AI-Integrated Urban Building Energy Simulation

A framework to forecast the morphological impact on daylight availability

Jayedi Aman[1], Jong Bum Kim[2], Debora Verniz[3]
[1,2,3]University of Missouri Columbia
[1]jayediaman@mail.misouri.edu [2]{kimjongb|vernizd}@missouri.edu

The research presents a computational framework to investigate the relationship between urban morphology and environmental performance metrics of buildings. Understanding how buildings interact with their surroundings is crucial in optimizing environmental performance. Current urban building energy simulation methods (UBES) often overlook the complex interaction between urban morphology and environmental performance across a diverse set of attributes, resulting in inaccuracies. The proposed framework integrates machine learning (ML) with physics-based simulations and includes Parametric Building Information Modeling, iterative physics-based simulations, Multi-Objective Optimization, and a graph neural network. The framework leverages the detailed analysis capabilities of physics-based simulations and the data processing strengths of ML to analyze urban morphological attributes. Evaluations indicate that the framework enhances prediction accuracy while considering the influence of urban morphology on environmental performance.

Keywords: *Urban Morphology, Urban Building Energy Modeling, Graph Neural Networks, Sustainable Urban Development, Environmental Performance, Multi-objective Optimization*

INTRODUCTION

Urban morphology has influences on the environmental performance of urban buildings. Understanding the interaction of buildings with their surroundings is necessary to improve their environmental performance and reduce the energy efficiency gap (Nutkiewicz et al. 2018). This necessitates the consideration of an array of morphological variables affecting building performance (Natanian 2021).

This research aims to investigate the interplay between urban morphology and environmental performance metrics for future constructions, with an emphasis on synergistically utilizing machine learning (ML) alongside physics-based simulations to enhance the predictability considering a wide range of urban morphological attributes.

Urban building energy simulation (UBES) workflows commonly utilize physics-based simulations or data-driven estimation methods, each with inherent limitations (Barbour et al. 2019). Physics-based simulations, though precise, are computationally intensive and time-consuming due to the incorporation of extensive building details (Kim et al. 2023). On the other hand, data-driven models expedite results but may compromise accuracy by not accounting for building physics. Consequently, UBES approaches often face trade-offs among detail, computational cost, and time, fostering inaccuracies and uncertainties (Hong et al. 2020).

This study aims to harness the capabilities of both physics-based simulations and ML by integrating them into a comprehensive computational framework. The developed framework includes a parametric building information model (BIM), iterative physics-based simulation, multi-objective optimization (MOO), and a graph neural network (GNN) as the ML algorithm. This approach is devised to enable an accurate evaluation of the environmental impact of different morphological attributes.

To gather data for the study, detailed information from various sources such as planning authority records, land use, and zoning regulations was collected. This allowed for the incorporation of a detailed three-dimensional context of both existing and future morphological developments on parametric BIM platform. Subsequently, an iterative physics-based simulation was used to analyze the daylight availability of future scenarios, and a large number of simulations were iterated to generate accurate simulation data for training and testing the ML algorithm.

The use of MOO fine-tuned four primary hyperparameters of the GNN algorithm. Specifically, MOO optimized the hyperparameters related to the number of layers and units in the neural network, learning rate, and dropout rate. The GNN algorithm was chosen for its ability to effectively model the relational structure of urban environments. Finally, regression analysis was utilized to compare the prediction accuracy and processing time of the proposed model with those of physics-based simulations.

This study presents an integrated framework that combines physics-based simulations and machine learning to comprehensively explore the relationship between urban morphology and environmental performance metrics in future constructions. This integrated approach aims to advance sustainable urban development and energy-efficient building design by leveraging the strengths of both methodologies.

BACKGROUND

Morphology & Daylight Availability

Urban morphology and the built environment are crucial for sustainable urban development. The spatial structure and organization of urban forms, such as buildings, streets, and open spaces, influence the interaction between future developments and their surroundings (Shi et al. 2017). A deep understanding of the complex relationship between urban morphology and the built environment is essential for optimizing environmental performance and fostering sustainable urban growth.

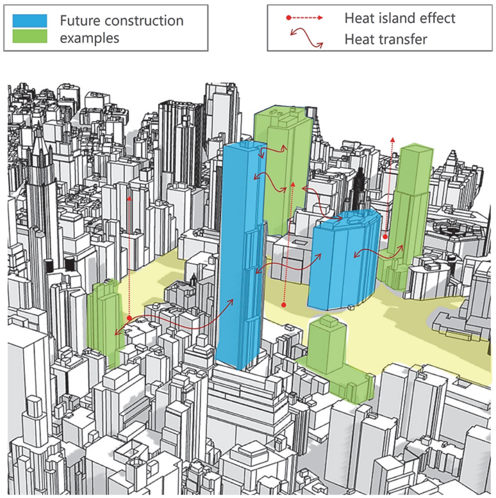

Figure 1
Morphological impact on daylight availability

Urban morphology significantly impacts the availability of daylight, which is a key aspect of environmental performance. The inter-building effects such as reflection, glare, and heat gain, as well as microclimatic effects like urban heat islands (as shown in figure 1), contribute to this impact. In addition, urban areas can experience sprawl, lengthened commutes, and increased infrastructure costs, which result in higher energy requirements.

Assessing the impact of urban morphology on daylight availability requires considering building and urban-level morphological variables. Essential factors include building height, width, floor area ratio, setback, and window-to-wall ratio among others (Kim et al. 2021). Morphological data can be collected from various sources, such as GIS, remote sensing, and sensors. Existing surrounding conditions, like neighboring building height, setbacks, streetscapes, and landscape, affect the environmental context of future constructions. Information about these conditions can be gathered from geospatial datasets. Finally, urban planning and design guidelines, such as neighborhood master plans, and land-use regulations, provide valuable information for these parameters.

ML algorithm & hyperparameters

ML algorithms, such as Artificial Neural Networks (ANN), Multiple Linear Regression (MLR), Deep Belief Networks (DBN), Support Vector Regression (SVR), Random Forests, Decision Trees, and Graph Neural Networks (GNN), have shown promising capabilities in predicting building energy consumption. For instance, Nutkiewicz et al. (2018) employed residual networks in conjunction with EnergyPlus to study the effects of geometry and physical efficiency retrofits on energy consumption. Similarly, Touzani et al. (2021) utilized a hybrid method combining gradient boosting and random forests to predict electricity consumption in commercial buildings. These studies exemplify the potential of ML algorithms in predicting urban environmental performance and underscore the importance of their ongoing development and application in the realm of sustainable urban planning and design.

Among the array of ML models, GNNs were specifically chosen for inclusion in the computational framework due to their adeptness in handling graph-structured data. This proficiency of GNNs to encode relationships and capture intricate interactions is crucial for generating accurate predictions, provided that an accurate training dataset is used. Furthermore, GNNs are inherently skilled at modeling spatial relationships among urban morphological attributes. This attribute is of vital importance for deciphering the nuanced interplay that ultimately shapes a building's environmental performance (Hu et al. 2022).

The effectiveness of GNNs, as with other ML algorithms, depends largely on the optimal selection of hyperparameters such as the number of epochs, hidden layer size, learning rate, and weight decay. These hyperparameters significantly influence the model's ability to generalize and accurately predict unseen data. MOO techniques, including evolutionary algorithms, are becoming increasingly popular for optimizing these hyperparameters. By optimizing for multiple objectives simultaneously, such as accuracy and generalization, MOO can help find the optimal set of hyperparameters for ML algorithms, including GNNs. This optimization is critical for ensuring that the predictions generated are reliable and relevant for practical application in urban building performance analysis.

PROPOSED UBES FRAMEWORK

The proposed framework comprises four essential steps (see figure 2):

- construct a parametric urban model in BIM to represent future developments and their surrounding context (**S1**);
- perform physics-based simulations to generate training, testing, and validation data (**S2**);
- optimize the hyperparameters of the GNN algorithm through MOO (**S3**);
- predict test results utilizing the optimal hyperparameters of the algorithm and validate the outcomes using the validation data (**S4**).

Parametric BIM

In this research, parametric BIM was employed to capture detailed information on surrounding conditions and to consider future developments based on urban design regulation plans. The process involved creating 3D models of the urban context, taking into account a diverse set of morphological

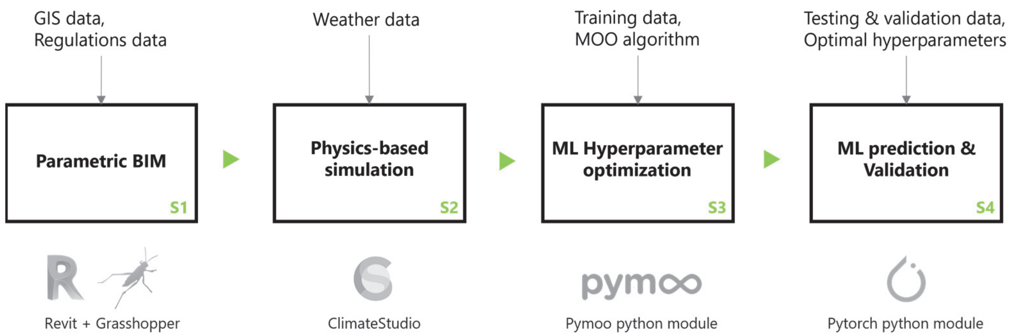

Figure 2
Workflow of the Proposed UBES

attributes. The models were combined with urban regulations and architectural data to represent both existing and future urban environments.

The parametric urban models were constructed using Autodesk Revit and its Application Programming Interface (API), drawing upon information gleaned from urban design regulation plans and GIS data. This enabled the creation of detailed 3D models that represent both the current and future urban developments. Revit's BIM-oriented capabilities were utilized for efficiently managing and integrating complex data, such as building heights, footprints, materials, and usage, which is particularly vital for the long-term objectives of this research.

Physics-based simulation

In this research, iterative physics-based simulation was used to develop training, testing, and validation datasets for prediction. Leveraging fundamental physics principles, these simulations accurately model complex systems such as building energy use and environmental performance. Through multiple simulation runs, the iterative approach refines model parameters (Haque et al. 2022).

The Revit Building Information Modeling (BIM) representation was imported into Grasshopper, following which the ClimateStudio simulation module was engaged to conduct daylight simulations. This module executed iterative simulations encompassing a diverse set of parameters to assess the impact of the surrounding environment on prospective developments. In doing so, a comprehensive dataset was produced to train and test the machine learning predictive model. Additionally, a separate set of simulations was conducted to create a validation dataset, which was crucial for evaluating the accuracy and reliability of the machine learning model in question.

ML hyperparameter optimization

The GNN algorithm was selected in the study for the development of the ML prediction model. Four key hyperparameters, including the number of epochs, hidden layer size, learning rate, and weight decay, significantly impact the performance of the GNN model regarding its learning process, capacity, and generalization (Yuan et al. 2021). tuning these hyperparameters with precision is crucial as they govern the model's ability to learn complex patterns

ML	Hyperparameter	Range
GNN Regression	Number of epochs	10- 500
	Hidden layer size	32- 512
	Learning rate	.0001- .01
	Weight decay	.0- .1

Table 1
specific ranges of each hyperparameter for MOO process.

and relationships in the data without overfitting. To optimize these hyperparameters, the study employed the Non- dominated Sorting Genetic Algorithm II (NSGA-II), a multi-objective optimization evolutionary algorithm that efficiently explores trade-offs between conflicting objectives.

NSGA-II was chosen to facilitate the balancing of R-squared (R^2), which measures the explained variance, and Root Mean Square Error (*RMSE*), which assesses predictive accuracy. This balance is critical in avoiding overfitting while ensuring generalizability. Table 1 shows the specific ranges selected for each hyperparameter based on literature, ensuring a comprehensive search for the optimal hyperparameter combination.

ML prediction & validation

Once the optimal GNN model was developed, it was utilized for predicting building performance metrics within the context of the specified urban environment. The trained model was subsequently applied to a testing dataset, with the intent of evaluating its generalization capabilities and accuracy in simulating real-world scenarios.

The validation process entailed a comparison of the predictions made by the GNN model with the actual performance metrics derived from physics-based simulations. This involved employing R^2 and *RMSE* metrics as assessment tools. These metrics allowed the research team to gauge the model's accuracy and its adeptness at generalizing to unseen data. Furthermore, by cross-examining these statistical measures, the research team could discern patterns and trends that would serve to highlight areas where the model may require further calibration. The validation process instilled confidence in the reliability of the ML prediction model, thus affirming its viability for application in urban planning and design tasks.

IMPLEMENTATION

This research conducted a case study of the Overland Park Downtown district in Kansas City to demonstrate the effectiveness of the proposed framework. Specifically, the study focused on two blocks (B01 and B02) and their surrounding buildings in the area. It was assumed in this study that B01 and B02, each consisting of two buildings, would be part of the future development, while the surrounding context would remain the same. The study incorporated the information of both the existing conditions and future development plans in parametric BIM platform (see figure 3).

To initiate the case study of the Overland Park Downtown district, detailed information on existing buildings was collected through the GIS mapping platform of Overland Park open datahub. The platform facilitated the collection of 2D dimensional information, such as building footprints, latitude-longitude, and area, from property and zoning shape files, while the height information was gathered from LiDAR data. These data were then integrated into the 3D parametric BIM platform to create accurate representations of the existing built forms. By utilizing the parametric BIM platform, the team captured detailed information on the surrounding conditions.

Once the necessary information was obtained, the research team analyzed the regulating map, transect types, and building design standards of the Overland Park Downtown district to construct two block models (B01 and B02) using Revit and its custom modeling interface. The district subdivision information, obtained from GIS data, was incorporated into the models, allowing for an accurate representation of the surroundings context and future development plans as per the Smart Growth regulations. By considering the building design standards outlined in the regulation, the team was able to model the building sections and ensure the accuracy of the 3D block models. The physics-based simulation process in this study utilized several input parameters to generate training, testing, and validation datasets for the two block models (B01 and B02) in the Overland Park Downtown district.

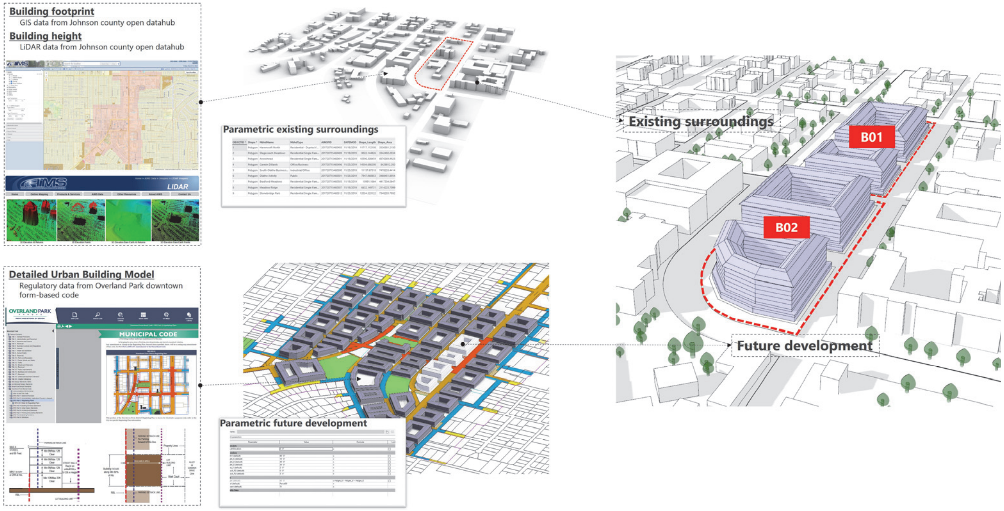

Figure 3
Data flow for developing 3D urban context in BIM incorporating future development and existing surroundings

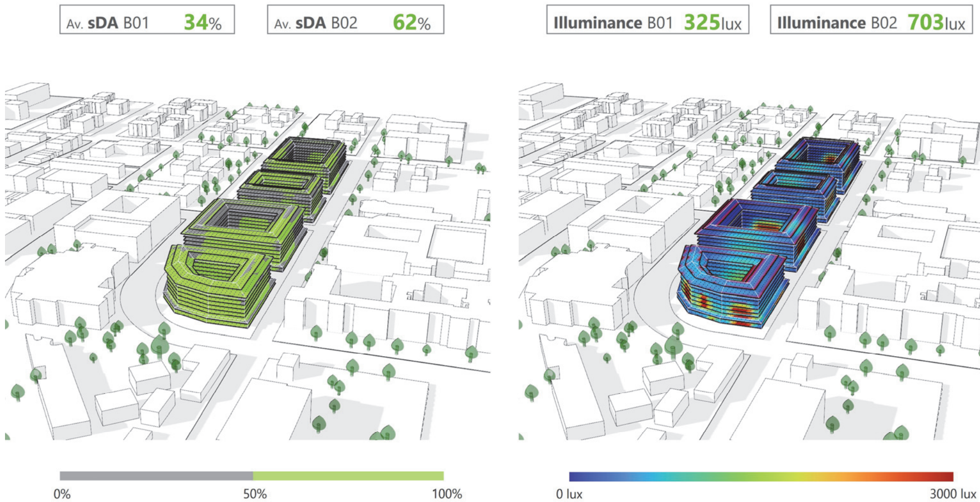

Figure 4
Visualization of sDA and Mean Illuminance for future construction

The parameters included:

- **Height:** ranging from 24-60ft for B01 and B02
- **Depth:** ranging from 60-80ft for B01 and B02
- **Setback**: ranging from 5-10ft for B01 and B02
-
- **Orientation**: ranging from 0-30 degrees for B01 and B02
- **Fenestration:** South_WWR and West_WWR ranging from 10-90%

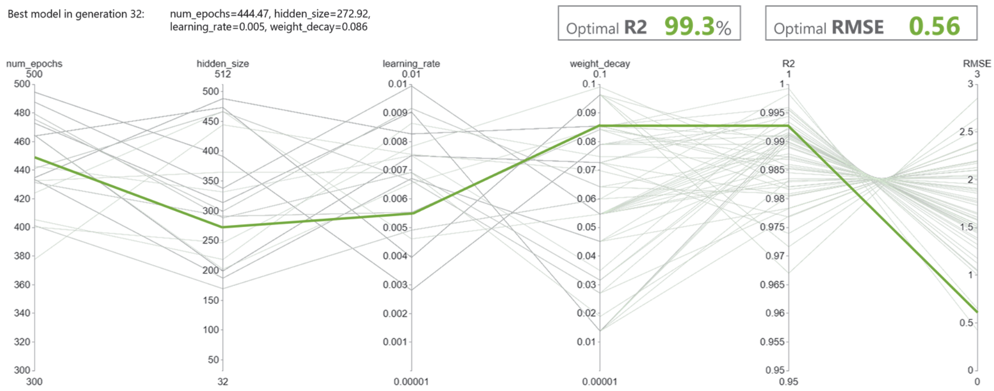

Figure 5
MOO generation results and the optimal hyperparameter values

The simulation process in this study utilized these input parameters to compute the spatial daylight autonomy (sDA) values for each iteration. The sDA measures the percentage of the occupied area of a building receiving a minimum level of daylight during a specific time, typically 300 lux for at least 50% of the annual occupied hours. The simulation process involved the creation of a sampling dataset, with 1000 iterations conducted for training and testing. An additional 500 simulations were conducted for validation purposes.

An iterative process was conducted to simulate and store average sDA values for B01 and B02 in specified datasets, as shown in figure 4. Using this training data, MOO was employed to optimize the hyperparameters of the GNN model with the trade-off model retained for subsequent predictions. The Pareto-optimal solutions yielded through MOO offered insights into the interplay between R^2 and *RMSE*.

A code snippet shows the schematic flow of the prediction model generation and storage, including the use of PyTorch to implement the GNN regressor model and Pymoo for the multi-objective optimization process.

```
total_generations = 50 #varies based on problem

for gen in range(1, total_generations + 1):
    res = minimize(problem, algorithm)
    toff_model_params = save_results(res)
toff_model_epochs = int(toff_model_params[0])

for epoch in range(toff_model_epochs):
    optimizer.zero_grad()
    outputs = toff_model(train_x, edge_index)
    loss = criterion(outputs, train_y)
    loss.backward()
    optimizer.step()
torch.save(toff_model.state_dict(), "toff_model.pth")
```

After identifying the optimal hyperparameter combination, the GNN regression model was trained using the data generated from iterative physics-based simulation. The saved GNN regressor model was then employed for prediction and validation tasks, significantly reducing computational resources and prediction time while maintaining high accuracy.

RESULTS

A number of validation and comparison analyses were carried out to assess the effectiveness of the proposed UBES framework. The optimization and prediction accuracy analyses showed that the framework had commendable accuracy. This result is particularly noteworthy in light of the system's complexity, particularly in the context of energy simulations for urban buildings.

Optimization accuracy

The MOO algorithm underwent 50 generations, and the tradeoff model was identified in generation 32 with these hyperparameter values: number of epochs=444.47, hidden layer size=272.92, learning rate=0.005, and weight decay=0.086. The optimal R^2 value achieved was an impressive 99.3%, while the optimal root mean squared error (*RMSE*) was 0.56. Figure 5 depicts a parallel plot that visualizes the generation results and the optimal values (see figure 5).

Additionally, R^2 and *RMSE* plots were generated to provide a deeper insight into the optimization accuracy. The R^2 plot consistently maintained high accuracy throughout the 50 generations, whereas the *RMSE* plot showcased a favorable decrease in the error metric across the generations. This is indicative of the framework's robustness and ability to learn effectively. Figures 6 and 7 display the evolution of R^2 and *RMSE* values across 50 generations (see figure 6 and figure 7, respectively).

Prediction accuracy

For the prediction accuracy assessment, the UBES framework was put to test using a dataset composed of various building design parameters and the corresponding simulated sDA values. The prediction results underwent a comparison with the simulation results via regression analysis, resulting in an R^2_{adj} value of 96.78%. This high value reflects a strong correlation between the predicted and simulated sDA values, providing preliminary evidence of the UBES framework's efficacy in predicting building sDA values based on design parameters (as shown in figure 8). The high R^2_{adj} value signifies that a substantial portion of the variability in the sDA

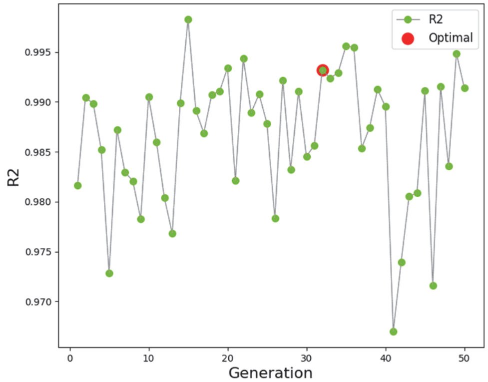

Figure 6
Tradeoff R2 values over 50 generations and optimal R2 result

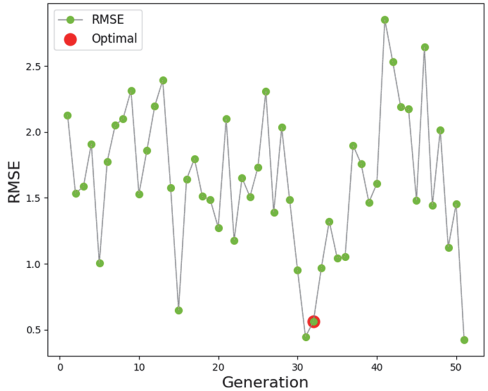

Figure 7
Tradeoff RMSE values over 50 generations and optimal RMSE result

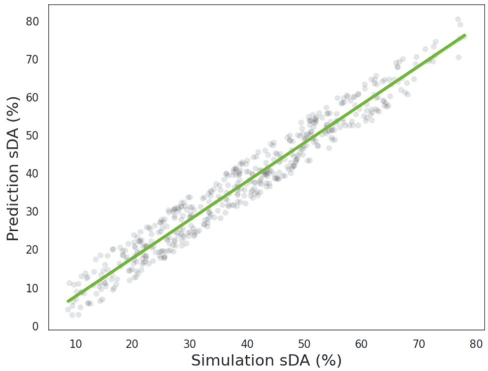

Figure 8
Regression analysis of simulation and prediction results

values can be explained by the model, highlighting the model's ability to capture the underlying trends and relationships in the data.

Further, the RMSE was employed to gauge the accuracy of the predicted values. The *RMSE* value stood at 1.54, which is relatively low considering the range of sDA values in the dataset. This result suggests that the predictions made by the framework are closely aligned with the actual values, demonstrating its potential accuracy.

In summary, these results are indicative of the proposed UBES framework's potential for high prediction accuracy. The integration of GNNs with physics-based simulations in this framework suggests a step in the direction of potentially aiding sustainable urban development.

DISCUSSION & FUTURE WORK

The proposed UBES framework addresses the critical role of urban morphology in building environmental performance by integrating machine learning models with physics-based simulations through parametric BIM. This integration enhances the predictability of environmental impact assessments of various morphological attributes, as the machine learning models benefit from the physics-based simulations' rigor. The framework is particularly valuable during the early design stages, making it a significant contribution to sustainable urban development and building energy simulation.

While the framework demonstrated promising results in terms of predictability, there are limitations, such as the small area analyzed and reliance on multiple software dependencies, and the challenges associated with the interpretability of ML models. Moreover, it is essential to recognize that the integration of ML models does not supersede physics-based simulations but rather works in concert with them to capture a broader range of morphological attributes. Data availability and interpretability of ML models may also pose challenges.

Future work could focus on expanding the framework's scope by including more blocks or different case areas for a rigorous validation process. Integrating multiple conflicting objectives, such as energy use intensity and carbon emissions, into the framework could provide recommendations for future optimal sustainable design development. Efficiency in prediction processing time, improving the interpretability of ML models, and reducing the number of software dependencies required for implementation could further enhance the framework's application and impact.

In conclusion, the proposed UBES framework signifies an attempt at fostering sustainable urban development and energy-conscious building design by synergistically integrating machine learning with physics-based simulations. The approach aims to enhance the predictive capabilities with a focus on accuracy in building energy simulations during the early design stages. While still in its nascent stages, it may provide some degree of insights for stakeholders, energy modelers, and designers. The framework aspires to aid in informed decision-making and support sustainable urban development through an inclusive consideration of various urban morphological attributes.

REFERENCES

Barbour, E. *et al.* (2019) ‘Planning for sustainable cities by estimating building occupancy with mobile phones’, *Nature Communications*, 10(1), p. 3736. Available at: https://doi.org/10.1038/s41467-019-11685-w.

Blank, J. and Deb, K. (2020) ‘Pymoo: Multi-Objective Optimization in Python’, *IEEE Access*, 8, pp. 89497–89509. Available at: https://doi.org/10.1109/ACCESS.2020.2990567.

Haque, Md.O. *et al.* (2022) ‘Evaluating the Daylighting and Energy Performance of Container Housing Systems in Coastal Regions: An Iterative Modeling Approach’, in H. Altan et al. (eds) *Advances in Architecture, Engineering and Technology*. Cham: Springer International Publishing (Advances in Science, Technology & Innovation), pp. 153–165. Available at: https://doi.org/10.1007/978-3-031-11232-4_13.

Hong, T. *et al.* (2020) 'State-of-the-art on research and applications of machine learning in the building life cycle', *Energy and Buildings*, 212, p. 109831. Available at: https://doi.org/10.1016/j.enbuild.2020.109831.

Hu, Y. *et al.* (2022) 'Times series forecasting for urban building energy consumption based on graph convolutional network', *Applied Energy*, 307, p. 118231. Available at: https://doi.org/10.1016/j.apenergy.2021.118231.

Kim, J.B., Aman, J. and Balakrishnan, B. (2021) 'Forecasting performance of Smart Growth development with parametric BIM-based microclimate simulations', in A. Globa, J. van Ameijde, A. Fingrut, N. Kim, T.T.S. Lo (eds.), PROJECTIONS - Proceedings of the 26th CAADRIA Conference - Volume 1, The Chinese University of Hong Kong and Online, Hong Kong, 29 March - 1 April 2021, pp. 411-420. CUMINCAD. Available at: https://doi.org/10.52842/conf.caadria.2021.1.411.

Kim, J.B., Kim, S. and Aman, J. (2023) 'An Urban Building Energy Simulation Method Integrating Parametric BIM and Machine Learning', in HUMAN-CENTRIC - Proceedings of the 28th CAADRIA Conference - Volume 1, Navrangpura, India, 18 March - 24 March 2023, pp. 665-674. CUMINCAD. Available at: https://caadria2023.org/conference

Natanian, J. (2021) *Beyond Zero Energy Districts: A Holistic Energy and Environmental Quality Evaluation Workflow for Dense Urban Contexts in Hot Climates*. Technische Universität München. Available at: https://mediatum.ub.tum.de/?id=1554761.

Nutkiewicz, A., Yang, Z. and Jain, R.K. (2018) 'Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow', *Applied Energy*, 225, pp. 1176–1189. Available at: https://doi.org/10.1016/j.apenergy.2018.05.023.

Shi, Z., Fonseca, J.A. and Schlueter, A. (2017) 'A review of simulation-based urban form generation and optimization for energy-driven urban design', *Building and Environment*, 121, pp. 119–129. Available at: https://doi.org/10.1016/j.buildenv.2017.05.006.

Touzani, S. *et al.* (2021) 'Determination of Phenolic Compounds in Various Propolis Samples Collected from an African and an Asian Region and Their Impact on Antioxidant and Antibacterial Activities', *Molecules*, 26(15), p. 4589. Available at: https://doi.org/10.3390/molecules26154589.

Yang, L. and Shami, A. (2020) 'On hyperparameter optimization of machine learning algorithms: Theory and practice', *Neurocomputing*, 415, pp. 295–316. Available at: https://doi.org/10.1016/j.neucom.2020.07.061.

Yuan, Y. *et al.* (2021) 'A Novel Genetic Algorithm with Hierarchical Evaluation Strategy for Hyperparameter Optimisation of Graph Neural Networks'. arXiv. Available at: https://doi.org/10.48550/arXiv.2101.09300.