

Dự đoán và đánh giá hiệu suất mặt dựng tích hợp quang điện bằng các mô hình trí tuệ nhân tạo

Trần Đức Học¹, Hoàng Hải Triều¹, Nguyễn Ngọc Thoan², Hồ Ngọc Khoa^{2,*}

¹Khoa Kỹ Thuật Xây dựng, Trường Đại học Bách khoa – Đại học Quốc gia TP. Hồ Chí Minh

²Khoa Xây dựng Dân dụng và Công nghiệp, Trường Đại học Xây dựng Hà Nội

TỪ KHOÁ

BIPV
Tiết kiệm năng lượng
Dự đoán
Trí tuệ nhân tạo

TÓM TẮT

Sự gia tăng tiêu thụ năng lượng toàn cầu do quá trình phát triển công nghiệp và đô thị hóa đang góp phần làm trầm trọng thêm tình trạng biến đổi khí hậu. Trong bối cảnh đó, năng lượng tái tạo, đặc biệt là năng lượng mặt trời, đóng vai trò then chốt trong việc hướng tới mục tiêu xây dựng xanh và bền vững. Nghiên cứu này đề xuất giải pháp ứng dụng các thuật toán trí tuệ nhân tạo gồm tăng cường độ dốc (GB), tăng cường độ dốc cực hạn (XGBoost), tăng cường độ dốc nhẹ (LightGBM) và tăng cường phân loại (CatBoost) để xây dựng mô hình dự đoán mức tiêu thụ năng lượng và sản lượng điện của hệ thống quang điện tích hợp trong tòa nhà (Building-Integrated Photovoltaics – BIPV) với lớp vỏ bao bằng kính. Dữ liệu đầu vào được thu thập thông qua mô phỏng trên phần mềm DesignBuilder. Kết quả cho thấy mô hình CatBoost đạt hiệu suất dự đoán cao nhất với hệ số xác định R^2 đạt 0,972 cho tiêu thụ năng lượng và 0,974 cho sản lượng điện. Ngoài ra, việc tích hợp BIPV vào mặt đứng có thể giúp giảm đến 34,3 % mức tiêu thụ năng lượng của tòa nhà, qua đó khẳng định tiềm năng ứng dụng của BIPV trong thiết kế công trình hiệu quả năng lượng.

KEYWORDS

BIPV
Energy saving
Prediction
Artificial intelligence

ABSTRACT

The continuous growth of global energy consumption due to industrialization and urban development has significantly contributed to climate change. In response, renewable energy, particularly solar energy, has emerged as a key solution toward green and sustainable building design. This study proposes the application of artificial intelligence algorithms, including Gradient boosting (GB), Extreme gradient boosting (XGBoost), Light gradient boosting machine (LightGBM), and Categorical boosting (CatBoost), to develop predictive models for building energy consumption and photovoltaic electricity generation in office buildings equipped with glass-based building-integrated photovoltaics (BIPV) facades. Input data were generated through simulation using DesignBuilder software. Among the models, CatBoost yielded the best performance, with R^2 values of 0.972 for energy consumption and 0.974 for electricity generation. Furthermore, the integration of BIPV systems into building facades has the potential to reduce energy consumption by up to 34.3%, highlighting its practical benefits for energy-efficient building design.

1. Giới thiệu

Năng lượng đóng vai trò cốt lõi trong sự tiến bộ và phát triển của toàn nhân loại cũng như sự tồn tại của con người. Trước bối cảnh tình hình sử dụng năng lượng cũng như lượng phát thải CO_2 toàn cầu liên tục tăng khiến nhiệt độ trung bình toàn cầu tăng 1,5 °C trong những năm gần đây [1]. Điều này đã đặt ra nhiều thách thức lớn trong việc cân bằng giữa phát triển kinh tế và bảo vệ môi trường, đồng thời cần đẩy mạnh quá trình chuyển dịch từ sử dụng các nguồn năng lượng hóa thạch sang các nguồn năng lượng sạch. Ở Việt Nam, năng lượng mặt trời là một trong những nguồn năng lượng sạch - năng lượng tái tạo mang nhiều tiềm năng phát triển trong tương lai. Trong đó việc sử dụng các phương pháp tích hợp nguồn năng lượng mặt trời vào trong các tòa nhà được xem là giải pháp hiệu quả giúp giảm phát thải CO_2 ,

hướng tới xây dựng bền vững [2].

Trước đây, các hệ thống pin năng lượng mặt trời được tích hợp rộng rãi nhưng diện tích mái có giới hạn, trong khi đó các tòa nhà trung tầng hoặc cao tầng có diện tích bề mặt lớp vỏ bao che lớn hơn rất nhiều so với diện tích mái. Đặc biệt hơn, các tòa nhà hiện đại thường sử dụng lớp vỏ bao tòa nhà bằng kính không chỉ nâng cao tính thẩm mỹ kiến trúc mà còn tận dụng hiệu quả nguồn ánh sáng tự nhiên [3]. Từ đó có một xu hướng mới đầy tiềm năng trong lĩnh vực xây dựng bền vững là tích hợp các hệ thống pin năng lượng mặt trời vào các lớp vỏ bằng kính của tòa nhà được gọi là BIPV (Building Integrated Photovoltaic). BIPV không chỉ giữ nguyên tính thẩm mỹ cho tòa nhà mà còn trở thành một nguồn năng lượng tái tạo dồi dào nhờ tận dụng bức xạ mặt trời để chuyển hóa thành điện năng cung cấp cho tòa nhà.

Perez, et al. [4] đánh giá hiệu quả sau 5 năm vận hành hệ thống

*Liên hệ tác giả: khoahn@huce.edu.vn

Nhận ngày 03/09/2025, sửa xong ngày 06/10/2025, chấp nhận đăng ngày 07/10/2025

Link DOI: <https://doi.org/10.54772/jomc.05.2025.1143>

BIPV của một tòa nhà ở New York với sản lượng cao nhất đạt 89 kWh/m² và mức phát thải CO₂ thấp chỉ 10,2gCO₂/kWh vào năm 2012. Ghazali, et al. [5] đánh giá hiệu quả hệ thống BIPV tích hợp mặt đứng tòa nhà sau 3 năm vận hành với sản lượng cao nhất 600 kWh/kW/năm và 94 kWh/m² ở vùng khí hậu nhiệt đới tại Kuala Lumpur, Malaysia vào năm 2017. Gholami, et al. [6] phân tích hiệu quả sau 4 năm vận hành hệ thống BIPV với sản lượng trung bình 40 kWh/m² với tuổi thọ sử dụng 30 năm ở vùng ôn đới lạnh vào năm 2020. Nguyen, et al. [7] đã xây dựng mô hình đánh giá hiệu quả của hệ thống Solar Cell tích hợp vào lớp vỏ kính của tòa nhà ở Đà Nẵng, Việt Nam với kết quả lượng điện sản sinh tối đa lên tới 493 922 kWh/năm. Từ lợi ích đó đã cho thấy việc tích hợp hệ thống BIPV vào các tòa nhà sẽ trở thành một hướng đi quan trọng và đầy tiềm năng trong thiết kế công trình bền vững ở tương lai trên thế giới nói chung và Việt Nam nói riêng.

Tuy nhiên để ước tính và đánh giá hiệu quả năng lượng cũng như sản lượng điện từ phương án tích hợp BIPV thường đòi hỏi mô phỏng phức tạp thông qua các phần mềm chuyên dụng như EnergyPlus, DesignBuilder... và những mô phỏng này cần rất nhiều tài nguyên, đặc biệt khi cần khảo sát nhiều phương án thiết kế với nhiều thông số tổ hợp để có được kết quả với độ chính xác cao [8]. Trước những khó khăn đó, học máy đã được ứng dụng rộng rãi trong lĩnh vực kỹ thuật để xây dựng các mô hình dự đoán, thay thế hoặc hỗ trợ cho các mô hình truyền thống. Trong mô phỏng năng lượng, học máy có thể tự học từ dữ liệu mô phỏng trước đó để dự đoán các kết quả nhanh chóng, đầy nhanh quá trình ra quyết định. Các mô hình học máy cho phép rút ngắn thời gian đánh giá hàng trăm phương án thiết kế khác nhau, giảm đáng kể tài nguyên mà vẫn duy trì độ chính xác cao. Từ đó, sử dụng học máy để đánh giá hiệu quả năng lượng là một giải pháp tiềm năng giúp các kỹ sư có thể đánh giá sơ bộ các giải pháp thiết kế trong giai đoạn đầu của dự án.

Trên thế giới đã có nhiều nghiên cứu về việc tích hợp hệ thống BIPV vào tòa nhà. Kabilan, et al. [9] đã xây dựng mô hình dự đoán năng lượng tiêu thụ của tòa nhà và kết quả tốt nhất là thuật toán mạng nơ ron nhân tạo (ANN-Artificial neural network) vào năm 2021. Imalka, et al. [10] mô phỏng năng lượng tòa nhà tích hợp hệ thống BIPV bằng Grasshopper và dự đoán năng lượng bằng ANN với độ chính xác cao và giảm 20% thời gian tính toán vào năm 2024. Học [11] và cộng sự dự đoán mức tiêu thụ năng lượng của các căn hộ chung cư tại Việt Nam thông qua các thuật toán trí tuệ nhân tạo với hiệu quả tốt nhất là mô hình đơn cây phân loại và hồi quy (CART- Classification and regression tree) vào năm 2020. Long [12] đã xây dựng mô hình dự đoán dự mức năng lượng tiêu thụ dựa trên các thuật toán học máy và thuật toán tăng cường độ dốc (GB - Gradient boosting) đạt hiệu quả tốt nhất với R²=0,994.

Mặc dù các kết quả nghiên cứu đều mang đến kết quả hiệu quả cao, tuy nhiên các nghiên cứu này có một điểm chung là tận dụng được các thuật toán học máy hiện đại trong những năm gần đây. Ngoài ra, mục tiêu dự đoán đơn lẻ tập trung vào mức tiêu thụ năng lượng, các kịch bản xây dựng còn hạn chế về các tham số mô phỏng. Bên cạnh đó, tại Việt Nam – nơi có khí hậu nhiệt đới đặc thù, nhận lượng bức xạ mặt

trời lớn đặc biệt ở miền Nam Việt Nam, với xu hướng sử dụng kính cho mặt đứng tòa nhà ngày càng tăng - cũng còn rất ít nghiên cứu triển khai mô hình học máy đối với các dữ liệu có các tham số đặc trưng cho hệ lớp vỏ bao che tòa nhà bằng kính tích hợp BIPV. Do đó, nghiên cứu được thực hiện với mục tiêu xây dựng mô hình học máy hiện đại, hiệu quả để dự đoán đồng thời năng lượng tiêu thụ và sản lượng từ hệ thống BIPV thông qua bộ dữ liệu được thu thập thông qua mô phỏng bằng phần mềm DesignBuilder. Các thuật toán được sử dụng để so sánh bao gồm tăng cường độ dốc (GB) và các biến thể là tăng cường độ dốc cực hạn (XGBoost - eXtreme gradient boosting), tăng cường độ dốc nhẹ (LightGBM - Light gradient-boosting machine) và tăng cường phân loại (CatBoost - Categorical boosting). Kết quả so sánh nhằm tìm ra thuật toán có độ chính xác và độ tổng quát hóa cao, tạo tiền đề cho các nghiên cứu tối ưu hóa đa mục tiêu hay hỗ trợ thiết kế công trình xanh tại Việt Nam trong tương lai.

2. Mô hình các thuật toán học máy và phương pháp đánh giá

2.1. Thuật toán tăng cường độ dốc (GB)

Tăng cường độ dốc là một thuật toán học máy thuộc nhóm học tổ hợp, sử dụng hiệu quả với bài toán phân loại và bài toán hồi quy do Jerome H.Friedman phát triển vào năm 1999, công bố năm 2001 [13]. GB dựa trên hai kỹ thuật Boosting và Gradient descent. Boosting là kỹ thuật dùng để chi phương pháp học tuần tự bằng cách kết hợp các mô hình yếu cùng loại để tạo ra mô hình mới tốt hơn thông qua sự thay đổi của pseudo-residual – hay còn gọi là “phần dư”. “Phần dư” của các dữ liệu dự đoán đúng sẽ được giữ nguyên và “phần dư” của các dữ liệu dự đoán sai sẽ được trả về kết quả tốt hơn ở các mô hình sau. Gradient descent là kỹ thuật tính toán để điều chỉnh xu hướng dự đoán “phần dư” giảm xuống thông qua đạo hàm của hàm mất mát. Quá trình được lặp đi lặp lại khi mỗi mô hình nhỏ được thêm vào để tinh chỉnh giá trị dự đoán đến khi đạt kết quả mong muốn. Đối với các thuật toán sử dụng mô hình cơ bản là cây quyết định (decision tree), GB đã được Friedman [14] cải tiến bằng cách thêm hệ số học (learning rate) để kiểm soát tốc độ của mô hình. Khi đó phương trình tổng quát của thuật toán được viết thành:

$$F_m(x) = F_{m-1}(x) + v \cdot \gamma_{lm} 1(x \in R_{lm}) \quad (1)$$

$$\gamma_{lm} = \arg \min_{\gamma} \sum_{x_i \in R_{lm}} L(y_i, F_{m-1}(x_i) + \gamma) \quad (2)$$

trong đó $F_m(x)$ là mô hình tổng hợp từ mô hình trước $F_{m-1}(x)$;

hệ số tối ưu γ_{lm} được tính toán thông qua tối ưu hàm mất mát L ; v là hệ số học.

2.2. Thuật toán tăng cường độ dốc cực hạn (XGBoost)

Thuật toán tăng cường độ dốc cực hạn (XGBoost) là một giải thuật mới dựa trên kỹ thuật Gradient Boosting nhằm tăng cường hiệu suất và tốc độ, được phát triển bởi Chen and Guestrin [15] tại trường Đại học Washington. Tương tự như thuật toán GB, XGBoost sử dụng cây quyết định là mô hình cơ bản, mô hình sau khắc phục lỗi cho mô

hình trước đó với mục tiêu tối ưu hóa hàm mất mát. XGBoost cải tiến hiệu suất bằng cách áp dụng các kỹ thuật như điều chuẩn (regularization) để giảm quá khớp (overfitting), học song song (parallelization) để tăng tốc độ và tối ưu hóa bộ nhớ nhằm xử lý hiệu quả dữ liệu lớn. Các bước thực hiện giải thuật XGBoost tương tự như GB, tuy nhiên hàm mất mát được áp dụng kỹ thuật điều chuẩn, thêm “hàm phạt” vào hàm mục tiêu giúp mô hình giảm hiện tượng quá khớp. Khi đó hàm mất mát trở thành:

$$L_{XGBoost} = \sum_{i=1}^N L[y_i, F(x_i)] + \sum_{m=1}^M \Omega(h_m) \quad (3)$$

với hàm phạt là:

$$\Omega(h) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (4)$$

trong đó T là số lượng lá của mỗi cây quyết định; w là giá trị dự đoán của các đặc trưng; γ là hệ số phạt theo số lượng lá; λ là hệ số điều chỉnh độ lớn của trọng số tránh trọng số quá lớn khiến mô hình dự đoán hội tụ nhanh. Nếu số lượng lá càng nhiều, mô hình càng phức tạp, hệ số γ càng cao và hàm bị “phạt” nhiều hơn; từ đó thuật toán sẽ ưu tiên các cây đơn giản hơn, giảm nguy cơ quá khớp.

2.3. Thuật toán tăng cường độ dốc nhẹ (LightGBM)

Máy tăng cường độ dốc nhẹ - LightGBM là một thuật toán học máy thuộc họ Cây quyết định tăng cường độ dốc được phát triển bởi Microsoft và công bố lần đầu tiên vào năm 2017 [16]. LightGBM được thiết kế để xử lý các tập dữ liệu lớn và các không gian đặc trưng có số chiều cao, mang lại hiệu suất vượt trội về tốc độ tính toán và độ chính xác. Hai kỹ thuật chính được áp dụng trong LightGBM là kỹ thuật lấy mẫu một phía (GOSS) và gom nhóm đặc trưng (EFB).

GROSS giữ lại các mẫu có sai số dự đoán lớn, đóng vai trò quan trọng trong quá trình huấn luyện và chỉ lấy một phần nhỏ các mẫu có sai số dự đoán nhỏ. Sau đó, hệ số chuẩn hóa sẽ được áp dụng cho nhóm mẫu có sai số dự đoán nhỏ nhằm cân bằng lại phân phối dữ liệu. Kỹ thuật này giúp giảm đáng kể số lượng mẫu cần xử lý mà vẫn đảm bảo độ chính xác của mô hình.

EFB giải quyết vấn đề dữ liệu có nhiều đặc trưng thừa thớt. Các đặc trưng không hoạt động đồng thời được gom nhóm lại thành một đặc trưng duy nhất – gọi là bundle, nhờ đó làm giảm chiều không gian dữ liệu từ $\#feature$ xuống $\#bundle$ (với $\#bundle < \#feature$). Điều này giúp giảm độ phức tạp tính toán từ $O(\#data \times \#feature)$ xuống $O(\#data \times \#bundle)$, đồng thời vẫn duy trì hiệu quả phân loại. LightGBM sử dụng thuật toán tham lam (greedy algorithm) để tìm ra cách gom nhóm tối ưu dựa trên phân phối tần suất.

2.4. Thuật toán tăng cường phân loại (CatBoost)

Máy tăng cường phân loại - CatBoost cũng là một thuật toán cải tiến từ GB được phát triển bởi Yandex và được đề xuất bởi nhóm nghiên cứu của Prokhorenkova, et al. [17] vào năm 2018. CatBoost được xây dựng nhằm xử lý hiệu quả các đặc trưng phân loại và giảm hiện tượng lệch dự đoán (prediction shift) trong quá trình huấn luyện.

Để mã hoá các đặc trưng phân loại, Catboost sử dụng kỹ thuật thống kê mục tiêu (target encoding) dựa trên giá trị của đặc trưng, kết hợp với giá trị ưu tiên (prior) và trọng số để tính toán giá trị đặc trưng mới. Trong quá trình tính toán đặc trưng cho mẫu i , CatBoost sử dụng các mẫu trước đó (tức $< i$) để thực hiện một hoán vị ngẫu nhiên và tính toán trên các mẫu hoán vị này. Ý tưởng này đảm bảo tất cả các dữ liệu đều tham gia quá trình huấn luyện mà không sợ bị quá khớp. Quá trình được lặp lại nhiều lần với hoán vị và trung bình kết quả. Công thức tính giá trị đặc trưng phân loại được thể hiện như sau:

$$\hat{x}_k^i = \frac{\sum_{j=1}^n 1_{\{x_j^i = x_k^i\}} y_j + a^P}{\sum_{j=1}^n 1_{\{x_j^i = x_k^i\}} + a} \quad (5)$$

trong đó y_i là giá trị mục tiêu của mẫu i ; P là giá trị ưu tiên (thường là giá trị trung bình của toàn mục tiêu trên toàn tập dữ liệu); a là trọng số làm mượt của P .

2.5. Đánh giá hiệu suất mô hình

Hiệu suất của mô hình dự đoán được đánh giá thông qua các chỉ số: hệ số xác định - R^2 , sai số phần trăm tuyệt đối trung bình – MAPE (Mean absolute percentage error), sai số tuyệt đối trung bình – MSE (Mean squared error) và căn bậc hai của sai số bình phương – RMSE (Root mean squared error).

Cụ thể, hệ số xác định R^2 phản ánh mức độ phù hợp của mô hình dữ đoán với dữ liệu thực nghiệm, có giá trị nằm trong khoảng 0 đến 1. Giá trị của R^2 càng gần tới 1 thì mô hình càng tốt.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

Sai số phần trăm tuyệt đối - MAPE đánh giá độ chính xác tương đối của mô hình dự đoán bằng cách đo lường phần trăm trung bình giữa giá trị dự đoán và giá trị thực tế.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \bar{y}|}{y_i} \quad (7)$$

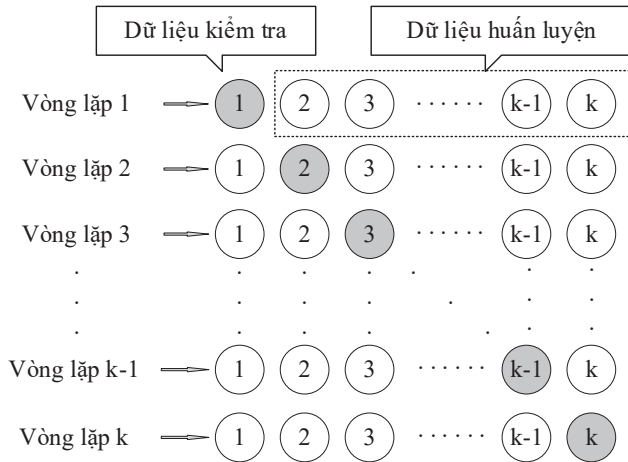
Sai số tuyệt đối trung bình - MSE đo lường độ sai lệch trung bình tuyệt đối giữa giá trị dự đoán và giá trị thực tế

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (8)$$

Căn bậc hai của sai số bình phương - RMSE thể hiện độ nhạy với các sai số lớn, thể hiện mức độ sai lệch tổng thể của mô hình dưới dạng trung bình bình phương có căn bậc hai.

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

Để đảm bảo hiệu suất và mức độ tổng quát hoá cho các mô hình học máy, nghiên cứu áp dụng kỹ thuật xác thực chéo (k - fold cross validation). Kỹ thuật này chia dữ liệu thành k nhóm ngẫu nhiên, lần lượt ($k-1$) nhóm dữ liệu tham gia quá trình huấn luyện và nhóm dữ liệu còn lại dùng để đánh giá hiệu suất kết quả mô hình. Kết quả hiệu suất mô hình học máy tổng hợp thường là giá trị trung bình của k lần đánh giá. Hình 1 mô tả sơ đồ phân chia dữ liệu thành k nhóm. Nghiên cứu sử dụng $k = 10$.

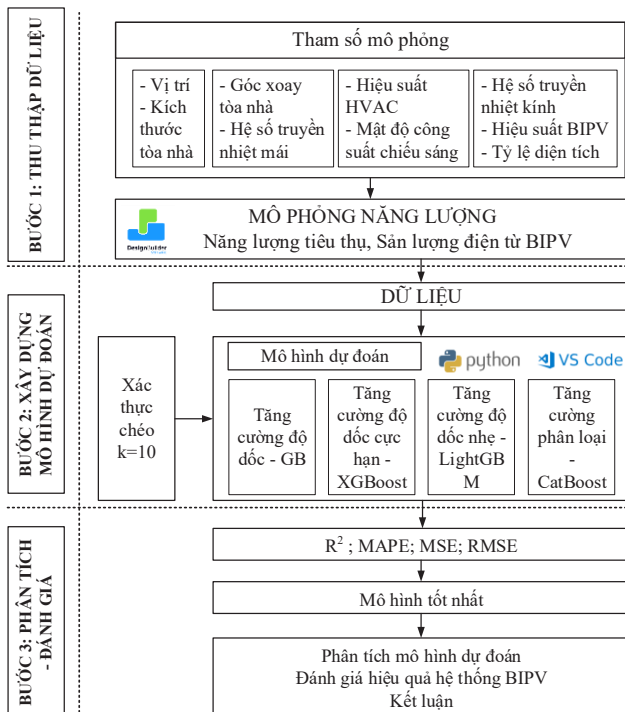


Hình 1. Kỹ thuật xác thực chéo.

3. Quy trình đề xuất và dữ liệu

3.1. Quy trình nghiên cứu

Hình 2 thể hiện quy trình nghiên cứu bao gồm ba bước. Thu thập dữ liệu thông qua quá trình mô phỏng năng lượng bằng phần mềm DesignBuilder. Bộ dữ liệu thu thập được phân chia thành 10 bộ dữ liệu con sử dụng cho các mô hình dự đoán. Hiệu suất các mô hình được đánh giá dựa trên các chỉ số R^2 , MAPE, MSE, RMSE. Hiệu quả hệ thống BIPV được đánh giá dựa trên mô hình dự đoán tốt nhất.



Hình 2. Quy trình nghiên cứu.

3.2. Thu thập dữ liệu

Dữ liệu được thu thập thông qua quá trình mô phỏng năng lượng bằng phần mềm DesignBuilder. Mỗi trường hợp mô phỏng tương ứng cho một kịch bản dữ liệu. Các tham số sử dụng để mô phỏng được thể hiện qua Bảng 1. Các tham số tương ứng cho các yếu tố ảnh hưởng đến các giải pháp tiết kiệm năng lượng tòa nhà. Với mỗi kịch bản mô phỏng, giá trị các tham số được xác định ngẫu nhiên và tổ hợp lại, kết hợp với kết quả mô phỏng để biểu thị cho một bộ dữ liệu.

Trường hợp nghiên cứu là một tòa nhà văn phòng gồm 1 tầng trệt và 7 tầng nổi với tổng diện tích sàn là 7593 m² với đầy đủ các chức năng của một tòa nhà thực tế. Lớp vỏ bao tòa nhà được làm bằng kính với diện tích kính từng mặt là 102,7 m², 96,6 m² và 102,7 m², nơi thực hiện nghiên cứu tích hợp hệ thống BIPV. Tầng trệt sử dụng kính thông thường, các tấm BIPV được tích hợp từ tầng 1 lên tới tầng 7 của tòa nhà. Tỷ lệ diện tích giữa tấm BIPV và kính thông thường được chỉnh sửa dựa trên tham số $S_n = (S_1; S_2; S_3)$ nằm trong đoạn giá trị 0 và 1. Với S_n là tỷ lệ diện tích tấm BIPV thì $(1 - S_n)$ là tỷ lệ diện tích kính.

Dữ liệu thời tiết được thiết lập tại trạm Tân Sơn Nhất TP.HCM theo dữ liệu của climate.onebuilding.org. Hình 3(a) thể hiện mô hình mô kết xuất của mô hình mô phỏng năng lượng bằng phần mềm DesignBuilder với hướng mặt trời di chuyển thay đổi theo từng tháng trong năm và hướng xoay của tòa nhà. Hình 3(b) thể hiện mô hình thông tin kỹ thuật của tòa nhà. Các tấm kính và tấm BIPV được mô tả trong hình lần lượt có màu xanh nước biển đậm và màu xanh da trời. Số lượng tấm BIPV và số lượng tấm kính được mô phỏng đúng theo các trường hợp thông số S_1, S_2, S_3 đã được xác định trong các kịch bản mô phỏng.

4. Kết quả nghiên cứu

4.1. Kết quả mô phỏng

Kết quả sau quá trình mô phỏng là bộ dữ liệu gồm 600 kịch bản tương ứng với từng trường hợp. Dữ liệu trích xuất từ quá trình mô phỏng bao gồm E (kWh/m²), S(m²) và PVE (kWh) lần lượt là tổng năng lượng tiêu thụ của tòa nhà, tổng diện tích sàn của tòa nhà và sản lượng điện của hệ thống BIPV. Năng lượng tiêu thụ ròng của tòa nhà tính toán khi sản lượng điện từ hệ thống BIPV sẽ được cung cấp cho quá trình vận hành tòa nhà theo công thức sau:

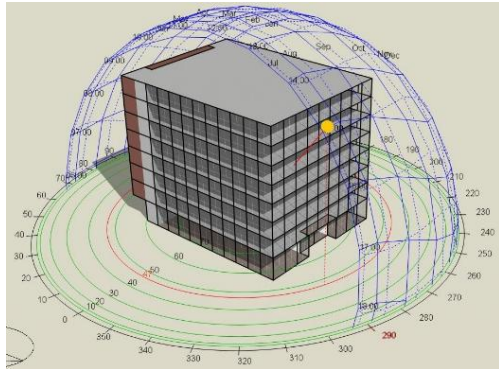
$$NetE = \frac{E \times S - PVE}{S} \quad (10)$$

với NetE là năng lượng tiêu thụ ròng của tòa nhà với nguồn từ lưới điện (kWh/m²/năm).

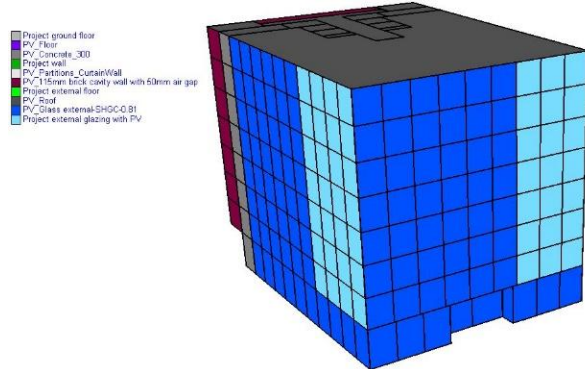
Kết quả của quá trình mô phỏng năng lượng là bộ dữ liệu bao gồm các tham số mô phỏng và kết quả mô phỏng sau khi xử lý dữ liệu bao gồm năng lượng tiêu thụ ròng (NetE) và sản lượng điện từ hệ thống BIPV (PVE). Bảng 2 thể hiện một phần của bộ dữ liệu này. Dữ liệu mô tả 600 kịch bản được xây dựng, giá trị tham số của mỗi kịch bản được xác định ngẫu nhiên và nằm trong khoảng cho phép. NetE và PVE là các giá trị mục tiêu cho quá trình huấn luyện mô hình dự đoán.

Bảng 1. Tham số mô phỏng năng lượng.

Tham số	Kí hiệu	Giá trị	Đơn vị
Góc xoay tòa nhà	Ori	0 – 360	độ
Hệ số hiệu quả năng lượng của hệ thống HVAC	COP	2,6 – 7,0	
Mật độ công suất chiếu sáng	LPD	7 - 13	W/m ²
Hệ số truyền nhiệt mái	UvR	0,239 – 1,923	W/(m ² .K)
Hệ số hấp thụ nhiệt của kính	SHGC	0,17 – 0,9	
Hiệu suất tấm BIPV	Effi	0,10 – 0,25	%
Tỷ lệ diện tích tấm BIPV	S ₁ , S ₂ , S ₃	0,0 – 1,0	%



(a) Mô hình kết xuất



(b) Mô hình thông tin kỹ thuật

Hình 3. Mô hình mô phỏng năng lượng.

Bảng 2. Kết quả mô phỏng năng lượng.

STT	Ori	LPD	COP	UvR	SHGC	Effi	S1	S2	S3	NetE	PVe
1	160	8,4	4,3	0,675	0,8	0,11	0	0	0,9	92,59	35344
2	170	9,2	6,3	1,352	0,26	0,14	0,1	1	0,7	67,09	113533
3	330	11,8	3,6	1,654	0,36	0,12	0,1	0,22	0,4	101,3	45846
...	...	...	...	...	...	...	...	...	...	...	...
598	60	9	5,9	1,547	0,47	0,13	0,6	0	0,1	77,89	46489
599	220	10,2	6,2	0,95	0,85	0,21	0,9	0,11	0,6	69,41	174307
600	210	11	6,9	0,728	0,54	0,15	0,6	0,89	0	71,48	123507

Để có cách nhìn tổng quát về mối quan hệ giữa các tham số, ma trận tương quan được xây dựng dựa trên bộ dữ liệu thu thập được ở Bảng 2. Ma trận phản ánh mối tương quan giữa các biến với giá trị giao động từ -1 (tương quan âm) đến 1 (tương quan dương) tương ứng với màu xanh và màu đỏ theo độ đậm nhạt. Hình 4 thể hiện ma trận tương quan giữa các tham số và kết quả mô phỏng.

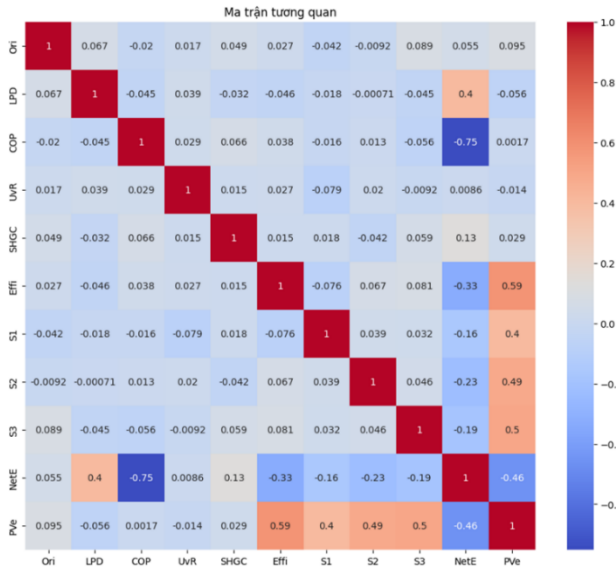
NetE có mối tương quan âm mạnh với COP (-0,75), trung bình với Effi (-0,33) và yếu với các tỷ lệ diện tích BIPV (từ -0,16 đến -0,23) cho thấy khi các giá trị COP, Effi, S₁, S₂, S₃ càng lớn thì năng lượng tiêu thụ ròng NetE càng nhỏ. Hệ số COP càng cao thì hiệu suất làm lạnh càng tốt, từ đó tiêu thụ ít năng lượng điện hơn; bên cạnh đó hiệu suất tấm BIPV cao tương ứng với diện tích lớn sẽ khiến sản lượng điện cung cấp cho quá trình vận hành lớn, giúp giảm năng lượng tiêu thụ ròng của tòa nhà. Trong khi đó LPD (0,4) có mối tương quan dương trung

bình và SHGC (0,13) có tương quan dương nhẹ với NetE cho thấy NetE tăng khi các giá trị này tăng. Điều này cho thấy mật độ công suất chiếu sáng cao sẽ tiêu tốn năng lượng nhiều, hệ số hấp thụ nhiệt của kính cao sẽ tổn năng lượng để làm mát hơn.

Sản lượng điện từ hệ thống BIPV được quyết định bởi hiệu suất Effi (0,59) và tỷ lệ diện tích S₁ (0,4), S₂ (0,49), S₃ (0,5) của các tấm BIPV. Hiệu suất càng cao, diện tích càng lớn sẽ mang lại sản lượng điện càng lớn. Các yếu tố còn lại có tương quan rất nhỏ và hầu như không ảnh hưởng đến sản lượng điện từ hệ thống BIPV. Mặc dù chỉ số LPD và COP là các yếu tố chính ảnh hưởng đến năng lượng tiêu thụ của tòa nhà nhưng các chỉ số Effi, S₁, S₂, S₃ cũng góp phần ảnh hưởng một phần tới NetE. Trong khi đó sản lượng điện Pve chỉ chịu ảnh hưởng lớn từ các chỉ số Effi, S₁, S₂, S₃ và chịu rất ít sự tác động bởi các tham số khác.

4.2. Kết quả các mô hình dự đoán

Mô hình dự đoán được xây dựng bằng ngôn ngữ Python qua môi trường VSCode. Các mô hình dự đoán được thiết lập theo cùng một điều kiện với các siêu tham số như nhau để có thể so sánh hiệu suất một cách khách quan nhất. Bảng 3 thể hiện các tham số khai báo trong các mô hình thuật toán học máy. Các tham số này bao gồm: “n_estimators” hoặc “interration” là số lượng cây hay số lần lặp được xây dựng trong mô hình boosting; “learning_rate” là hệ số học; “max_depth” là độ sâu tối đa của mỗi cây; “random_state” là chỉ số tái lập kết quả.



Hình 4. Ma trận tương quan giữa các tham số.

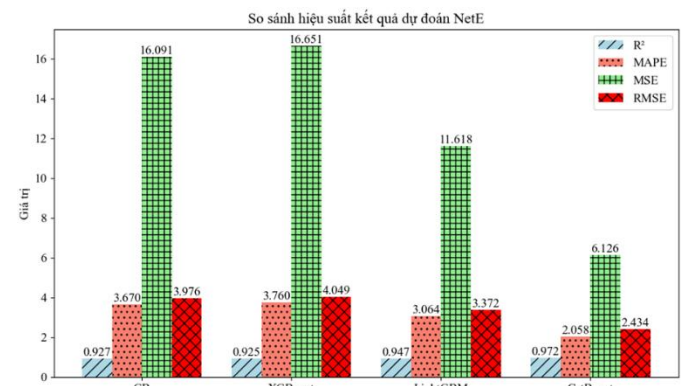
Bảng 3. Giá trị thiết lập cho các mô hình dự đoán.

Tham số	Giá trị
“n_estimators” , “interration”	500
“learning_rate”	0,05
“max_depth”	5
“random_state”	42

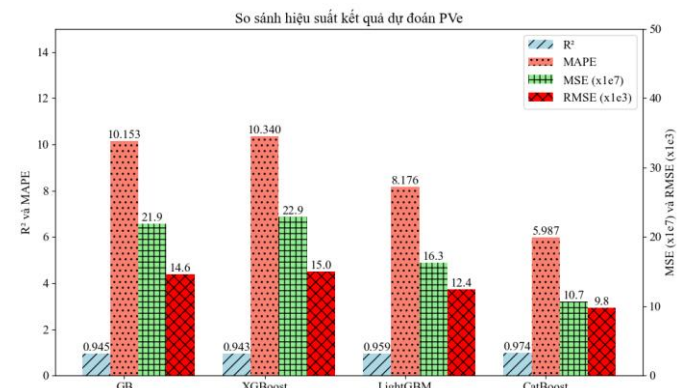
Hình 5 và Hình 6 thể hiện kết quả so sánh hiệu suất dự đoán năng lượng tiêu thụ ròng của tòa nhà và sản lượng điện từ hệ thống BIPV ứng với bốn thuật toán GB, XGBoost, LightGBM và CatBoost. Đối với mô hình dự đoán năng lượng tiêu thụ ròng NetE, kết quả cho thấy CatBoost vượt trội hơn hẳn các thuật toán còn lại với $R^2=0,972$ cao nhất và các chỉ số sai số thấp nhất $MAPE=2,058$, $MSE=6,126$, $RMSE=2,434$. GB và XGBoost có kết quả tương đương nhau với R^2 lần lượt là 0,927 và 0,925. LightGBM cũng đạt kết quả tốt với $R^2=0,947$. Tương tự mô hình dự đoán NetE, Catboost cũng là mô hình dự đoán PVE tốt nhất với $R^2=0,974$, sau đó tới LightGBM và cuối cùng là GB, XGBoost có kết quả tương đương nhau.

Hiệu suất của các mô hình dự đoán có sự khác biệt xuất phát từ các giải thuật được sử dụng. Mô hình dự đoán sử dụng thuật toán GB

và XGBoost đều sử dụng phương pháp học kết hợp (boosting) và bỏ túi (bagging). Ở đó phương pháp bỏ túi sẽ lấy ngẫu nhiên các bộ dữ liệu để xây dựng và phân tích dữ liệu dựa trên tất cả các tham số bao gồm các tham số có tương quan mạnh lẫn nhẹ. Đối với các biến có sự tương quan nhẹ được lựa chọn trong quá trình xây dựng cây quyết định sẽ làm cây học sai, không đóng góp quá nhiều vào trong quá trình huấn luyện. XGBoost cải tiến hơn so với GB với kỹ thuật điều chuẩn chi làm giảm quá khớp và tăng tốc độ học nên không đóng góp quá nhiều cho bộ dữ liệu này. Trong khi đó cả hai thuật toán LightGBM và CatBoost đều được cải tiến với các kỹ thuật xử lý dữ liệu nhằm giảm số nhiễu bằng EFB đối với LightGBM và Target Encoding đối với CatBoost. Các kỹ thuật này tập trung vào các dữ liệu có mối tương quan mạnh với mục tiêu dự đoán, lựa chọn các tham số có mức độ đóng góp lớn nên mang lại hiệu suất tốt hơn.

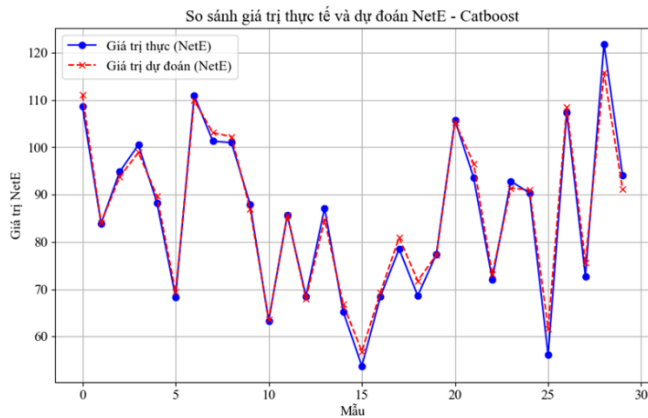


Hình 5. So sánh hiệu suất dự đoán NetE giữa các mô hình thuật toán.

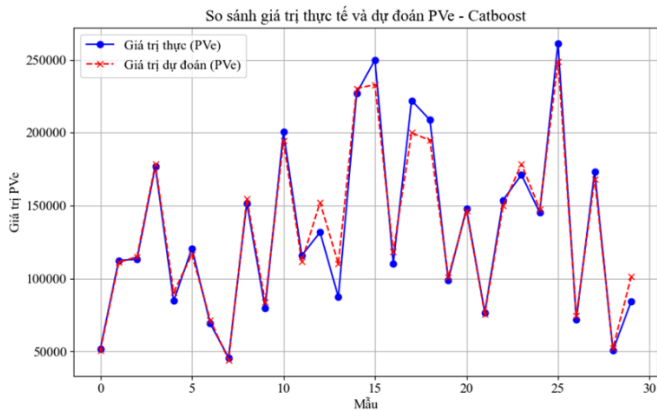


Hình 6. So sánh hiệu suất dự đoán PVE giữa các mô hình thuật toán.

Nhằm đánh giá mô hình dự đoán một cách toàn diện hơn, nghiên cứu xây dựng bộ dữ liệu gồm 30 kịch bản độc lập với 600 kịch bản dùng để huấn luyện mô hình và so sánh giá trị mô phỏng thực tế và giá trị dự đoán của mô hình tốt nhất. Hình 7 và Hình 8 thể hiện kết quả so sánh mô hình dự đoán NetE và PVE bằng thuật toán CatBoost 30 kịch bản này. Hiệu suất thể hiện qua Bảng 4. Chỉ số R^2 của kết quả dự đoán NetE và PVE lần lượt là 0,9825, 0,986 đều cao hơn so với hiệu suất mô hình CatBoost, các chỉ số sai số đều nhỏ hơn.



Hình 7. So sánh giữa giá trị dự đoán và thực tế của NetE sử dụng CatBoost.



Hình 8. So sánh giữa giá trị dự đoán và thực tế của PVe sử dụng CatBoost.

Bảng 4. Kết quả đánh giá dữ liệu kiểm tra.

Mục Tiêu	R^2	MAPE	MSE	RMSE
NetE	0,9825	2,17	5,1295	2,265
PVe	0,986	5,28	89924835	9482,87

Nghiên cứu sử dụng mô hình dự đoán để đánh giá sơ bộ hiệu quả của hệ thống BIPV khi tích hợp vào lớp vỏ tòa nhà văn phòng sử dụng kính ở điều kiện thời tiết tại trạm Tân Sơn Nhất, TP. Hồ Chí Minh. Các tham số góc xoay tòa nhà, mật độ công suất chiếu sáng, hệ số hiệu quả năng lượng của hệ thống HVAC, hệ số truyền nhiệt mái, hệ số hấp thụ nhiệt của kính và hiệu suất tấm BIPV được xác định như nhau tại ba trường hợp so sánh. Các trường hợp so sánh bao gồm khi thay thế toàn bộ hệ kính bằng hệ thống BIPV, thay thế một nửa và không sử dụng BIPV. Bảng 5 thể hiện kết quả so sánh của ba trường hợp tích hợp hệ thống BIPV. Kết quả cho thấy năng lượng tiêu thụ ròng của tòa nhà văn phòng có thể giảm từ 82,75 kWh/m²/năm xuống 54,35 kWh/m²/năm, giảm 34,3 % khi tích hợp toàn bộ hệ thống BIPV thay thế cho lớp kính tòa nhà. Khi tỷ lệ diện tích giữa kính và tấm BIPV bằng nhau thì năng

lượng tiêu thụ ròng là 68,84 kWh/m²/năm, giảm 16,8 % khi không tích hợp hệ thống BIPV.

Bảng 5. Đánh giá hiệu quả tòa nhà tích hợp BIPV.

Tham Số	Đơn vị	Không BIPV	50 % BIPV	100 % BIPV
Ori	độ	250	250	250
LPD	W/m ²	10	10	10
COP	-	6	6	6
UvR	W/(m ² .K)	1	1	1
SHGC	-	0,57	0,57	0,57
Effi	%	0,21	0,21	0,21
S1	m2	0	0,5	1,0
S2	m2	0	0,5	1,0
S3	m2	0	0,5	1,0
NetE	kWh/m ² /năm	82,75	68,84	54,34
PVe	kWh	0	150 478	309 636

5. Kết luận

Nghiên cứu này đề xuất các mô hình trí tuệ nhân tạo hiện đại bao gồm GB, XGBoost, LightGBM và CatBoost trong lĩnh vực dự đoán năng lượng tiêu thụ ròng của tòa nhà có tích hợp hệ thống BIPV cho các tòa nhà văn phòng hiện đại có lớp vỏ bao bằng kính. Dữ liệu được thu thập thông qua mô phỏng năng lượng tòa nhà văn phòng có ba mặt sử dụng kính làm lớp vỏ bao ở TP. Hồ Chí Minh bằng phần mềm DesignBuilder. Các mô hình dự đoán sử dụng với ngữ lập trình Python trong môi trường VSCode. Hiệu suất các mô hình dự đoán được đánh giá thông qua các hệ số xác định R^2 , sai số phần trăm tuyệt đối MAPE, sai số tuyệt đối trung bình MAE, căn bậc sai của sai số bình phương RMSE và kỹ thuật trao đổi chéo 10 lần nhằm nâng cao hiệu suất và giảm thiểu sai sót của mô hình dự đoán.

Kết quả nghiên cứu chỉ ra rằng mô hình dự đoán sử dụng thuật toán CatBoost là mô hình có hiệu quả tốt nhất so với các mô hình dự đoán sử dụng GB, XGBoost, LightGBM ở cả hai mục tiêu dự đoán là năng lượng tiêu thụ ròng (NetE) và sản lượng điện từ hệ thống BIPV (PVe) với chỉ số R^2 lần lượt là 0,972 và 0,974. Nghiên cứu cũng đánh giá sơ bộ hiệu quả của hệ thống BIPV khi được tích hợp vào trong lớp vỏ bao tòa nhà giúp giảm năng lượng tiêu thụ của tòa nhà lên tới 34,3% nhờ sản lượng điện từ hệ thống BIPV.

Kết quả nghiên cứu chứng minh được hiệu quả của hệ thống BIPV khi được tích hợp vào trong lớp vỏ bao tòa nhà bằng kính và hiệu quả của các mô hình trí tuệ nhân tạo trong việc dự đoán mức tiêu thụ năng lượng và sản lượng điện từ hệ thống BIPV. Kết quả nghiên cứu có thể làm cơ sở cho các nghiên cứu về tối ưu hóa đa mục tiêu cho tòa nhà được tích hợp hệ thống BIPV.

Lời cảm ơn

Chúng tôi xin cảm ơn Trường Đại học Bách Khoa, ĐHQG-HCM đã hỗ trợ cho nghiên cứu này.

Tài liệu tham khảo

- [1]. J. Davenport and N. Wayth, "Statistical review of world energy," *Energy institute*, 2023.
- [2]. H. T. Hoi, "Potential for solar energy development in Vietnam," *International Journal of Environmental Science and Development*, vol. 11, no. 7, pp. 358-364, 2020.
- [3]. M. Roginska-Niesluchowska, "Use of daylight and aesthetic image of glass facades in contemporary buildings," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 245, no. 8, p. 082035: IOP Publishing.
- [4]. M. J. Perez, V. Fthenakis, H. C. Kim, and A. O. Pereira, "Façade-integrated photovoltaics: a life cycle and performance assessment case study," *Progress in Photovoltaics: Research and Applications*, vol. 20, no. 8, pp. 975-990, 2012.
- [5]. A. Ghazali, L. C. Haw, S. Mat, and K. Sopian, "Performance and financial evaluation of various photovoltaic vertical facades on high-rise building in Malaysia," *Energy and Buildings*, vol. 134, pp. 306-318, 2017.
- [6]. H. Gholami, H. N. Røstvik, N. M. Kumar, and S. S. Chopra, "Lifecycle cost analysis (LCCA) of tailor-made building integrated photovoltaics (BIPV) façade: Solsmaragden case study in Norway," *Solar Energy*, vol. 211, pp. 488-502, 2020.
- [7]. Q. T. Nguyen, D. L. Luong, A. D. Pham, and Q. C. Truong, "Developing an optimization model of solar cell installation on building façades in high-rise buildings—A case study in Viet Nam," *GMSARN Int. J.*, vol. 15, no. 1, pp. 44-49, 2021.
- [8]. S. Attia, E. Gratia, A. De Herde, and J. L. Hensen, "Simulation-based decision support tool for early stages of zero-energy building design," *Energy and buildings*, vol. 49, pp. 2-15, 2012.
- [9]. R. Kabilan *et al.*, "Short-Term Power Prediction of Building Integrated Photovoltaic (BIPV) System Based on Machine Learning Algorithms," *International Journal of Photoenergy*, vol. 2021, no. 1, p. 5582418, 2021.
- [10]. S. T. Imalka, R. J. Yang, and Y. Zhao, "Machine learning driven building integrated photovoltaic (BIPV) envelope design optimization," *Energy and Buildings*, vol. 324, p. 114882, 2024.
- [11]. T. Đ. Học, "Tối ưu hóa ước tính mức tiêu thụ năng lượng trong các tòa nhà dựa trên các thuật toán trí tuệ nhân tạo," *Tạp chí Khoa học Công nghệ Xây dựng (TCKHCN XD)-ĐHXDH*, vol. 14, no. 1V, pp. 35-45, 2020.
- [12]. L. D. Long, "An AI-driven model for predicting and optimizing energy-efficient building envelopes," *Alexandria Engineering Journal*, vol. 79, pp. 480-501, 2023.
- [13]. J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189-1232, 2001.
- [14]. J. H. Friedman, "Stochastic gradient boosting," *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367-378, 2002.
- [15]. T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785-794.
- [16]. G. Ke *et al.*, "Lightgbm: A highly efficient gradient boosting decision tree," *Advances in neural information processing systems*, vol. 30, 2017.
- [17]. L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "CatBoost: unbiased boosting with categorical features," *Advances in neural information processing systems*, vol. 31, 2018.