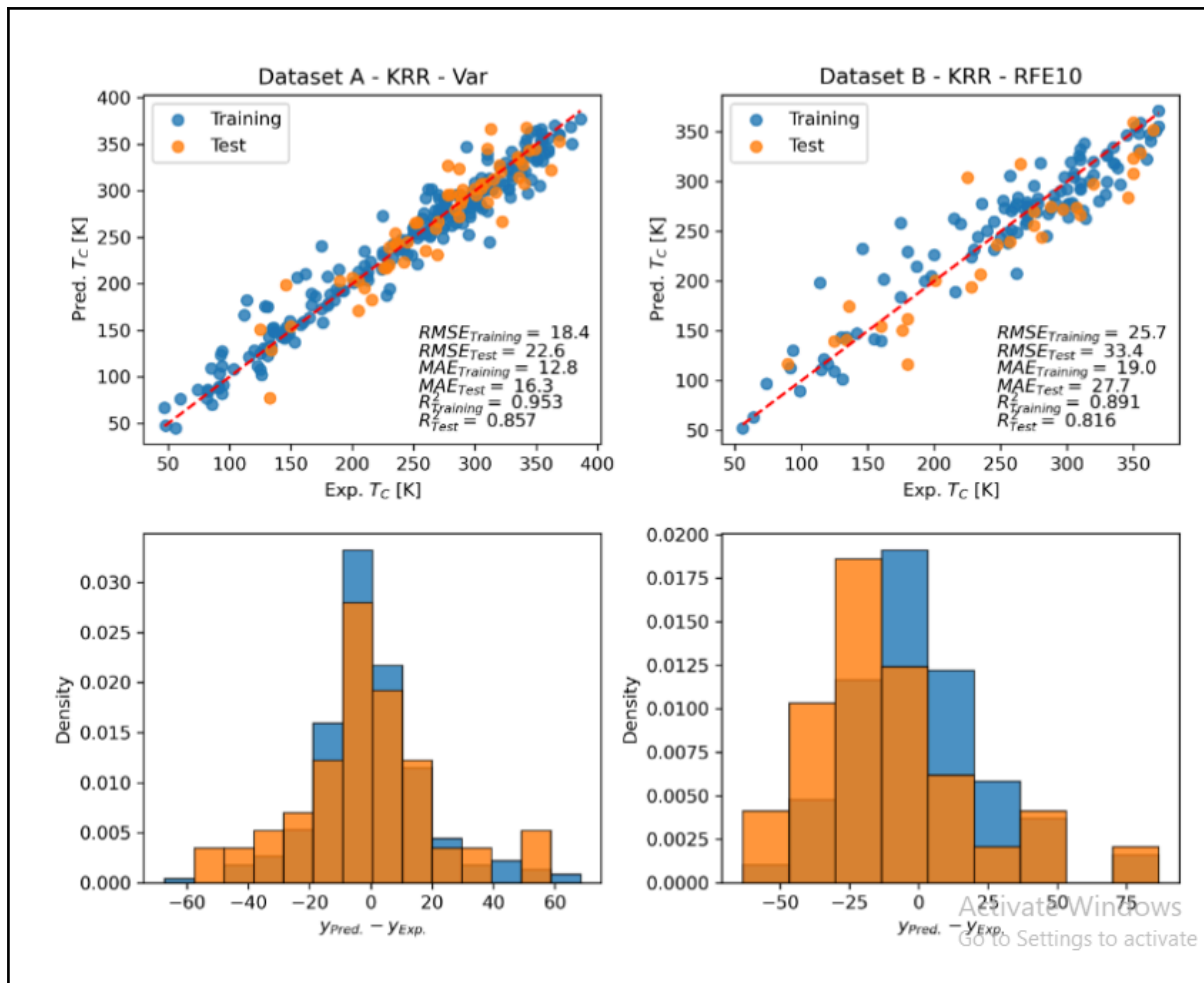
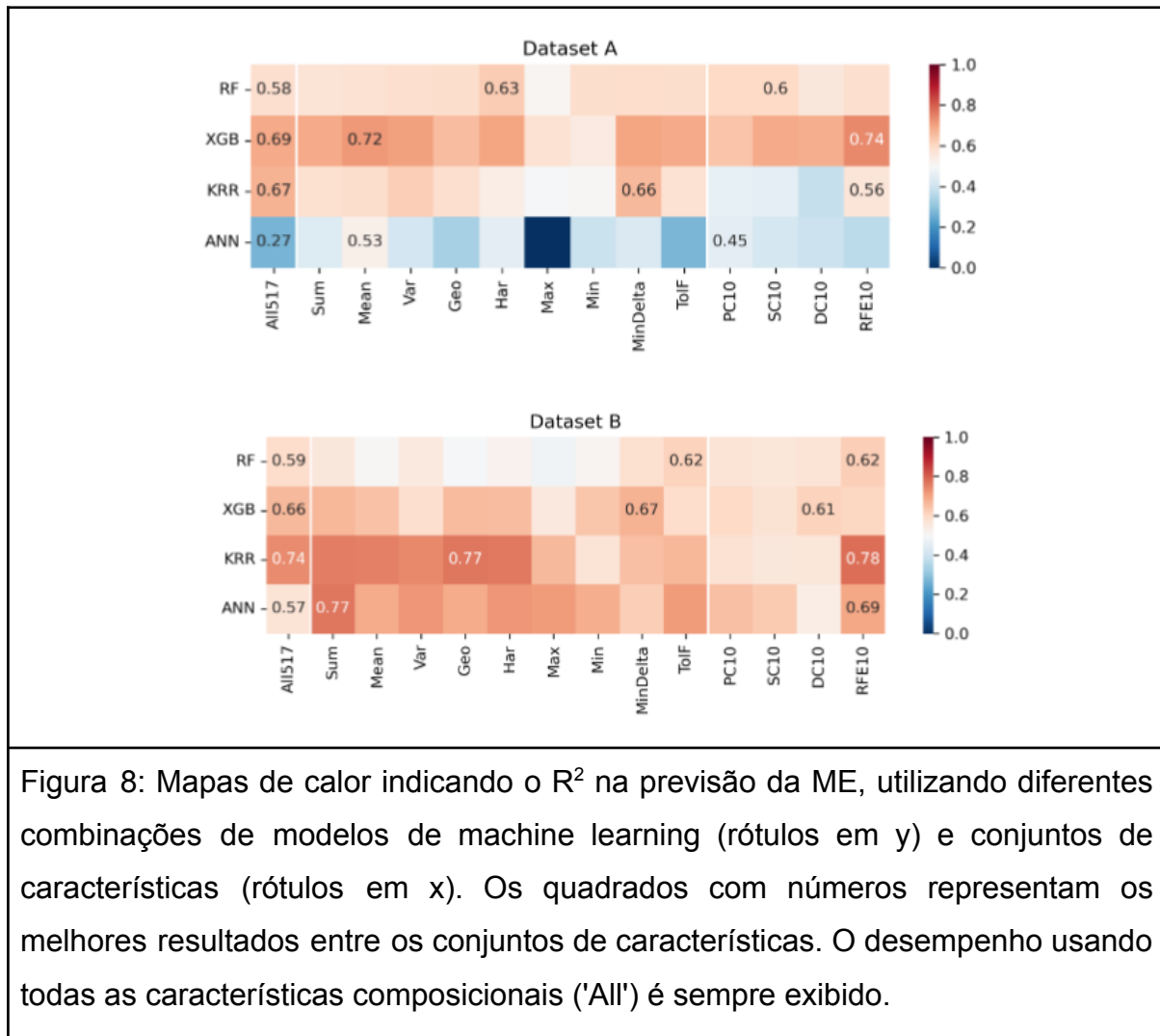


Figura 7: Gráficos de paridade da melhor combinação de modelos de machine learning e conjunto de características para T_C experimental versus previsto de perovskitas magnetocalóricas para dados de treinamento (azul) e teste (laranja). Abaixo de cada gráfico de paridade, é apresentado o histograma normalizado dos resíduos respectivos.

4.5 Mudança Máxima na Entropia Magnética



A Figura 8 indica que, para os Conjuntos de dados A e B, nenhum conjunto de características se destaca dentro das características agrupadas (grouped features), em contraste com as características classificadas, onde o descritor RFE10 emerge como o grupo de características que treina de forma ótima a maioria dos modelos de acordo com o valor de R-quadrado. Quanto à previsão da ME, modelos treinados no Conjunto de dados A demonstram que as propriedades atômicas de agrupamento máximo e mínimo (descritores Max e Min, respectivamente) exibem desempenho inferior em comparação com outros descritores em todos os quatro modelos de machine learning testados. Especificamente sobre o Conjunto de dados A, o modelo RNA consistentemente apresenta desempenho negativo em comparação com outros modelos, independentemente do conjunto de características empregado. No entanto, o modelo XGB se destaca positivamente para este conjunto de dados, demonstrando uma capacidade preditiva superior.

A Figura 9 apresenta a melhor combinação do modelo de ML e conjunto de características para a previsão da ME, os gráficos de paridade e as métricas de desempenho para os conjuntos de dados A e B. A Figura 9 também mostra histogramas de resíduos, indicando a diferença entre os dados coletados e os dados previstos pelos modelos. Ao contrário dos modelos para Tc, os histogramas de resíduos dos modelos para ambos os conjuntos de dados exibem a mesma moda tanto para os dados de teste quanto para os de treinamento, indicando consistência na distribuição dos erros do modelo entre diferentes conjuntos de dados. No entanto, há uma diferença significativa entre os resíduos dos dados de treinamento e teste para o conjunto de dados A, indicando uma situação de sobreajuste (overfitting) que não pôde ser tratada apenas com um esquema de validação cruzada.

De acordo com as métricas do coeficiente de determinação, o melhor modelo foi: Regressão por Kernel Ridge (KRR), utilizando o conjunto de dados "B" e "RFE10" como características, resultando em um RMSE de $0,9 \text{ J kg}^{-1} \text{ K}^{-1}$, MAE de $0,6 \text{ J kg}^{-1} \text{ K}^{-1}$ e R^2 de 0,78. Isso é comparável com os modelos XGBoost construídos por Castro *et al.* (2020) e Court *et al.* (2021), que previram a ME de materiais magnetocalóricos utilizando características geradas pela biblioteca XenonPy e obtiveram um MAE de $1,8 \text{ J kg}^{-1} \text{ K}^{-1}$ (164 pontos de dados de teste) e MAE de $1,2 \text{ J kg}^{-1} \text{ K}^{-1}$ (junto com um R^2 de 0,87 usando 152 pontos de dados de teste), respectivamente. Adicionalmente, Ucar *et al.* (2022) também construíram um regressor Random Forest (RF) para uma variedade de materiais magnetocalóricos e obtiveram um MAE de $2,26 \text{ J kg}^{-1} \text{ K}^{-1}$, RMSE de $3,23 \text{ J kg}^{-1} \text{ K}^{-1}$ e R^2 de 0,82. Mantendo em mente as diferenças na quantidade de dados de cada trabalho, os resultados da nossa seleção de modelo e conjunto de características parecem promissores, mesmo quando comparados aos resultados com tamanho de cristalito, que também apresenta um MAE menor ($0,6 \text{ J/(kg K)}$). Embora o MAE obtido neste trabalho seja menor que o relatado por Court *et al.* (2021)., o R^2 é pior. Esses resultados devem ser verificados com mais dados, pois o tamanho de cristalito não é comumente relatado como descrito na seção de descrição do conjunto de dados, afetando a precisão do modelo de ML.

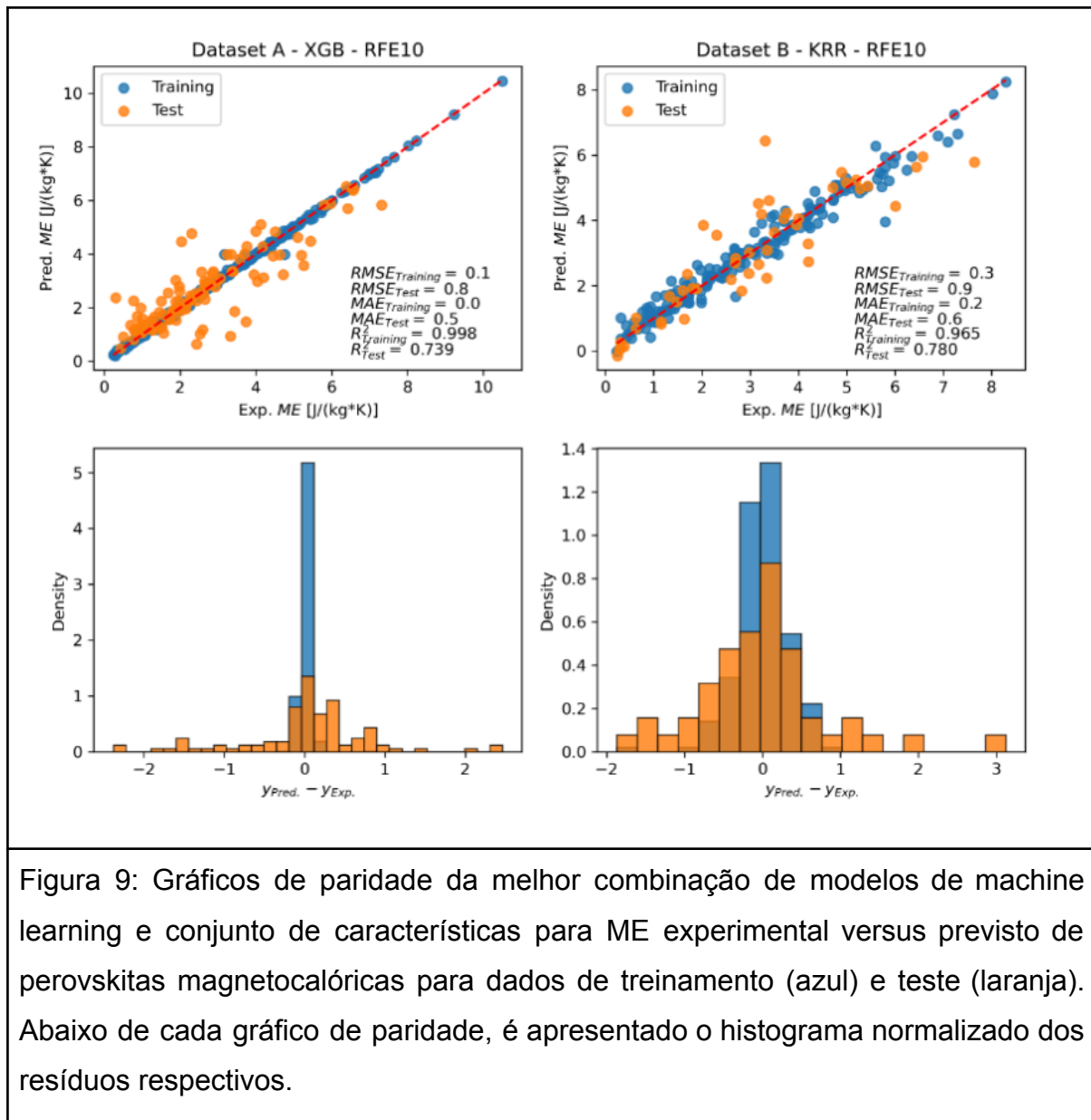


Figura 9: Gráficos de paridade da melhor combinação de modelos de machine learning e conjunto de características para ME experimental versus previsto de perovskitas magnetocalóricas para dados de treinamento (azul) e teste (laranja). Abaixo de cada gráfico de paridade, é apresentado o histograma normalizado dos resíduos respectivos.

4.6 Relative Cooling Power

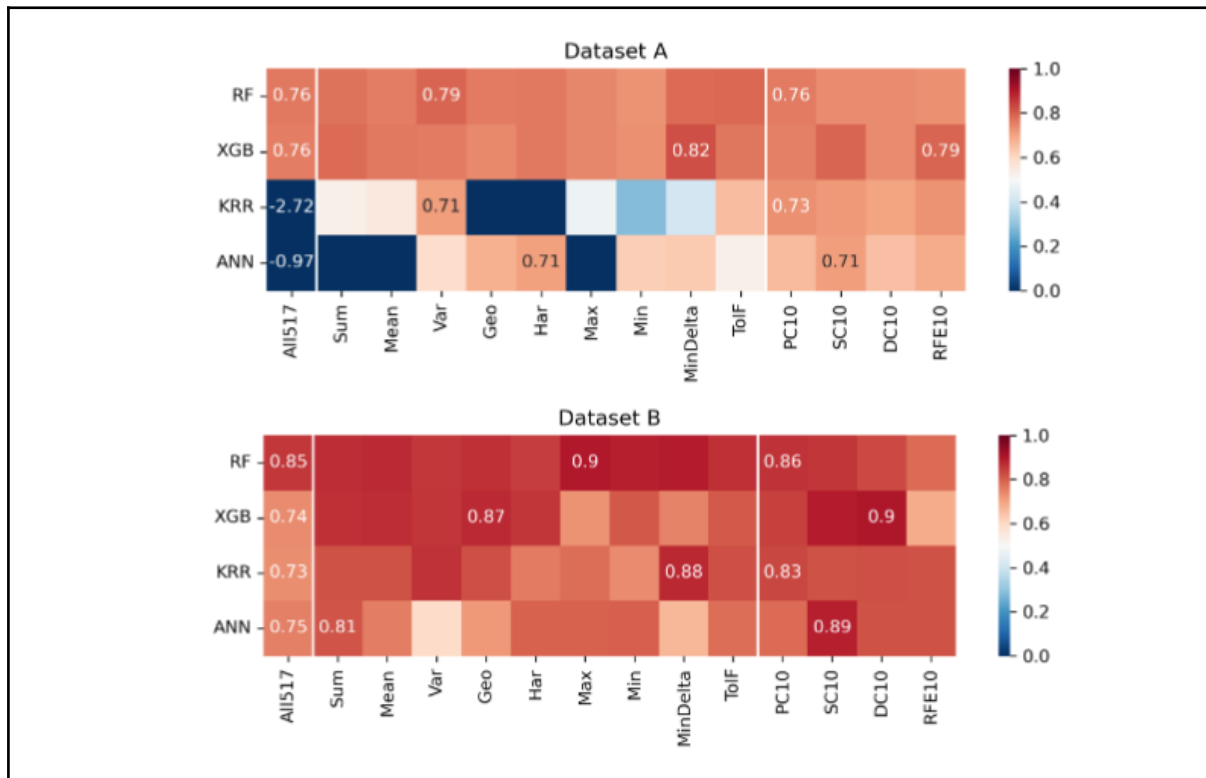


Figura 10: Mapas de calor indicando o R^2 na previsão do RCP, utilizando diferentes combinações de modelos de machine learning (rótulos em y) e conjuntos de características (rótulos em x). Os quadrados com números representam os melhores resultados entre os conjuntos de características. O desempenho usando todas as características composicionais ('All') é sempre exibido.

Os resultados para esta propriedade específica (Figura 10) revelam que os modelos treinados com o conjunto de dados B exibiram um desempenho superior em comparação com aqueles treinados com o conjunto de dados A. Notavelmente, tanto os modelos KRR quanto RNA mostraram inferioridade em relação aos outros quando aplicados ao conjunto de dados A. Curiosamente, em ambos os conjuntos de dados, nenhum conjunto de características se destacou significativamente entre as características agrupadas e classificadas. Isso sugere a importância da seleção de características entre os vários conjuntos, indicando a complexidade dos conjuntos de dados e os desafios inerentes à identificação de características dominantes para a modelagem preditiva. A análise do histograma de resíduos (apresentado na Figura 11) mostra a mesma moda para os dados de treinamento e teste no conjunto de dados A, enquanto para o conjunto de dados B, isso não está claro. Em ambos os casos, foi observada uma situação de sobreajuste.

Tabela 4: Resumo das métricas para cada propriedade e cada conjunto de dados, abrangendo os dados de treinamento e teste.

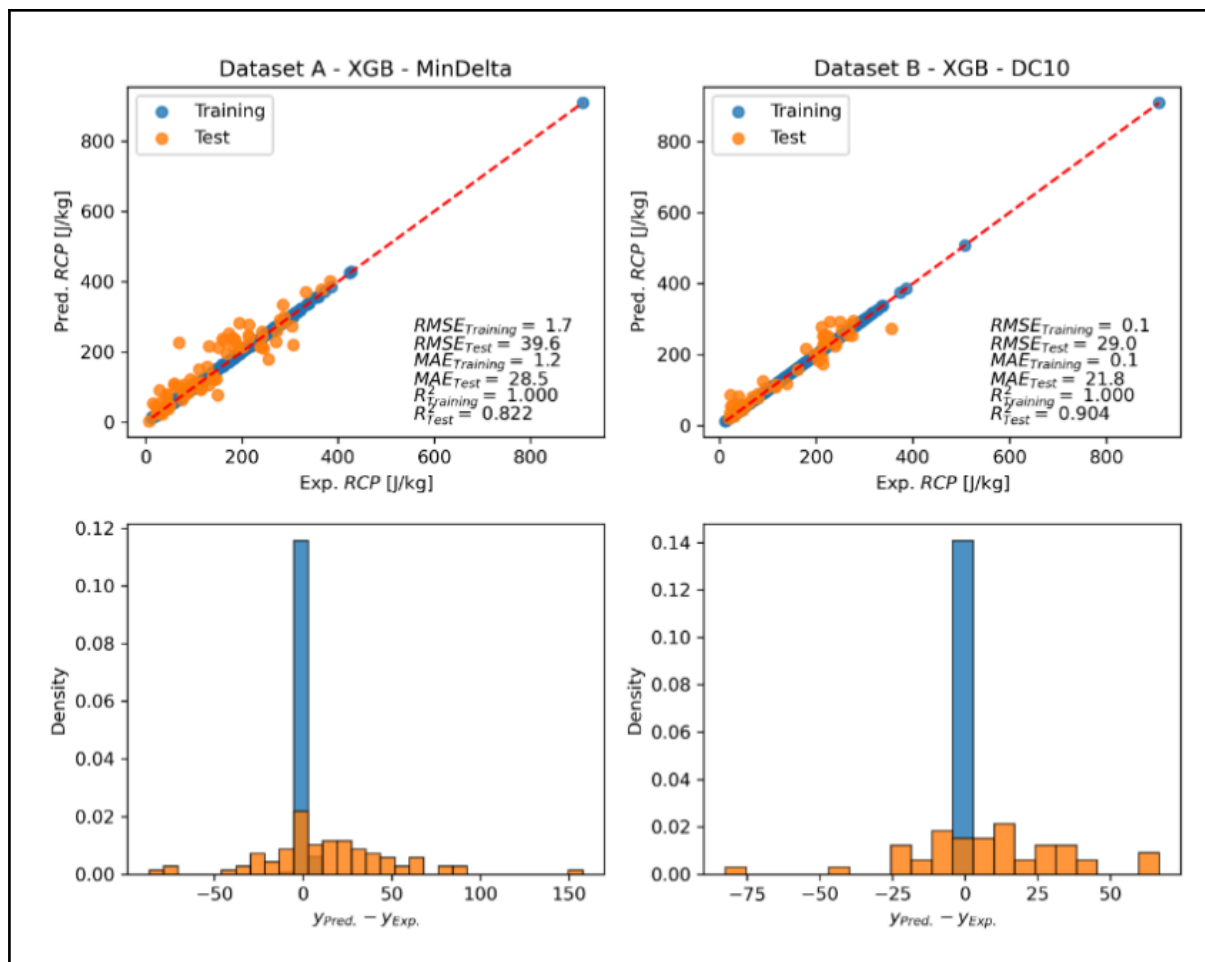


Figura 11: Gráficos de paridade da melhor combinação de modelos de machine learning e conjunto de características para RCP experimental versus previsto de perovskitas magnetocalóricas para dados de treinamento (azul) e teste (laranja). Abaixo de cada gráfico de paridade, é apresentado o histograma normalizado dos resíduos respectivos.

De acordo com todas as métricas de avaliação, o melhor modelo foi o XGBoost usando o conjunto de dados "B" e "DC10" como o melhor grupo de características, resultando em um RMSE de $29 \text{ J kg}^{-1} \text{ K}^{-1}$, MAE de $21.8 \text{ J kg}^{-1} \text{ K}^{-1}$ e R^2 de 0.9. Isso é comparável ao modelo XGBoost construído por Court *et al.* (2021), que apresentou um MAE de $16.8 \text{ J kg}^{-1} \text{ K}^{-1}$ e um R^2 de 0.92 (88 pontos de dados de teste).

Tabela 4: Resumo das métricas para cada propriedade e cada conjunto de dados, abrangendo os dados de treinamento e teste.

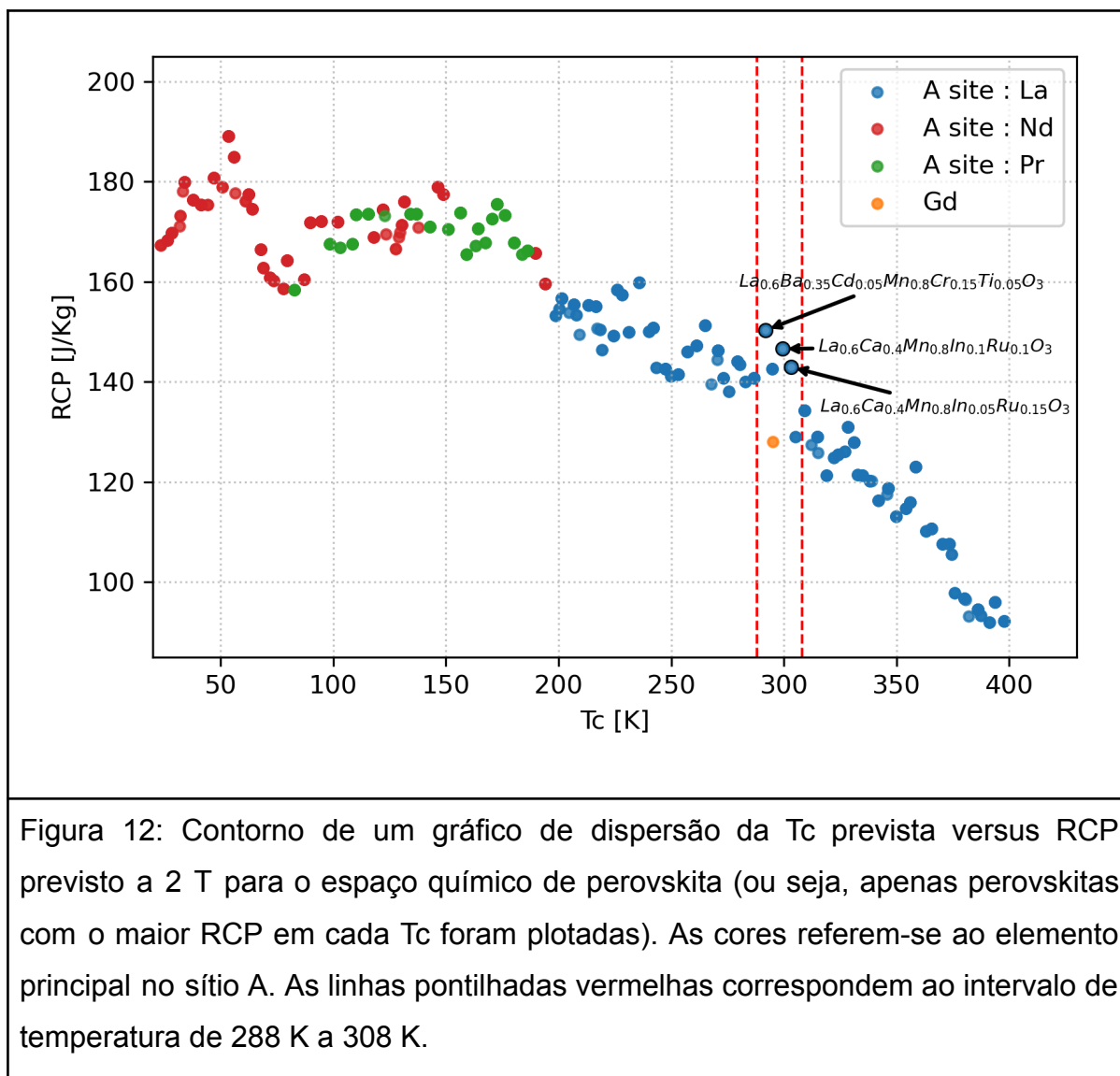
	Tc		ME		RCP	
Metric						
Performance	Dataset A	Dataset B	Dataset A	Dataset B	Dataset A	Dataset B
RMSE train	18.4	25.7	0.1	0.3	1.7	0.1
RMSE test	22.6	33.4	0.8	0.9	39.6	29.0
MAE train	12.8	19.0	0.0	0.2	1.2	0.1
MAE test	16.3	27.7	0.5	0.6	28.5	21.8
R2 train	0.953	0.891	0.998	0.965	1.00	1.00
R2 test	0.857	0.816	0.739	0.780	0.822	0.904

4.7 Análise dos Modelos

Nesta seção, modelos de machine learning foram utilizados para explorar o espaço químico das perovskitas magnetocalóricas, conforme descrito na seção de métodos. Para isso, optou-se por modelos treinados com o conjunto de dados A, devido ao seu maior volume de dados, o que proporciona maior robustez em comparação com os modelos treinados no conjunto de dados B. Além disso, como o parâmetro morfológico (tamanho do cristalito) não aparece como um parâmetro livre, podemos focar apenas na composição química. Para conduzir essa análise, utilizamos a figura de mérito esquematizada por Gottschall *et al.* (2019), onde a T_c está localizada no eixo x e um indicador da intensidade do efeito magnetocalórico, no eixo y. Dessa forma, as Figuras 12 e 13 apresentam exclusivamente as composições com o RCP ou ME previstos mais altos, respectivamente, para cada valor de T_c previsto. As composições foram categorizadas pelo elemento químico predominante presente no sítio A (La, Nd, Pr). Além disso, a análise destaca composições que se sobressaem dentro de uma faixa de temperatura de 288 K a 308 K, ideal para aplicações magnetocalóricas em temperatura ambiente. O gadolínio, material padrão usado neste contexto (COURT *et al.*, 2021), também é plotado para comparação.

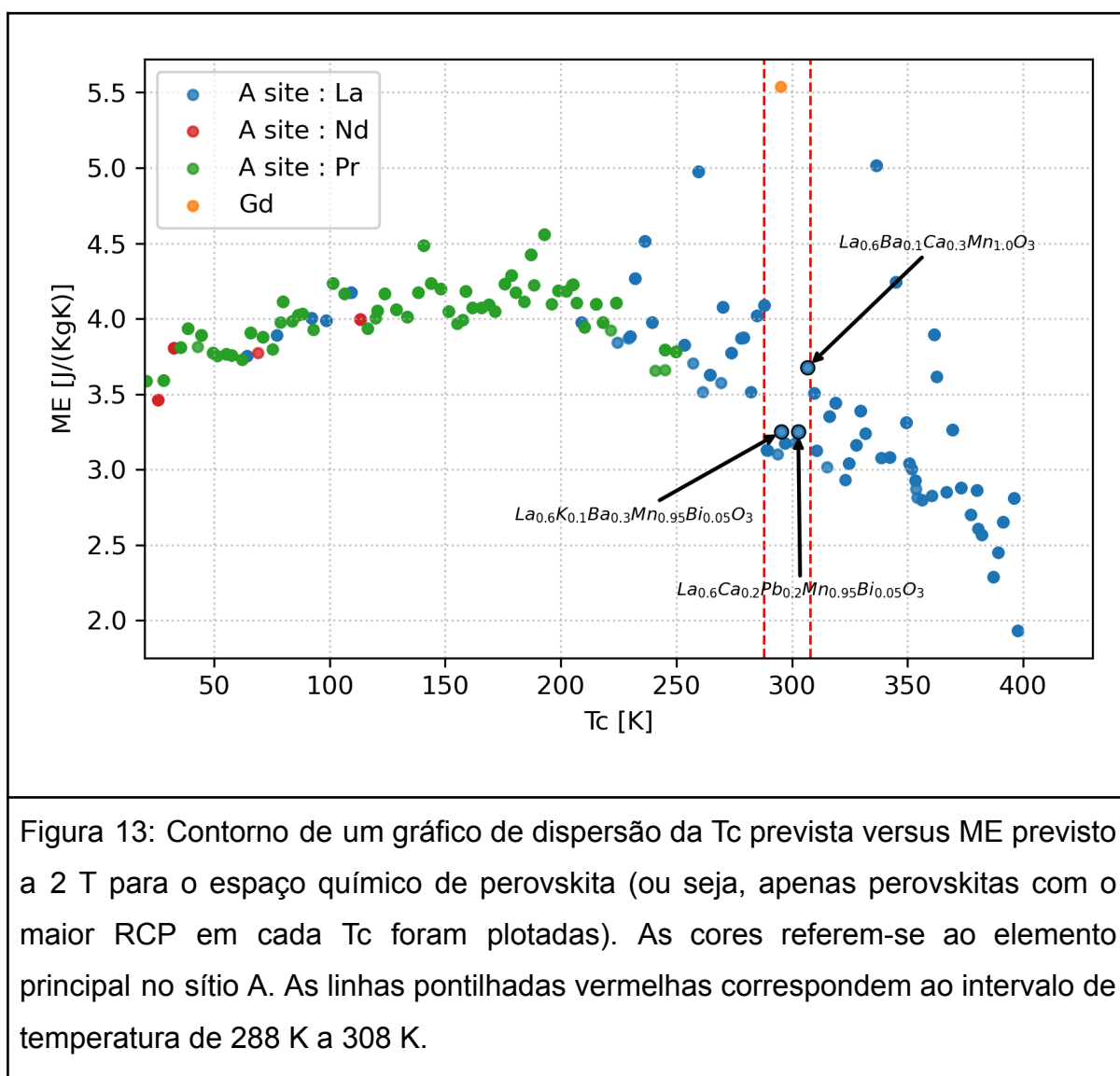
De acordo com a Figura 12, as manganitas com Nd como principal elemento no sítio A apresentam o maior RCP para aplicações em temperaturas próximas a 50 K. Em uma faixa de temperaturas intermediárias (em torno de 160 K), as manganitas com praseodímio são apontadas como as perovskitas mais adequadas. As manganitas com lantânio são mais indicadas para aplicações em temperaturas mais altas. Outra tendência interessante prevista pelos modelos é a diminuição do RCP à medida que a temperatura aumenta para o espaço químico estudado.

Para o intervalo específico de temperatura ambiente, três composições de manganitas foram reportadas com um RCP significativamente maior do que o exibido pelo Gadolínio, indicando seu potencial promissor como composições alternativas para tais aplicações. Nenhum desses compostos foi sintetizado até o momento da escrita deste trabalho; portanto, essas propriedades não podem ser garantidas; no entanto, dentro do banco de dados coletado, foram encontrados exemplos semelhantes que comprovam que esses compostos possuem tais características. Esse é o caso do $\text{La}_{0.6}\text{Ca}_{0.4}\text{MnO}_3$, que exibiu uma T_c de 250 K e um RCP de $160 \text{ J kg}^{-1} \text{ K}^{-1}$.



Quanto à ME prevista (Figura 13), a situação é diferente. Primeiramente, a tendência de diminuir com a temperatura neste caso deve ser esclarecida. Em contraste, as maiores ME são alcançadas em alguns casos de alta temperatura. Em relação às categorias do sítio A, a faixa de temperaturas baixas a intermediárias é dominada pelas manganitas com Pr na sua composição, e existem apenas algumas manganitas com Nd com valores de ME notável. Na faixa de temperaturas de 250 K e acima, as manganitas com La são as candidatas mais promissoras.

Nenhuma das composições à temperatura ambiente apresentou magnitude comparável à do gadolínio. Apesar de divergir da análise anterior, é compreensível considerando que a curva de ME do gadolínio é estreita e sua ME de pico é excepcionalmente elevada. Assim, sua entropia magnética geralmente é muito alta, enquanto seu RCP não é tão proeminente.



4.8 Efeito do tamanho do Cristalito

Nesta seção, utilizou-se a melhor combinação de modelo e conjunto de características para cada propriedade a fim de explorar o impacto do tamanho do cristalito para cada propriedade. Foram selecionadas quatro composições que possuíam valores para pelo menos dois tamanhos de cristalito. As Figuras 14 - 16 apresentaram as previsões como linhas contínuas, com a área sombreada representando o RMSE.

Para a T_c , o efeito é positivo ou, pelo menos, constante, o que está alinhado com a discussão na seção de descrição do conjunto de dados. Quanto à ME, o comportamento não é monótono para uma das composições analisadas, apresentando uma redução da ME em tamanhos de cristalito maiores. Isso pode ocorrer devido à escassez de dados na região de tamanhos de cristalito maiores, tornando o modelo menos robusto nesta região. Em relação ao RCP, a curva resultante exibe um comportamento não monótono e descontínuo, assemelhando-se a um padrão de escada. Essa particularidade deve-se à natureza do modelo XGBoost, que emprega árvores de decisão, fazendo com que cada previsão percorra um caminho através dos ramos das árvores de decisão. Essa natureza não contínua é considerada inadequada para o efeito contínuo que o tamanho do cristalito exerce.

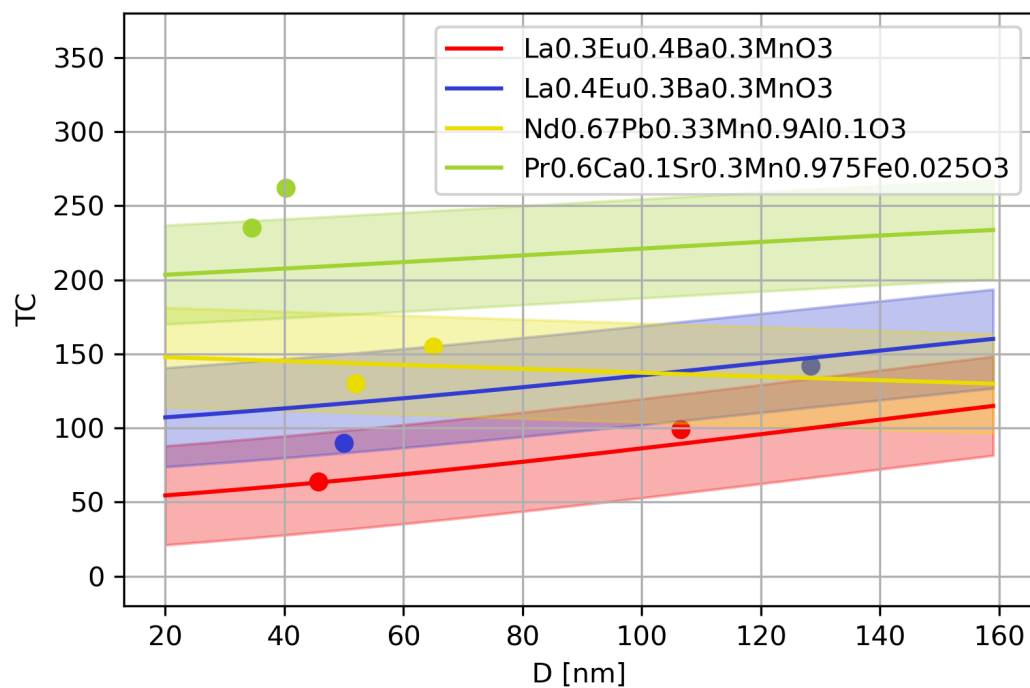


Figura 14: Previsões de T_c como função do tamanho do cristalito para o modelo KRR em combinação com o conjunto de características RFE10. Os pontos correspondem aos dados experimentais.

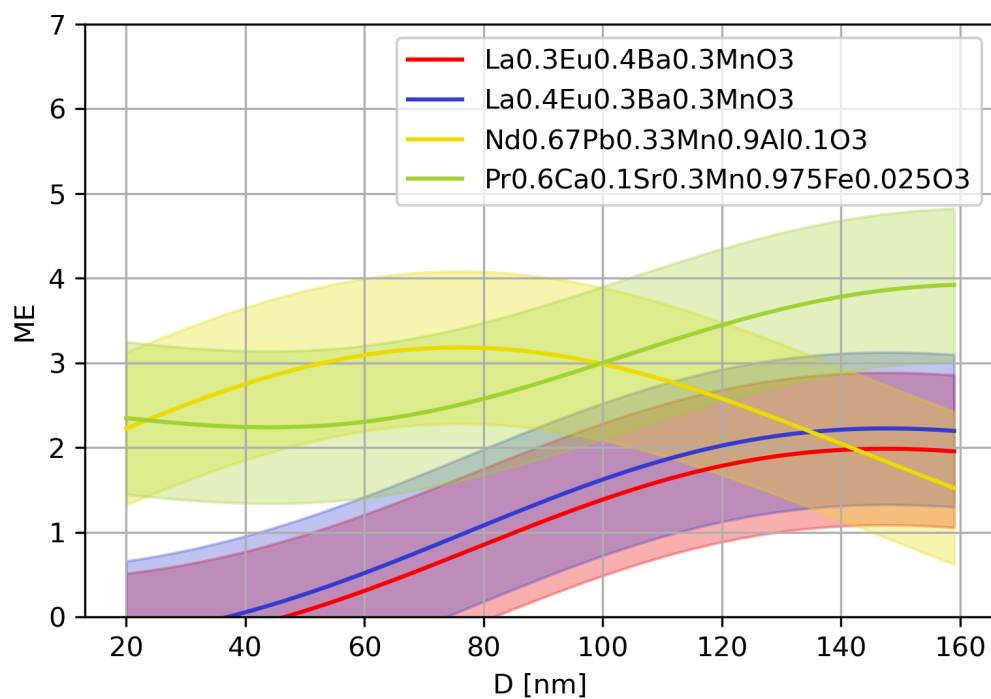


Figura 15: Previsões de ME como função do tamanho do cristalito para o modelo KRR em combinação com o conjunto de características RFE10. Para um campo magnético de 2 T.

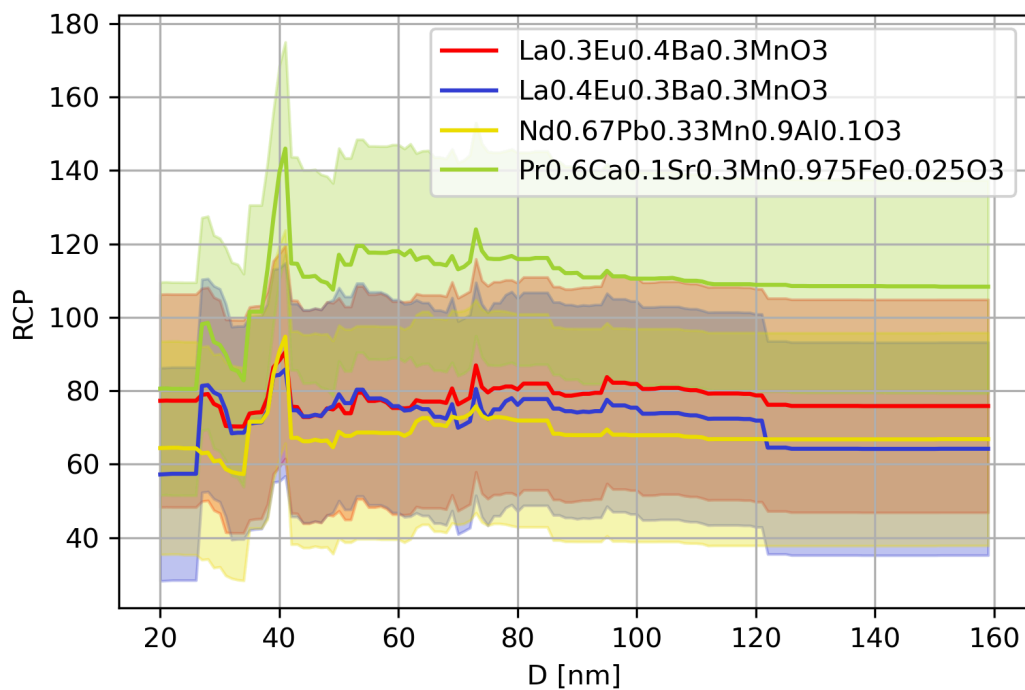


Figura 16: Previsões de RCP como função do tamanho do cristalito para o modelo XGB em combinação com o conjunto de características DC10. Para um campo magnético de 2 T.

5. CONCLUSÃO

Este trabalho envolveu o desenvolvimento e avaliação de 4 diferentes modelos de ML, utilizando conjuntos variados de características para calcular as principais propriedades magnetocalóricas de perovskitas. O estudo focou na exploração de um subespaço químico de manganitas para aplicações magnetocalóricas. Um banco de dados das propriedades magnetocalóricas de centenas de perovskitas foi compilado usando dados coletados da literatura científica. A análise dos dados revelou a importância do tamanho do cristalito como descritor morfológico e sua ausência em mais da metade dos dados coletados. Isso resultou na condução deste estudo com base nos dados completos, excluindo as características relacionadas ao tamanho do cristalito e utilizando o conjunto de dados que não incorpora este parâmetro morfológico.

Quanto aos conjuntos de características, observou-se que níveis similares de precisão podem ser alcançados tanto utilizando todos os 517 descritores composicionais quanto selecionando os 10 melhores dentre eles, demonstrando que modelos de machine learning mais simples podem ser construídos com menos características. A melhor precisão para quase todas as propriedades foi obtida com os subconjuntos das 10 melhores características classificadas. Entre os métodos de seleção de características, a RFE+DT mostrou-se uma alternativa melhor do que os métodos de classificação baseados em coeficientes de correlação para T_c e ME. Para RCP, nenhum método específico de seleção de características foi predominante.

As capacidades preditivas dos melhores modelos foram utilizadas para explorar o efeito magnetocalórico dentro de um subespaço químico de manganitas. De acordo com as previsões, as perovskitas ainda se mostram inferiores ao gadolínio em aplicações magnetocalóricas à temperatura ambiente devido ao seu elevado valor de ME. No entanto, composições específicas de perovskita exibiram um RCP superior ao do Gd, sugerindo que elas poderiam superar o Gd em uma faixa mais ampla de variações de temperatura.

Para futuros trabalhos envolvendo DRX, seria extremamente benéfico reportar o tamanho de cristalito, visando obter mais dados e construir bases de dados mais robustas. Além disso, como continuidade deste trabalho, recomenda-se a síntese das composições de perovskitas que apresentaram elevados RCPs à

temperatura ambiente. Esta abordagem permitirá verificar se os valores previstos pelo nosso modelo correspondem aos valores obtidos experimentalmente, proporcionando uma validação mais rigorosa dos resultados e potencial aplicação prática das descobertas.

6. REFERÊNCIAS

ANDRADE, Vivian Maria Campos Soares de. Manganitas nanoestruturadas: um estudo do Efeito Magnetocalórico. 2015. 1 v. Dissertação (Mestrado)- Curso de Física, Universidade Federal Fluminense, Niterói, 2015.

AYAŞ, Ali Osman et al. Magnetic refrigeration: Current progress in magnetocaloric properties of perovskite manganite materials. **Materials Today Communications**, p. 105988, 2023.

BALACHANDRAN, Prasanna V. et al. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. **Nature communications**, v. 9, n. 1, p. 1668, 2018.

BELICH, Nikolai et al. Perovskite puzzle for revolutionary functional materials. **Frontiers in Chemistry**, v. 8, p. 550625, 2020.

BELO, J. H. et al. On the Curie temperature dependency of the magnetocaloric effect. **Applied Physics Letters**, v. 100, n. 24, 2012

BRUNTON, Steven L.; KUTZ, J. Nathan. **Data-driven science and engineering: Machine learning, dynamical systems, and control**. Cambridge University Press, 2022..

CASTRO, Pedro Baptista de et al. Machine-learning-guided discovery of the gigantic magnetocaloric effect in HoB₂ near the hydrogen liquefaction temperature. **NPG Asia Materials**, v. 12, n. 1, p. 35, 2020.

Chen, T. He 2023, 1–4.

COURT, Callum J.; JAIN, Apoorv; COLE, Jacqueline M. Inverse design of materials that exhibit the magnetocaloric effect by text-mining of the scientific literature and generative deep learning. **Chemistry of Materials**, v. 33, n. 18, p. 7217-7231, 2021.

ELYACOUBI, AS Erchidi; MASROUR, R.; JABAR, A. Magnetocaloric effect and magnetic properties in SmFe_{1-x}Mn_xO₃ perovskite: Monte Carlo simulations. **Solid State Communications**, v. 271, p. 39-43, 2018.

FRANCO, V. et al. Magnetocaloric effect: From materials research to refrigeration devices. **Progress in Materials Science**, v. 93, p. 112-232, 2018.

GAO, Chaochao et al. Innovative materials science via machine learning. **Advanced Functional Materials**, v. 32, n. 1, p. 2108044, 2022.

GÉRON, Aurélien. **Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow**. " O'Reilly Media, Inc.", 2022.

GLADKIKH, Vladislav et al. Machine learning for predicting the band gaps of ABX₃ perovskites from elemental properties. **The Journal of Physical Chemistry C**, v. 124, n. 16, p. 8905-8918, 2020.

GOTTSCHELL, Tino et al. Making a cool choice: the materials library of magnetic refrigeration. **Advanced Energy Materials**, v. 9, n. 34, p. 1901322, 2019.

GSCHNEIDNER, Karl A.; PECHARSKY, V. K.; TSOKOL, A. O. Recent developments in magnetocaloric materials. **Reports on progress in physics**, v. 68, n. 6, p. 1479, 2005.

HASSANZADEH-TABRIZI, S. A. Precise calculation of crystallite size of nanomaterials: A review. **Journal of Alloys and Compounds**, p. 171914, 2023.

HE, Jinrong et al. Kernel ridge regression classification, *2014 International Joint Conference on Neural Networks (IJCNN)*, Beijing, China, 2014, pp. 2263-2267, doi: 10.1109/IJCNN.2014.6889396.

HERNANDEZ, José Luis Valenzuela. Obtenção e caracterização de filmes de perovskitas do tipo SrFeO_{3-δ} para uso em spintrônica baseada em seleção aumentada de spins. 2018. 1 v. Tese (Doutorado)- Curso de Ciência dos Materiais, Instituto de Física, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018

KITANOVSKI, Andrej. Energy applications of magnetocaloric materials. **Advanced Energy Materials**, v. 10, n. 10, p. 1903741, 2020.

KÖRBEL, Sabine; MARQUES, Miguel AL; BOTTI, Silvana. Stability and electronic properties of new inorganic perovskites from high-throughput ab initio calculations. **Journal of Materials Chemistry C**, v. 4, n. 15, p. 3157-3167, 2016.

KOUBAA, W. Cheikh-Rouhou; KOUBAA, M.; CHEIKHROUHO, A. Effect of potassium doping on the structural, magnetic and magnetocaloric properties of La_{0.7}Sr_{0.3- x}K_xMnO₃ perovskite manganites. **Journal of alloys and compounds**, v. 470, n. 1-2, p. 42-46, 2009.

LOUDAINI, A. et al. Electronic, magnetic properties and magnetocaloric effect in La_{0.67}Sr_{0.33}MnO₃ compound: Ab initio calculations and Monte Carlo simulation. **Solid State Communications**, v. 295, p. 5-11, 2019.

LONG, Teng et al. An accelerating approach of designing ferromagnetic materials via machine learning modeling of magnetic ground state and Curie temperature. **Materials Research Letters**, v. 9, n. 4, p. 169-174, 2021.

MAHJOUB, Saoussen et al. Effect of synthesis techniques on structural, magnetocaloric and critical behavior of $\text{Pr}_{0.6}\text{Ca}_{0.1}\text{Sr}_{0.3}\text{Mn}_{0.975}\text{Fe}_{0.025}\text{O}_3$ manganites. **Ceramics International**, v. 41, n. 9, p. 12407-12416, 2015.

MARKOVIC, Nenad; GASTEIGER, Hubert; ROSS, Philip N. Kinetics of oxygen reduction on Pt (hkl) electrodes: implications for the crystallite size effect with supported Pt electrocatalysts. **Journal of the Electrochemical Society**, v. 144, n. 5, p. 1591, 1997.

MASROUR, R. et al. Monte Carlo simulation study of magnetocaloric effect in NdMnO_3 perovskite. **Journal of Magnetism and Magnetic Materials**, v. 401, p. 91-95, 2016

MUNIZ, Francisco Tiago Leitao et al. The Scherrer equation and the dynamical theory of X-ray diffraction. **Acta Crystallographica Section A: Foundations and Advances**, v. 72, n. 3, p. 385-390, 2016.

MURPHY, Kevin P. **Machine learning: a probabilistic perspective**. MIT press, 2012.

NANTO, Dwi; YU, Seong-Cho. RELATIVE COOLING POWER OF $\text{La}_{0.7}\text{Ca}_{0.3}\text{Mn}_{1-x}\text{Cu}_x\text{O}_3$ ($0.0 \leq x \leq 0.03$). **International Journal of Technology**, v. 3, p. 417-423, 2016.

NAVROTSKY, Alexandra; WEIDNER, Donald J. Perovskite: a structure of great interest to geophysics and materials science. **Geophysical monograph series**, v. 45, 1989

NELSON, James; SANVITO, Stefano. Predicting the Curie temperature of ferromagnets using machine learning. **Physical Review Materials**, v. 3, n. 10, p. 104405, 2019

NÓBREGA, Eduardo Pilad. Efeito Magnetocalórico em Compostos com Terras-Raras: Um Estudo Monte Carlo. 1 v. Tese (Doutorado)- Curso de Física, Centro Brasileiro de Pesquisas Físicas, Rio de Janeiro, 2006.

PEARSON, Karl. VII. Mathematical contributions to the theory of evolution.—III. Regression, heredity, and panmixia. **Philosophical Transactions of the Royal Society of London. Series A, containing papers of a mathematical or physical character**, n. 187, p. 253-318, 1896.

PECHARSKY, Vitalij K.; GSCHNEIDNER JR, Karl A. Magnetocaloric effect and magnetic refrigeration. **Journal of magnetism and magnetic materials**, v. 200, n. 1-3, p. 44-56, 1999.

PEDREGOSA, Fabian et al. Scikit-learn: Machine learning in Python. **the Journal of machine Learning research**, v. 12, p. 2825-2830, 2011

RIVNAY, Jonathan et al. Quantitative analysis of lattice disorder and crystallite size in organic semiconductor thin films. **Physical Review B**, v. 84, n. 4, p. 045203, 2011.

SALAZAR-MUNOZ, V. E.; GUERRERO, A. Lobo; PALOMARES-SANCHEZ, S. A. Review of magnetocaloric properties in lanthanum manganites. **Journal of Magnetism and Magnetic Materials**, v. 562, p. 169787, 2022.

SEDGWICK, Philip. Spearman's rank correlation coefficient. **Bmj**, v. 349, 2014.

SINGH, Prashant et al. Physics-informed machine-learning prediction of curie temperatures and its promise for guiding the discovery of functional magnetic materials. **Chemistry of Materials**, v. 35, n. 16, p. 6304-6312, 2023.

SMITH, Anders et al. Materials challenges for high performance magnetocaloric refrigeration devices. **Advanced Energy Materials**, v. 2, n. 11, p. 1288-1318, 2012.

SOARES, Ana Brígida. UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE. 2005. 1v. Dissertação (Mestrado)- Curso de Ciências Naturais, Universidade Estadual do Norte Fluminense, Campos dos Goytacazes, 2005.

SUTA, Prissadang et al. An overview of machine learning in chatbots. **International Journal of Mechanical Engineering and Robotics Research**, v. 9, n. 4, p. 502-510, 2020.

SZÉKELY, Gábor J.; RIZZO, Maria L.; BAKIROV, Nail K. Measuring and testing dependence by correlation of distances. 2007.

SZYMCZAK, H.; SZYMCZAK, R. Magnetocaloric effect. Physics and applications. **Materials Science-Poland**, v. 26, n. 4, 2008

TAHIRI, N. et al. Magnetocaloric and thermoelectric properties of the perovskite LaMnO₃ material: A DFT study and Monte Carlo technique. **Phase Transitions**, v. 94, n. 11, p. 826-834, 2021.

THANH, Tran Dang et al. Magnetocaloric effect in La_{0.7}Ca_{0.25}Ba_{0.05}MnO₃ nanocrystals exhibiting the crossover of first-and second-order magnetic phase transformation. **Materials Transactions**, v. 56, n. 9, p. 1316-1319, 2015.

TLILI, R. et al. Influence of crystallite size reduction on the magnetic and magnetocaloric properties of $\text{La}_{0.6}\text{Sr}_{0.35}\text{Ca}_{0.05}\text{CoO}_3$ nanoparticles. **Polyhedron**, v. 121, p. 19-24, 2017.

TOZRI, Anowar et al. Influence of Pr-doping on magnetic phase transition and magnetocaloric effect of $\text{La}_{0.7-x}\text{Pr}_x\text{Ba}_{0.3}\text{MnO}_3$ manganite. **Materials Chemistry and Physics**, v. 149, p. 728-733, 2015.

UCAR, Huseyin; PAUDYAL, Durga; CHOUDHARY, Kamal. Machine learning predicted magnetic entropy change using chemical descriptors across a large compositional landscape. **Computational Materials Science**, v. 209, p. 111414, 2022.

VASALA, Sami; KARPPINEN, Maarit. $\text{A}_2\text{B}'\text{B}'\text{O}_6$ perovskites: a review. **Progress in solid state chemistry**, v. 43, n. 1-2, p. 1-36, 2015.

YAMADA, Hironao et al. Predicting materials properties with little data using shotgun transfer learning. **ACS central science**, v. 5, n. 10, p. 1717-1730, 2019.

YING, Xue. An overview of overfitting and its solutions. In: **Journal of physics: Conference series**. IOP Publishing, 2019. p. 022022.

ZHAI, Xiuyun; CHEN, Mingtong; LU, Wencong. Accelerated search for perovskite materials with higher Curie temperature based on the machine learning methods. **Computational Materials Science**, v. 151, p. 41-48, 2018.

ZHANG, Fengqi et al. Magnetocaloric effect in the $(\text{Mn}, \text{Fe})_2(\text{P}, \text{Si})$ system: From bulk to nano. **Acta Materialia**, v. 224, p. 117532, 2022.

ZHANG, Yun; XU, Xiaojie. Curie temperature modeling of magnetocaloric lanthanum manganites using Gaussian process regression. **Journal of Magnetism and Magnetic Materials**, v. 512, p. 166998, 2020.

Apêndice A: Descrição das features

Feature	Descrição
applied_field	Campo magnético aplicado para os modelos de RCP e entropia magnética
atomic_number	Número de prótons encontrado no núcleo de um átomo
atomic_radius	Raio atômico
atomic_radius_rahm	Raio atômico por Rahm et al
atomic_volume	Volume atômico
atomic_weight	Massa de um átomo
boiling_point	Temperatura de ebulição
bulk_modulus	Módulo em massa
c6_gb	C ₆ Coeficiente de dispersão em a.u.
covalent_radius_cordero	Raio covalente por Cordero et al
covalent_radius_pyykko	Raio covalente de ligação simples por Pyykko et al
covalent_radius_pyykko_double	Raio covalente de ligação dupla por Pyykko et al
covalent_radius_pyykko_triple	Raio covalente de ligação tripla por Pyykko et al
covalent_radius_slater	Raio covalente por Slater
density	Densidade em T = 295 K
dipole_polarizability	Polarizabilidade dipolo
electron_affinity	Eletronegatividade
electron_negativity	Tendência de atrair um par compartilhado de elétrons
en_allen	Escala de eletronegatividade de Allen
en_ghosh	Escala de eletronegatividade de Ghosh
en_pauling	Escala de eletronegatividade de Mulliken
first_ion_e	Primeira energia de ionização
fusion_enthalpy	Calor de fusão
gs_bandgap	Energia de bandgap DFT do estado fundamental T = 0 K
gs_energy	Energia DFT por átomo de T = 0 K estado fundamental
gs_est_bcc_latent	Parâmetro de rede CCC estimado
gs_est_fcc_latent	Parâmetro de rede CFC estimado
gs_mag_moment	Momento magnético DFT do estado fundamental em T = 0 K
gs_volume_per	Volume atômico DFT do estado fundamental em T = 0 K
heat_capacity_mass	Capacidade de calor específico em massa
heat_capacity_molar	Capacidade térmica específica molar
heat_of_formation	Calor de formação
hhi_p	Valores de produção do Índice Herfindahl-Hirschman
hhi_r	Índice de valor de reservas de Herfindahl-Hirschman
icsd_volume	Volume Atômico no banco de dados ICSD
lattice_constant	Dimensão física de células unitárias em uma rede cristalina
melting_point	Ponto de fusão
mendeleeev_number	Número atômico na tabela periódica de Mendeleev
molar_volume	Volume molar
num_d_unfilled	Número de elétrons não preenchidos na camada d
num_d_valence	Número de elétrons de valência na camada d
num_f_unfilled	Número de elétrons não preenchidos na camada f
num_f_valence	Número de elétrons de valência na camada f

num_p_unfilled	Número de elétrons não preenchidos na camada p
num_p_valence	Número de elétrons de valência na camada p
num_s_unfilled	Número de elétrons não preenchidos na camada s
num_s_valence	Número de elétrons de valência na camada s
num_unfilled	Total de elétrons não preenchidos
num_valence	Elétrons totais de valência
period	Período na tabela periódica
poissons_ratio	Razão de Poisson
polarizability	Capacidade de formar dipolos instantâneos
sound_velocity	Velocidade no som
specific_heat	Calor específico à 20°C
ermal_conductivity	Condutividade térmica à 25°C
vdw_radius	Raio de Van der Waals
dw_radius_alvarez	Raio de Van der Waals segundo Alvarez
vdw_radius_mm3	Raio de Van der Waals de MM3 FF
vdw_radius_uff	Raio de Van der Waals de UFF
