XVIII ENCONTRO NACIONAL DE CONFORTO NO AMBIENTE CONSTRUÍDO
XIV ENCONTRO LATINO-AMERICANO DE CONFORTO NO AMBIENTE CONSTRUÍDO
AMBIENTE CONSTRUÍDO E USUÁRIO: PERSPECTIVAS LATINO-AMERICANAS

Aplicação de *machine learning* no *benchmarking* de eficiência energética em edificações: revisão bibliométrica e sistemática.

Aplicación de machine learning en el benchmarking de eficiencia energética en edificaciones: revisión bibliométrica y sistemática

Application of machine learning in energy efficiency benchmarking of buildings: bibliometric and systematic review

Eficiência energética / *Eficiencia energética / Energy efficiency*

Tomazelli, Saile
Mestrando, Universidade Federal do Espírito Santo, Vitória, Brasil, sailetomazelli@gmail.com

Jesus Luciana Aparecida Netto de Jesus
Doutora, Universidade Federal do Espírito Santo, Vitória, Brasil, luciana.a.jesus@ufes.br

Geraldi, Matheus Soares
Doutor, Universidade Federal de Santa Catarina, Florianópolis, Brasil, matheus.geraldi@posgrad.ufsc.br

Resumo

Este estudo tem como objetivo analisar a aplicação de *machine learning* no *benchmarking* de eficiência energética em edificações. Utilizando o método ProKnow-C, a pesquisa envolveu a seleção de artigos publicados entre 2015 e 2024 resultando na análise bibliométrica de 229 publicações. Em seguida, foram selecionados 47 artigos que abordaram em detalhe etapas de *benchmarking* utilizando *machine learning*, os quais foram submetidos à análise sistemática e à avaliação de lacunas de pesquisa. A análise revelou um aumento nas publicações, especialmente a partir de 2020. A maioria das pesquisas utiliza dados reais e aplicam os algoritmos de redes neurais artificiais, máquinas de vetor de suporte e florestas aleatórias. Também foram identificadas lacunas de pesquisa, como a necessidade de criar modelos representativos dos estoques. Conclui-se que o *benchmarking* pode desempenhar um papel importante na eficiência energética de edificações, e a ampliação das bases de dados pode tornar o método mais assertivo e aplicável.

Palavras-chave: Eficiência energética. Edificações. Benchmarking. Machine learning.

Resumen

Este estudio tiene como objetivo analizar la aplicación del machine learning en el benchmarking de eficiencia energética en edificaciones. Utilizando el método ProKnow-C, la investigación incluyó la selección de artículos publicados entre 2015 y 2024, resultando en un análisis bibliométrico de 229 publicaciones. Luego, se seleccionaron 47 artículos que abordaban en detalle las etapas de benchmarking con machine learning, los cuales fueron sometidos a un análisis sistemático y a la evaluación de brechas de investigación. El análisis reveló un aumento en las publicaciones, especialmente desde 2020. La mayoría de los estudios utilizan datos reales y aplican algoritmos de redes neuronales artificiales, máquinas de soporte vectorial y bosques aleatorios. También se identificaron brechas, como la necesidad de crear modelos representativos de los edificios existentes. Se concluye que el benchmarking puede desempeñar un papel clave en la eficiencia energética y que la ampliación de bases de datos puede hacer el método más preciso y aplicable.

Palabras clave: Eficiencia energética. Edificaciones. Benchmarking. Machine learning.

Abstract

This study aims to analyze the application of machine learning in energy efficiency benchmarking for buildings. Using the ProKnow-C method, the research selected articles published between 2015 and 2024, resulting in a bibliometric analysis of 229 publications. Then, 47 articles detailing benchmarking stages using machine learning were selected and subjected to systematic analysis and research gap assessment. The analysis revealed a growth in publications, especially from 2020 onwards. Most studies use real data and apply artificial neural networks, support vector machines, and random forests. Research gaps were also identified, such as the need to develop representative building stock models. It is concluded that benchmarking can play a key role in building energy efficiency, and expanding databases can make the method more accurate and applicable.

Keywords: Energy efficiency. Buildings. Benchmarking. Machine learning.

Introdução

A energia consumida durante a fase de uso das edificações, destinada à climatização e iluminação de ambientes, aquecimento de água, cocção de alimentos, entre outros, representa 30% de todo o consumo energético global e 28% das emissões mundiais de CO_2 (United Nations Environment Programme, 2022).Essa demanda pode ser significativamente reduzida quando os edifícios são projetados e operados de forma ótima (Villano; Mauro; Pedace, 2024). Portanto, a eficiência energética deve ser uma premissa fundamental nos empreendimentos da construção civil.

O *benchmarking*, que é uma análise comparativa do desempenho energético de um edifício com outros da mesma tipologia, é uma das soluções presentes na literatura para auxiliar no aumento de eficiência das edificações já construídas, em fase de operação. Pesquisas indicam que essa abordagem pode detectar comportamentos ineficientes e aprimorar as decisões de design para a renovação energética (Yussuf; Asfour, 2024).

Na revisão de Geraldi e Ghisi (2020), os autores classificam os métodos de *benchmarking* em normalização simples (estatística simples), mínimos quadrados ordinários (regressão simples para determinar a Intensidade de Uso da Energia – IUE), análise de fronteiras estocásticas (regressão incluindo elemento geométrico), análise de envoltória de dados (regressão determinada por um limite calculado por todos os dados) e outros métodos avançados (geoestatística e aprendizado de máquina). O diferencial mencionado é que esse último método é considerado capaz de otimizar a modelagem do conjunto de edifícios construídos, ou, estoque construído.

Aprendizado de máquina é o processo de estudo de grandes volumes de dados visando identificar padrões para tomar decisões complexas com base no histórico. A técnica é realizada através de códigos de programação denominados algoritmos, que permitem realizar tarefas de classificação, regressão e agrupamento (Tien *et al.*, 2022).

Para *benchmarking* utilizando aprendizado de máquina, Geraldi e Ghisi (2022a) propõem uma metodologia que pode ser resumida em definição do estoque, criação do inventário das características das edificações, análise de entropia dos parâmetros, análise de cluster para os parâmetros variáveis, modelagem e simulação de arquétipos no *EnergyPlus*, desenvolvimento do modelo de rede neural artificial e comparação do resultado real da edificação com o estoque.

Os arquétipos são simplificações de um estoque, onde as características geométricas e não geométricas utilizadas são as médias ou padrões observados nas edificações (Ali *et al.*, 2019). Portanto, quando não existe banco de dados reais das edificações, é possível definir um *benchmarking* por meio dessa técnica, utilizando um banco de dados simulado a partir de características conhecidas do estoque.

Diante do contexto apresentado, o presente artigo tem como objetivo mapear e analisar a aplicação de *machine learning* no *benchmarking* energético em edificações. O estudo busca identificar os países que realizam pesquisas sobre o tema e explorar a origem do dado de entrada, os algoritmos empregados e as lacunas existentes na pesquisa sobre este assunto.

Método

O método empregado neste estudo foi o ProKnow-C, desenvolvido pelo Laboratório MCDA da Universidade Federal de Santa Catarina (Ensslin, et al., 2010), que consiste em 4 etapas: seleção de portfólio bibliográfico; análise bibliométrica; análise sistemática; e pergunta da pesquisa, objetivo geral e objetivos específicos.

Na Etapa 1 (seleção do portfólio bibliográfico), definiu-se o tema de pesquisa e selecionaram-se as bases de dados que apresentaram maior quantidade de artigos relevantes, especificamente Scopus e Web of Science. Optou-se por utilizar apenas artigos em inglês, e as palavras-chave foram definidas, traduzidas e expandidas. O termo final de busca utilizado foi: *benchmark* AND (*algorithm* OR "*machine learning*" OR *model*) AND "*energy efficiency*" AND (*building* OR *construction*).

Estabeleceu-se que os artigos deveriam ser publicados entre 2015 e 2024 para garantir informações atuais, e que as palavras-chave deveriam estar no título, resumo ou palavras-chave dos artigos, para assegurar a relevância do tema. Após a seleção inicial dos artigos, eles foram agrupados das duas bases pelo gerenciador de artigos Mendeley, e as duplicidades foram removidas.

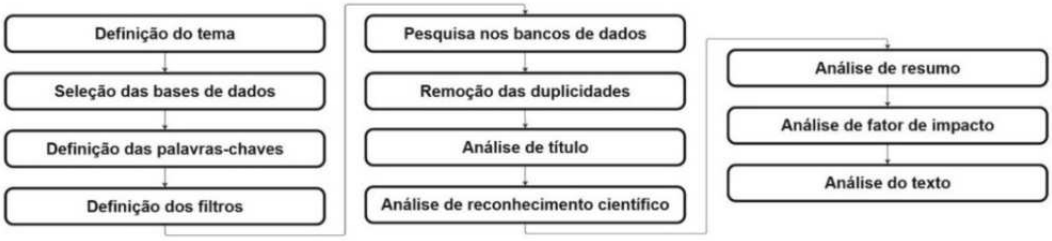
Posteriormente, avaliaram-se os títulos e as citações dos artigos, sendo selecionados aqueles com 21 ou mais citações. Incluíram-se também publicações com menos de dois anos, por ainda não

terem sido amplamente explorados pela comunidade acadêmica. Também foram considerados estudos com mais de 2 anos de autores com outro artigo com o número mínimo de 21 citações.

Em seguida, foram selecionados apenas os artigos cujo resumo mencionava *benchmarking* de eficiência energética de edificações com uso de algoritmos de aprendizado de máquina ou etapas do processo, como dados utilizados, seleção de parâmetros e comparação de desempenho.

Na fase de análise do fator de impacto, foram selecionados apenas artigos classificados como Q1 ou Q2 pelo *Scientific Journal Rankings*. Por fim, a análise do texto filtra apenas os artigos que possuem conteúdo sobre as etapas do *benchmarking* de eficiência energética em edificações utilizando técnicas de aprendizado de máquina. A Figura 1 apresenta o fluxograma do processo de seleção dos artigos.

Figura 1: Processo de seleção do portfólio.



Fonte: O autor.

Na Etapa 2 (análise bibliométrica), foi realizada uma análise quantitativa dos artigos resultantes da etapa da análise de título quanto ao ano de publicação, país do primeiro autor, periódico e autor.

Na Etapa 3 (análise sistemática), foram considerados os artigos após a análise do texto, verificando-se os tipos de dados utilizado na base de entrada dos modelos de benchmarking (real do estoque ou arquétipos simulados), os algoritmos de aprendizado de máquina utilizados (redes neurais artificiais, máquina de vetor de suporte, árvore de decisão etc.) e a tipologia de edificações que foram estudadas (residencial, comercial, educacional etc.).

Finalmente, na Etapa 4 (pergunta da pesquisa, objetivo geral e objetivos específicos), foram identificadas as lacunas de pesquisa dentro do tema em estudo.

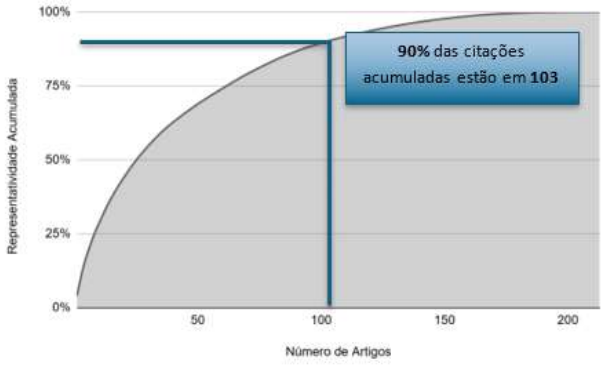
Resultados e discussões

Após a definição inicial do portfólio (tema, base de dados, palavras-chave e filtros), foi elaborado o termo de busca que resultou em 628 artigos: 276 provenientes da Scopus e 352 da Web of Science. Após a remoção de duplicidades, o total foi de 512 artigos.

Após a análise dos títulos, permaneceram apenas 229 artigos, uma vez que muitos estavam relacionados à *benchmarking* da construção de algoritmos eficientes e de sistemas isolados da edificação, como ar-condicionado, iluminação e data centers.

Na etapa de análise do reconhecimento científico, avaliou-se a quantidade de citações dos artigos do portfólio com base no Google Acadêmico. A Figura 2 apresenta o percentual acumulado de citações, a partir do qual foram selecionados 103 artigos com 21 ou mais citações, representando 90% do total. Adicionalmente, incluíram-se 52 artigos com menos de dois anos que não atendiam ao critério mínimo de citações e 7 artigos de autores com outros trabalhos com esse mínimo, totalizando 162 artigos.

Figura 2: Curva ABC – filtro do banco de artigos quanto ao reconhecimento científico.



Fonte: O autor.

Após a análise dos resumos, foram selecionados 85 artigos, visto que os demais eram focados em gerenciamento de energia, previsão de carga ou não utilizaram aprendizado de máquina.

Finalmente, foram selecionados 78 artigos Q1 e Q2, evidenciando que a base filtrada possuía poucos artigos de periódico classificado como Q3 ou Q4, e após a leitura dos textos completos

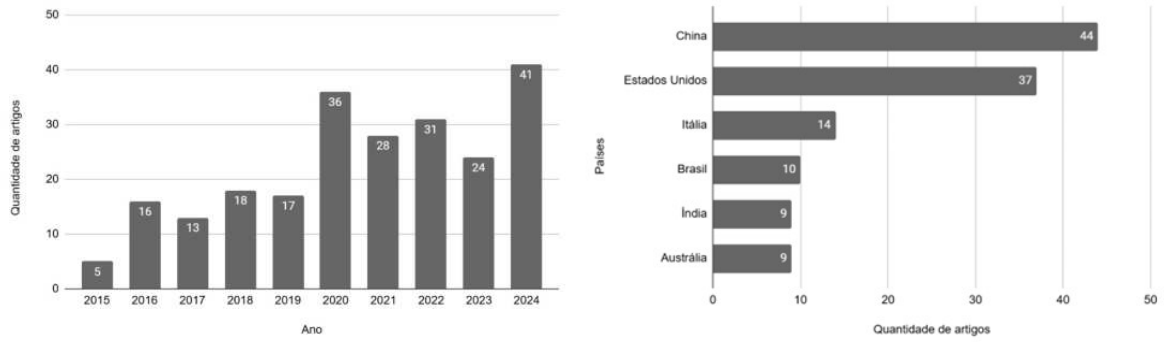
restaram 47 artigos, pois aqueles que não utilizavam aprendizado de máquina e não indicavam essa informação claramente no resumo foram excluídos.

Análise bibliométrica

A análise bibliométrica foi conduzida com base nos 229 artigos selecionados pelo termo de busca, após a remoção das duplicidades e seleção dos títulos.

Os artigos publicados nos últimos cinco anos representam 70% do total, indicando um aumento significativo na relevância do tema. Quanto à distribuição geográfica da produção científica, foram identificados 46 países realizando pesquisas sobre o tema. Os seis países com maior quantidade de publicações representam 54% do total. O Brasil, por sua vez, possui dez artigos publicados. Esses resultados são apresentados na Figura 3.

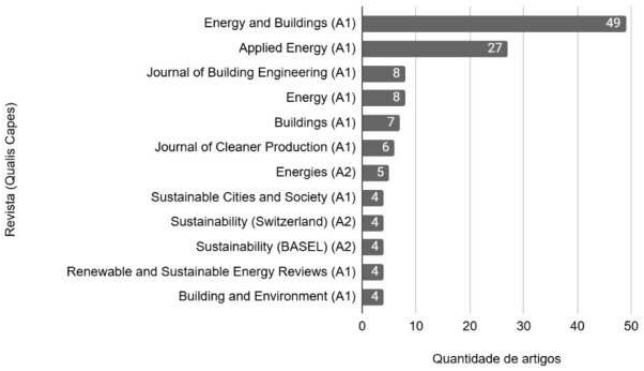
Figura 3: Quantidade de artigos por ano e top 6 países com maior número de artigos – Análise bibliométrica.



Fonte: O autor.

Foram identificadas 89 revistas diferentes, das quais as 12 que mais publicaram sobre o tema, apresentadas na Figura 4, representam 57% dos artigos. Isso indica uma concentração das publicações nesses periódicos. Além disso, a maioria dessas 12 revistas é classificada como A1 pelo Qualis Capes (critério vigente até 2024), de acordo com os dados da plataforma Sucupira para o quadriênio 2017-2020.

Figura 4: Top 12 revistas com maior quantidade de artigos.



Fonte: O autor.

Análise sistêmica

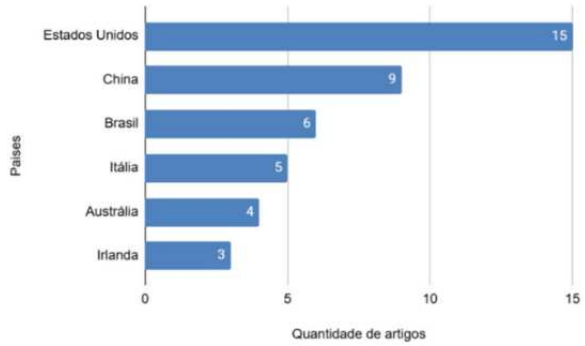
A revisão sistemática foi realizada com base nos 47 artigos selecionados após todas as etapas de seleção do portfólio. Ao analisar o tipo de dado utilizado para estabelecer a base de entrada dos estudos foi identificado que os estudos utilizando tanto dados reais do estoque quanto dados obtidos por arquétipos e simulação computacional.

Em resumo, para utilizar dados reais, é necessário que exista um banco que reúna todas as informações das edificações a serem estudadas, como arquitetura, localização geográfica, consumo de energia, em quantidade suficiente para treinar e testar os modelos de *machine learning*. A literatura aponta que obter uma base consistente para esse tipo de análise é desafiador e requer um esforço amplo de diversas partes, visto que o consumo de energia e os dados das edificações são tratados como dados sensíveis e são protegidos pela Lei de Proteção de Dados, tanto do Brasil quanto de outros países.

Quando isso não existe, é preciso desenvolver arquétipos representativos do estoque construído, ou seja, simplificar padrões e utilizar parâmetros médios das edificações, e simulá-los em diferentes cenários para criar um banco de dados representativo e possibilitar a aplicação do método. A principal vantagem é a redução do esforço de modelagem, mas, em contrapartida, eles podem não representar com precisão todo o universo de estudo, afetando a confiabilidade do modelo de *benchmark* (Ali *et al.*, 2019).

A maioria dos estudos utiliza dados reais do estoque (80,6%), o que só é possível realizar com maior facilidade em países que possuem bancos de dados sobre os estoques, como é o caso dos Estados Unidos, China e Itália, que são os países que mais pesquisam sobre o tema, conforme Figura 5.

Figura 5: Top 6 países com maior quantidade de artigos – Análise sistemática.



Fonte: O autor.

Já países que estão no processo de desenvolver seus bancos de dados, como o Brasil, majoritariamente são utilizados os dados de arquétipos para criar os modelos de *benchmarking*.

Quanto aos algoritmos de aprendizado de máquina, foram identificados 55 termos diferentes em 35 que citaram o algoritmo utilizado, uma média de quase dois modelos por artigo. Isso ocorre porque 60% da amostra utiliza dois ou mais tipos de código de programação para otimizar processos computacionais ou realizar análises comparativas de desempenho, utilizando os mesmos dados e premissas no treinamento, para determinar a melhor opção a ser utilizada.

Os algoritmos mais utilizados são redes neurais artificiais (RNA), máquinas de vetor de suporte (MVS) e florestas aleatórias (FA), presentes em 23%, 17% e 17% dos artigos, respectivamente.

As RNA são métodos de aprendizado estatístico não lineares inspirados no funcionamento do cérebro humano. Estruturalmente, é composta por três camadas principais — entrada, oculta (pode haver mais de uma) e saída — formadas por neurônios interconectados que utilizam funções de ativação para processar informações. A principal vantagem do método é a habilidade de identificar, de forma implícita, relações não lineares complexas entre variáveis de entrada e

saída. Porém, uma limitação é a incapacidade de representar diretamente as interações entre os parâmetros físicos do edifício e seu consumo energético, dificultando a adaptação do modelo a modificações nos sistemas ou componentes da edificação (Wang; Srinivasan, 2017).

As MVS são modelos utilizados para resolver problemas não lineares de regressão e classificação. Baseiam-se no princípio da Minimização do Risco Estrutural, buscando reduzir o erro de generalização ao equilibrar erro de treinamento e complexidade do modelo. Uma de suas principais vantagens é a capacidade de operar com dados reduzidos, sendo eficazes mesmo sem registros históricos. No entanto, uma limitação significativa é o alto custo computacional, que cresce aproximadamente com o cubo do número de amostras (Seyedzadeh *et al.*, 2018).

Já a FA aprimora o método de ensacamento ao introduzir aleatoriedade na seleção de amostras e preditores durante a construção de múltiplas árvores de decisão. Cada árvore é gerada a partir de uma amostra *bootstrap* diferente (ou seja, alguns dados do conjunto original podem se repetir, enquanto outros podem não ser incluídos), e os nós são divididos com base em subconjuntos aleatórios de variáveis, o que reduz o viés e torna o modelo menos suscetível ao sobreajuste. É considerada uma das técnicas de aprendizado estatístico mais precisas, porém é frequentemente criticada por ser uma "caixa preta", dificultando a interpretação dos resultados e a comunicação com públicos não técnicos (Kontokosta; Tull, 2017).

Além dos algoritmos, foram analisadas as tipologias das edificações mais estudadas na área. As edificações comerciais, residenciais e de escritório representaram respectivamente 22%, 21% e 14,5% dos estudos de caso. Além disso, 13% dos artigos não especificaram a tipologia por se tratar de revisão de literatura de temas aplicáveis a qualquer edifício.

Lacunas identificadas

Entre as principais lacunas identificadas sobre o tema está a criação de modelos de *benchmarking* que sejam representativos nacionalmente. O modelo criado por Arjunan, Poolla e Miller, por exemplo, apresenta essa limitação (Arjunan; Poolla; Miller, 2022). De forma similar, Quevedo, Geraldi e Melo citam que o modelo criado para universidades do Brasil não é representativo para instituições públicas, visto que o arquétipo utilizado foi baseado em instituições privadas (Quevedo; Geraldi; Melo, 2023). Além disso, Geraldi e Ghisi (2022) mencionam existir

subjetividade na modelagem geométrica e a desconsideração das condições de arredores das edificações nas modelagens, podem dificultar o estabelecimento de padrões realistas desse fator para o estoque.

Também se observou como lacuna a limitação da base de dados disponível para utilização nos modelos. O estudo de Ding e Liu cita a necessidade de expandir o tamanho da amostra para testar a consistência dos resultados obtidos (Ding; Liu, 2020). Já Geraldi e Ghisi citam que a base de dados precisa atualizar de acordo com a evolução do estoque (Geraldi; Ghisi, 2022b).

Conclusões

Este artigo evidenciou uma crescente relevância da aplicação de *machine learning* no *benchmarking* de eficiência energética em edificações, conforme observado o crescimento das publicações nos últimos anos.

A análise revelou que os estudos se concentram em países que possuem bancos de dados robustos sobre o estoque edificado, como Estados Unidos, China e Itália. Em contrapartida, nações em processo de desenvolvimento desses bancos, como o Brasil, recorrem majoritariamente a arquétipos simulados, que embora reduzam o esforço de modelagem, podem comprometer a precisão dos resultados.

No que diz respeito aos algoritmos, destacaram-se as RNA, as MVS e as FA, cada uma com vantagens específicas e limitações que afetam a aplicabilidade e interpretabilidade dos modelos. A escolha por múltiplos algoritmos em um mesmo estudo, prática recorrente na amostra, reforça a busca por otimização e confiabilidade dos resultados.

Além disso, a pesquisa identificou lacunas significativas, como a necessidade de modelos representativos em nível nacional, a subjetividade na modelagem geométrica das edificações e a limitação de bases de dados para treinar os modelos.

Dessa forma, futuras pesquisas devem focar na ampliação e aprimoramento das bases de dados utilizadas, na criação de modelos que representem melhor a diversidade do estoque edificado e na incorporação de fatores contextuais, como as condições dos arredores das edificações. Esses

avanços contribuirão para o desenvolvimento de ferramentas mais precisas e aplicáveis à realidade dos países, promovendo maior eficiência energética no setor da construção.

Referências

ALI, U. *et al.* A data-driven approach for multi-scale building archetypes development. **ENERGY AND BUILDINGS**, [*s. l.*], v. 202, 2019.

ARJUNAN, P.; POOLLA, K.; MILLER, C. BEEM: Data-driven building energy benchmarking for Singapore. **ENERGY AND BUILDINGS**, [*s. l.*], v. 260, 2022.

DING, Y.; LIU, X. A comparative analysis of data-driven methods in building energy benchmarking. **ENERGY AND BUILDINGS**, [*s. l.*], v. 209, 2020.

ENSSLIN, L et al. ProKnow-C: Processo de análise sistêmica. Brasil: Processo técnico com patente de registro pendente junto ao INPI, 2010.

GERALDI, M. S.; GHISI, E. Building-level and stock-level in contrast: A literature review of the energy performance of buildings during the operational stage. **ENERGY AND BUILDINGS**, [*s. l.*], v. 211, 2020.

GERALDI, M. S.; GHISI, E. Data-driven framework towards realistic bottom-up energy benchmarking using an Artificial Neural Network. **APPLIED ENERGY**, [*s. l.*], v. 306, 2022.

GERALDI, M. S.; GHISI, E. Integrating evidence-based thermal satisfaction in energy benchmarking: A data-driven approach for a whole-building evaluation. **Energy**, [*s. l.*], v. 244, 2022.

KONTOKOSTA, C. E.; TULL, C. A data-driven predictive model of city-scale energy use in buildings. **APPLIED ENERGY**, [*s. l.*], v. 197, p. 303–317, 2017.

QUEVEDO, T. C.; GERALDI, M. S.; MELO, A. P. Applying machine learning to develop energy benchmarking for university buildings in Brazil. **JOURNAL OF BUILDING ENGINEERING**, [*s. l.*], v. 63, 2023.

SEYEDZADEH, S. *et al.* Machine learning for estimation of building energy consumption and performance: a review. **Visualization in Engineering**, [*s. l.*], v. 6, n. 1, 2018.

TIEN, P. W. *et al.* Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality – A Review. **Energy and AI**, [*s. l.*], v. 10, 2022.

UNITED NATIONS ENVIRONMENT PROGRAMME. **2022 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector**. [*S. l.: s. n.*], 2022.

VILLANO, F.; MAURO, G. M.; PEDACE, A. A Review on Machine/Deep Learning Techniques Applied to Building Energy Simulation, Optimization and Management. **Thermo**, [*s. l.*], v. 4, n. 1, p. 100–139, 2024.

WANG, Z.; SRINIVASAN, R. S. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. **Renewable and Sustainable Energy Reviews**, [*s. l.*], v. 75, p. 796–808, 2017.

YUSSUF, R. O.; ASFOUR, O. S. Applications of artificial intelligence for energy efficiency throughout the building lifecycle: An overview. **Energy and Buildings**, [*s. l.*], v. 305, 2024.