BIM-BASED ENERGY CONSUMPTION ESTIMATION USING A DATA-DRIVEN MODEL

HAMID KIAVARZ MOGHADAM

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATE STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN EARTH AND SPACE SCIENCE YORK UNIVERSITY
TORONTO, ONTARIO

May 2024

ABSTRACT

Building Information Modeling (BIM) is undergoing rapid technological evolution in the building construction industry. Recently, employing BIM as a building 3D digital model in Building Energy Consumption Estimation (BECE) has gained momentum because of the enriched geometric and semantic information. Indeed, indoor BECE notably depends on the semantics, geometry (building elements and shapes), and topology information of the building's elements to recognize the spaces in a building with high energy demand.

However, despite extensive studies on applying the BIM and Industry Foundation Classes (IFC) as an open standard data model for BIM in BECE analysis, employing the full potential of the BIM remains poor due to its data model complexity and incompatibility with BECE data-driven algorithms. There is a significant lack of building energy modeling in using the detailed geometry, semantic, and 3D topology information in BECE data-driven models. The objective of this dissertation is to develop an innovative and comprehensive framework called space-based precise building energy consumption estimation using BIM. In this research, a framework is developed to convert the IFC model into a space-based graph, including the geometry, semantic, and topology information on the proposed graph nodes and edges. The graph is compatible with the machine learning algorithm. A graph-based classification algorithm is suggested in this research to find critical spaces in the building for energy consumption. This research proposed a prescriptive model by integrating building energy simulation with optimization techniques, using BIM data and a Genetic Algorithm (GA) to develop a prescriptive model for indoor building design. The study focuses on space-based BECE analysis, leveraging BIM interoperability to recommend optimal solutions. The proposed model employs the value engineering method to

balance energy consumption, functionality, and cost, providing engineers and designers with insights to optimize building performance effectively. This approach enhances energy efficiency and offers substantial design optimization solutions, bridging the gap between energy prediction and practical application in the architectural, engineering, and construction (AEC) industry.

The outcomes of this study are conducive to contemporary data-driven models in BIM and indoor BECE analysis. This provides a comprehensive perspective on both present and prospective requirements for BIM in the estimation of building energy consumption. The study integrates various sectors, including architecture, construction, machine learning, and 3D geospatial analysis, aiming to derive comprehensive and optimal solutions. Furthermore, it underscores the necessity for future multidisciplinary research by unfolding existing gaps and limitations.

ACKNOWLEDGMENTS

I sincerely thank Dr. Mojgan Jadidi, my supervisor, for giving me the invaluable opportunity to undertake my Ph.D. research under her expert guidance. The past five years have been marked by Dr. Jadidi's unwavering support, profound patience, and insightful recommendations, shaping my research trajectory.

I am deeply thankful for the commendable recommendations from Dr. Gunho Sonh and Dr. Abbas Rajabifard, whose expertise has further enriched my academic pursuits. The autonomy and encouragement bestowed upon me by Dr. Jadidi and my committee members have been pivotal, allowing me to work independently and discover my unique research direction—an aspect of my academic experience that I hold in the highest regard.

The support extended by the Lassonde School of Engineering, particularly the Department of Earth and Space Science and Engineering at York University, has been a cornerstone of my academic journey. The numerous graduate support and mentorship opportunities provided by my supervisors have not only facilitated my academic growth but have also contributed significantly to my overall development as a researcher.

In addition, I express my gratitude to Payam Esmaili, our esteemed industry collaborator, for his invaluable contributions. His expertise and collaboration in testing and analyzing our dataset and outputs using RETScreen software have elevated the impact and relevance of our achievements. His perspective as a knowledge expert has been instrumental in broadening the scope of our research. I extend heartfelt gratitude to my family, especially my wife, Samin, for unwavering support throughout my PhD journey.

In conclusion, my heartfelt thanks go out to Dr. Mojgan Jadidi, Dr. Gunho Sonh, Dr. Abbas Rajabifard, the Lassonde School of Engineering, and Payam Esmaili for their unwavering support and contributions to my academic and research endeavors. Each of you has played a crucial role in shaping my academic journey, and I am sincerely grateful for the privilege of working alongside such esteemed mentors and collaborators.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

ACID	Atomicity, Consistency, Isolation, Durability
AEC	Architectural, Engineering, and Construction
AGC	Associated General Contractors of America
ANN	Artificial Neural Network
API	Application Programming Interface
BECE	Building Energy Consumption Estimation
BEM	Building Energy Modeling
BGD	Batch Gradient Descent
BIM	Building Information Modeling
BSA	BuildingSMART Alliance
CEP	Cost Evaluation Phase
CI	Cost Index
DNN	Deep Neural Network
EUI	Energy Use Intensity
FA	Firefly Algorithm
FI	Functional Index

GA	Genetic Algorithm
GAT	Graph Attention Networks
GCM	Graph Convolutional Network
GIS	Geographical Information Systems
GNN	Graph Neural Network
GSA	Gravitational Search Algorithm
HVAC	Heating, Ventilation, and Air Conditioning
IAI	International Alliance for Interoperability
IFC	Industry Foundation Classes
LINMAP	Linear Programming Techniques for Multidimensional Analysis of Preference
LOD	Level of Details
LSTM	Long Short-Term Memory
MLP	Multi-Layer Perceptron neural network
MSE	Mean Squared Error
NBIMS-US	National BIM Standard-United States
GraphSAGE	Graph Sample and Aggregation
NE-GraphSAGE	Node-Edge GraphSAGE

NF	Normalization Function
NIBS	National Institute of Building Sciences
NSGA-II	Non-Dominated Sorting Genetic Algorithm-II
OWL	Web Ontology Language
PEP	Prescription Evaluation Phase
PSO	Particle Swarm Optimization
PVI	Prescription Value Index
R-Value	Resistance Value
SE	Squared Error
SFEP	Space Functionality Evaluation Phase
SFI	Space Functionality Index
SHAP	SHapley Additive exPlanations
SVM	Support Vector Machine
TOPICS	Order Preference by Similarity to Ideal Solution

CHAPTER 1: INTRODUCTION

1.1. BACKGROUND AND MOTIVATION

For over a decade, building 3D modeling has fundamentally transformed the perspective of Building Information Modeling (BIM) [1]. Within the realm of BIM, the complex representation of physical building elements, their nuanced properties, and the complex inter-element relationships have become the cornerstone [2]. The National BIM Standard-United States™ (NBIMS-US™) defines BIM as the digital articulation of a facility's physical and functional attributes. It stands at the forefront of evolving technologies, utilizing diverse software tools to streamline the design process [3]. BIM is a collaborative platform fostering joint planning, design, construction, and maintenance, which is a life cycle effort for new buildings. This collaborative data environment is instrumental for developers, businesses, and government bodies to engage seamlessly with consultants and contractors throughout various project phases [3].

Employing 2D geometry data in traditional methods presents several challenges in design and construction. One primary limitation is the lack of depth and spatial information inherent in 2D representations [4], [5]. This deficiency can lead to difficulty accurately visualizing and communicating complex three-dimensional structures and spatial relationships. Additionally, 2D data often results in a fragmented representation of the project, making it challenging to capture and manage complex details and dependencies within a building or infrastructure[6]. Beyond the limitations associated with the lack of spatial information, using 2D data in design and construction comes with additional challenges. One significant challenge is the potential for misinterpretation and miscommunication. Two-dimensional drawings often require a high level of technical expertise to interpret correctly, and the risk of errors or misunderstandings increases when stakeholders, such as architects, engineers, and contractors, rely solely on 2D representations. This can lead to discrepancies in the construction process, resulting in rework, delays, and increased costs [4], [6].

Another challenge is the difficulty in building maintenance and managing changes efficiently. Alterations to the design in a 2D context may necessitate multiple revisions across different drawings, making tracking and maintaining accuracy cumbersome. This lack of version control can lead to discrepancies between drawings, and ensuring that all stakeholders are working with the latest and most accurate information becomes challenging [7].

In summary, the challenges of using 2D data extend beyond spatial limitations, encompassing interpretation, change management, and collaboration, which can impede the efficiency and accuracy of design and construction processes. These challenges highlight the need for more advanced, three-dimensional approaches, such as BIM, to address these shortcomings and enhance overall project outcomes.

The Associated General Contractors of America (AGC) further delineates BIM as utilizing computer software to model and simulate the construction and operation of a facility. This involves creating a digital representation of a building that is not only rich in detail but also object-oriented, intelligent, and parametric. Such a model caters to many user needs and facilitates planners' decision-making. Also, the BIM definition emphasizes the significance of extracting and analyzing data to enhance project delivery [3].

According to the US National BIM Standard, BIM is a digital portrayal that comprehensively encapsulates a building's physical and functional traits. It operates as an information-sharing hub that proves invaluable for decision-making across the entire lifespan of a building. This collaborative tool allows stakeholders to seamlessly access, modify, and exchange data, reinforcing their collective responsibilities. Throughout different project stages, BIM proves to be an indispensable asset for architects, engineers, and construction professionals who rely on popular BIM authoring tools such as Revit, ArchiCAD, Tekla, Allplan, and MicroStation for efficient design, planning, and management of building projects [9].

While several standards have emerged to encapsulate BIM's data complexities, the Industry Foundation Classes (IFC) have become the industry's premier open standard, initially introduced by the International Alliance for Interoperability, now recognized as BuildingSMART Alliance [8]. The IFC model serves as a sophisticated object-oriented data framework, encompassing not only the three-dimensional geometry but also spatial relations (topology), precise quantities, and the semantic properties of building elements [9]. The essence of IFC lies in its ability to address the challenges of interoperability across diverse disciplines, playing a pivotal role in the seamless exchange of construction and building management components [10]. Disciplines ranging from facility management to energy efficiency studies critically rely on the precise geometric, semantic, and topological information encoded in IFC models to ensure the accuracy and reliability of their outcomes. Moreover, the relevance of 3D information within IFC models has significantly expanded, finding resonance in contemporary applications, particularly in the dynamic domains of smart cities and buildings, where its potential is increasingly being harnessed [11].

Despite its extensive applications, BIM data are underutilized in data-driven and machine-learning models due to the complexity of the data model and incompatibility with machine-learning algorithms. In a recent study, N. Fumo [8] emphasizes the growing importance of the BIM and IFC models as 3D data sources for Building Energy Consumption Estimation (BECE), highlighting their crucial role in shaping the future. BECE relies on the geometry, topology, and material information of building elements to identify high-energy-demand spaces within a building [6], [9], enabling the evaluation of different building design alternatives and operational strategies.

Physical modeling and data-driven approaches are widely used in BECE [10]. The physical models, also known as engineering methods or white-box models, rely on thermodynamic rules for detailed energy modeling and analysis. These models calculate energy consumption based on detailed building and environmental parameters such as construction details, operation schedules, and HVAC design [12]. However, two shortcomings limit the BECE physical models: Firstly, there are limitations in inputting physical parameters. Secondly, these systems are not designed to utilize geometric information and modern data models, such as BIM, as influential parameters for estimating energy consumption [13], [14].

Conversely, data-driven methods estimate building energy consumption by learning from historical data and building properties [6], [8]. While these models can handle any input parameters, they require sufficient data for accurate BECE analysis. The identified research gap is to explore the potential of using BIM in data-driven (machine learning) models for indoor BECE analysis by leveraging detailed 3D geometry, material, and topology information. By utilizing BIM's comprehensive digital representation of a building's physical and functional characteristics, along with detailed material properties, data-driven models can achieve higher accuracy in energy simulations. This integration enables the seamless incorporation of diverse datasets, including architectural, structural, mechanical, and electrical data, allowing for holistic analyses of energy performance.

Furthermore, the parameterization of building components facilitated by detailed material information empowers sensitivity analyses, enabling the evaluation of various design and material options for optimal energy efficiency. Through dynamic information that considers factors such as weather conditions and occupancy patterns, these models can provide precise predictions of energy consumption and demand. Ultimately, BIM, 3D information, topology, and detailed material data not only enhance the accuracy of energy modeling but also facilitate effective visualization, communication, and lifecycle analysis, empowering stakeholders to make informed decisions toward sustainable building design and operation [1], [15], [16]. However, there is limited information on how BIM can enhance a building's energy efficiency because of the complexity and incompatibility of the BIM data model with data-driven methods in BECE data-driven methods which motivates this study to find the solution to tackle these limitations.

1.2. PROBLEM STATEMENT

BIM is a robust data source that can provide building element information for accurately estimating indoor energy consumption in buildings. The input process becomes more efficient and detailed when BIM is integrated as the primary data source for data-driven energy estimation. This paradigm shift allows for the adoption of 3D spatial whole-building models, deviating from the reliance on space-based models in current energy estimation practices. Through BIM, energy estimation gains access to extensive building elements and topology data, representing a substantial improvement in the precision and depth of information available for analysis. The optimized generation of a building energy model from a BIM model can enhance the integration of the energy simulation process in design, increase its efficiency and efficacy, and save time. Multiple studies across the world have recently studied the impact of BIM in respective local contexts.[17]. Although the potential benefits of BIM in the energy estimation process are significant, relatively little research has been employed on BIM in energy estimation using data-driven models.

Additionally, limited information is available on how BIM can enhance a building's energy efficiency. The primary reason is that the complexity and incompatibility of the BIM data model with data-driven methods restrict its use in BECE data-driven methods. Because of the mentioned challenges of using the IFC model in BECE data-driven methods, most of the published methodologies solely focus on one specific part of BIM data (e.g., geometry or HVAC) and use traditional physical simulation engines (e.g., EnergyPlus) [18]. Therefore, utilizing an efficient data model to have the following three characteristics is necessary: 1) the capability of the data model to be connected to different data sources; 2) the capability of the data model to preserve geometric, semantic, and relationship information; and 3) compatibility of the data model with the machine learning algorithm. The primary issue addressed in this research is broken down into three subsidiary problems:

1.2.1. Challenges in Extracting The BECE-Based Geometry and Topology Information From BIM and IFC Model

The current state of the art in estimating building energy consumption highlights the advantageous use of IFC models for extracting building semantics, geometric information, and 3D spatial relationships between spaces [15]. Efficient 3D analysis of building energy consumption relies on the semantic information of the indoor environment, the geometric description of IFC entities, and their spatial relationships. The geometry and material information of entities such as walls, windows, roofs, and floors are crucial for each space in the IFC, affecting the space's heat gain and loss. Therefore, it is essential to accurately capture the geometry and material properties of these entities for each space. Most IFC-based building energy modeling depends on the information stored in the model. However, incomplete or erroneous implementation of the IFC and implicit topological information, commonly found in many IFC export modules, lead to inaccurate geometric information, such as the wall and window areas of each space.

Additionally, extracting accurate topological information, such as the spatial relationships between spaces, is challenging from IFC models. These errors result in invalid graphs for BECE-based information extraction and incorrect spatial relationships. Consequently, advanced geometric processing is necessary to derive the desired graphs from pure geometric information. Specifically, the challenges regarding the geometry and topology information of spaces in the IFC model can be categorized into two main problems: invalid geometry information in the IfcSpace object, which represents Spaces of areas or volumes that provide for certain functions within a building [19] and inaccurate space adjacency (topology) information, which are detailed in the following sections.

- *Illustrating invalid geometry information in the IfcSpace object*

To evaluate the energy consumption of different building spaces using data-driven approaches and the IFC model, we adopt the concept of space, which represents the idea of space [5], [14]. Spaces divide buildings into distinct parts and represent areas or volumes that are either physically or theoretically bounded. Therefore, it is crucial to accurately extract the geometric information of each space and its related entities, such as windows, walls, doors, roofs, and floors, as they significantly impact the energy load. This information enhances the analysis of a space's energy efficiency [16]. The primary geometric information of these entities includes the Window-Wall Ratio (WWR), Exterior-Wall Ratio (EWR), Interior-Wall Ratio (IWR), Floor-Space Ratio (FRR), Roof-Space Ratio (RRR), the total area of the space, and the total height of the space. spacespaceSpaceSpacespacespaceDespite the standard definitions and implementation guidelines for proper IFC models, incorrect geometry for IfcSpace objects is often expected. In the early stages of 3D indoor geometry information extraction, querying IfcSpace objects and their geometry information in an IFC model is typically sufficient to extract the geometry properties of each space's walls, windows, roofs, and floors. Figure 1.1 illustrates an example of high-quality semantic and geometric modeling of IfcSpaces, where walls are geometrically split for each space (indicated by yellow lines). A semantic-based query using the IfcOpenShell library can extract geometry information, such as wall area for each space, as the walls are split geometrically based on the boundaries of each space. However, most IFC models lack the secondary boundary definition (split geometry) for spaces and have inaccurate geometry calculations [17]. Figure 1.2a shows an IFC model with non-topological definitions of walls and windows for the corresponding spaces, depicting a shared wall (W143) between three spaces (S202, S201, and S203). In this case, the query returns an equal value of 3.84 m^2 for wall-143 for each space, totaling the entire area of wall-143 because it is not split based on space boundaries. Relying on such erroneous geometry values in the BECE process impacts the analysis of each space's energy consumption.

Additionally, windows play a crucial role in a building's energy gain or loss. They account for more than 10% of the building's energy load and significantly affect space energy consumption. If a space has an exterior, low-efficiency window, its energy loss increases substantially. Therefore, a space's energy consumption is directly influenced by the window's size, location, and insulation efficiency [18]. For instance, Figure 3.2b shows a shared window (WIN923) between spaces S155 and S156. Querying the existing IFC model returns an equal area value for window WIN923 for both spaces.

In contrast, the window size in S156 is significantly larger than in S155. Therefore, the impact of window 923 on the heating loss and gain in space 156 is more than its impact on space 155, which is not recognized if we rely on the current IFC query result. This leads to the necessity of geometric computation to extract accurate geometry information for space geometry information before using them in data-driven methods.

- Identifying inaccurate space adjacency (topology) information in the IFC model

In a 3D model, two spaces are considered adjacent if they share at least one wall, window, roof, or floor. In the context of BECE, the energy transferred between adjacent spaces impacts the energy demand of each space.

For example, if a space has a shared wall or floor (like the basement) with a cold space (with low energy efficiency), the heating loss rate increases significantly from the warm to the cold area. Therefore, extracting accurate adjacent information is essential. Nevertheless, the IFC semantic-based query cannot return the valid adjacent information because of each space's invalid (non-split geometry) wall, roof, window, and floor geometry. Figure 1.3 illustrates the instance of this error, which results in the wrong adjacency list.

Figure 3.4. describes a semantic-based query using the current information in the IFC model. It employs the "Bounded By" property of space objects to retrieve the IfcWall and IfcCurtainWall objects bounded to each space. By iterating through all space around the target space (B202 in Figure 1.3), the neighbor space objects are added to the adjacent list and considered

adjacent. The result shows the wrong adjacent spaces for the target space because there is a continued wall 143 between the target space, spaces A202 and B203.

The accurate adjacency information is important in BECE analysis, which means no geometrical relationship causes no energy transfer. However, in Figure 1.4, the IFC semantic query considers their adjacent spaces, which is not valid. Therefore, retrieving the topological information based on the semantic query from the IFC model is inaccurate and causes erroneous results when generating the BECE space-based graph.

1.2.2. Non-Topological Data-driven Models for BECE Analysis

Numerous studies review have been conducted to analyze and evaluate existing data-driven approaches. These studies contribute to advancing knowledge and informing the building industry, for example, categorizing building energy consumption prediction methods into several categories, including elaborate engineering methods, simplified engineering methods, statistical methods, Artificial Neural Network (ANN)-based methods [20], Support Vector Machine (SVM)-based methods [21], and grey models [14]. They conducted a comparative analysis considering various factors such as model complexity, ease of use, running speed, required inputs, and accuracy of the methods. Their analysis provided insights into the strengths and weaknesses of different approaches for energy consumption prediction in buildings [22]–[24].

Despite the promising accuracy of data-driven models, they often ignore the interconnected relationships between building elements, such as the topology of the spaces inside the building. This limitation leads to inaccuracies in BECE models. Most researchers rely on the existing data in IFC and extract just geometry and material information for BECE analysis. However, in BECE analysis, considering the shared elements (i.e., topology information) is essential because the amount of energy transferred from one space to another is directly related to the surface area of shared walls, windows, roofs, floors, and material properties. For instance, space A201 in Figure 1.3. A cold space on this floor significantly impacts the heating transfer to adjacent spaces. In principle, the heating transformation from space A204 to space A201 (from a warm to a cold area) is more than the heating transformation from A205 to A201 because of the differential in the shared wall area. Unfortunately, this information is not explicitly available in the IFC model. They must be derived from the available entities by advanced geometric processing and material isolation information to obtain the desired topology information and involve them in the learning process.

1.2.3. Lack of Explanatory in Data-Driven Models for BECE Analysis

On the other hand, despite accurate data-driven models such as ANN and SVM, they lack transparency and interpretability for BECE analysis. Indeed, the most accurate models, such as ANN and SVM models, are often considered black boxes, limiting their trustworthiness and applicability in critical applications that require a clear understanding of the underlying mechanisms behind estimation. To deploy data-driven models trustfully, it is necessary to provide both accurate predictions and human-intelligible feature explanations, especially for the AEC and facility management industry. These facts raise the need to develop a feature explanatory model for BECE analysis to explain the results of the model by calculating the role and weight of each (feature) in the result. Recent researchers in BECE analysis have often faced a trade-off between accuracy and interpretability when selecting a model [25]. Some advanced models, such as ANN, while accurate and flexible, can be challenging to interpret.

In contrast, simpler models like logistic regression or decision trees provide more uncomplicated explanations but may sacrifice some accuracy. While it is true that complex models offer greater flexibility and can achieve high levels of accuracy, it is important to acknowledge that simpler models also have their advantages. Simple models are often more straightforward to implement and interpret, and they can provide useful insights and solutions in scenarios where extreme precision is not necessary or practical. Additionally, complex models may require significant computational resources and expertise to develop and maintain, which could pose challenges in certain contexts. Therefore, while complex models undoubtedly offer advantages in terms of flexibility and accuracy, it is essential to carefully consider the trade-offs and select the most appropriate modeling approach based on the specific requirements and constraints of the project. However, there is a trade-off involved: these models may sacrifice interpretability. Despite their effectiveness, comprehending the rationale behind prediction models can be challenging. Even data scientists who developed and trained these intricate models may struggle to explain results, such as assigning a building space to an energy-efficient or inefficient class. Understanding the outcomes of data-driven models holds significant value for several reasons: (i) it fosters trust in the model, (ii) it offers transparency for building designers and decision-makers, and (iii) it facilitates the detection and rectification of model errors and patterns prior to real-world implementation. Furthermore, grasping network characteristics, including topology information, empowers designers to gain insights and make informed decisions [26].

1.2.4. Lack of Optimized Solution in Data-Driven Models

Developing energy-efficient buildings presents a complex task within the Architecture, Engineering, and Construction (AEC) industry. While Building Information Modeling (BIM) enables precise digital construction, optimizing designs either in the initial stages of a project or during retrofitting poses challenges [7], [27]. This difficulty stems from the necessity to account for numerous design variables and performance objectives. Despite the common use of building codes and engineering expertise in building design, they may not always yield the most energy-efficient outcomes. Moreover, while data-driven models can forecast energy consumption, they do not inherently provide design optimization solutions or recommendations for engineers [28]. This study addresses this gap by leveraging BIM Data and an optimization model to develop a prescriptive and explainable approach for decision-making and optimization in indoor building design.

1.3. RESEARCH OBJECTIVES

Considering the distinguished needs and gaps outlined in the preceding literature, the primary objective of this research is to develop a comprehensive framework called Space-based precise building energy consumption estimation using BIM. To achieve this overarching objective, the research is structured around several sub-objectives:

1. Develop an intermediate data model to generate a space-based graph based on the IFC model to tackle the compatibility issues between the IFC model and machine learning methods.

    1.1. Design an object-oriented model using a proposed framework that includes the building objects' effect on BECE analysis.

    1.2. Develop a geometry computation-based method to extract accurate geometry and space-relationship information from the IFC model.

    1.3. Establish an uncertainty analysis method to determine the reliability of the proposed model computation.

2. Design a graph-based machine learning algorithm based on Building Energy Consumption Estimation space-based graphs to consider detailed information about building objects and their adjacent information in the learning process.

    2.1. Design an optimized graph-based learning algorithm Graph Neural Network (GNN) based on a weighted knowledge graph.

    2.2. Train the proposed GNN algorithm based on different IFC models.

2.3. Establish a validation method for accurate assessment of the model.

3. Develop an explainable and prescriptive model to interpret the result of the graph-based prediction model so we can have a quantified-based interpretation for each critical space.

   3.1. Define an interpretable model to explain the result of the proposed graph-based classification model. The interpretation process should consider both spaces and the neighborhood's properties.

   3.2. Design a prescriptive model to use the result of interpretation models and suggest decision options to architectures and engineers using a BIM scale in perspective of geometry or material suggestions, considering the cost of each recommendation.

1.4. SUMMARY OF METHODOLOGY

In this section, we outline the methodology employed in conducting the research, detailing the approach to address the research questions and achieve the study's objectives. The detailed exposition of the methodology aims to achieve the main goals of the research for data-driven models for BECE analysis using the BIM data model. To achieve the three main research objectives, which are discussed in the previous section, the research methodology is divided into three parts (Figure 1.1).

1) A proposed framework to generate a space-based graph from BIM (IFC by extracting accurate semantic, geometry, and topology information from the space-level (space) IFC schema by introducing new geo-computation algorithms. Also, we define a new topological relationship between spaces in different stories by introducing two-dimensional vector values of common geometry area and energy resistance value (R-Value) between the spaces and a vector for each edge to represent the space's properties (node attribute). Eventually, the proposed space-based graph will have a graph-based data structure to decrease the original complexity of the IFC model

and be compatible with graph-based machine learning algorithms for BECE-based analysis.

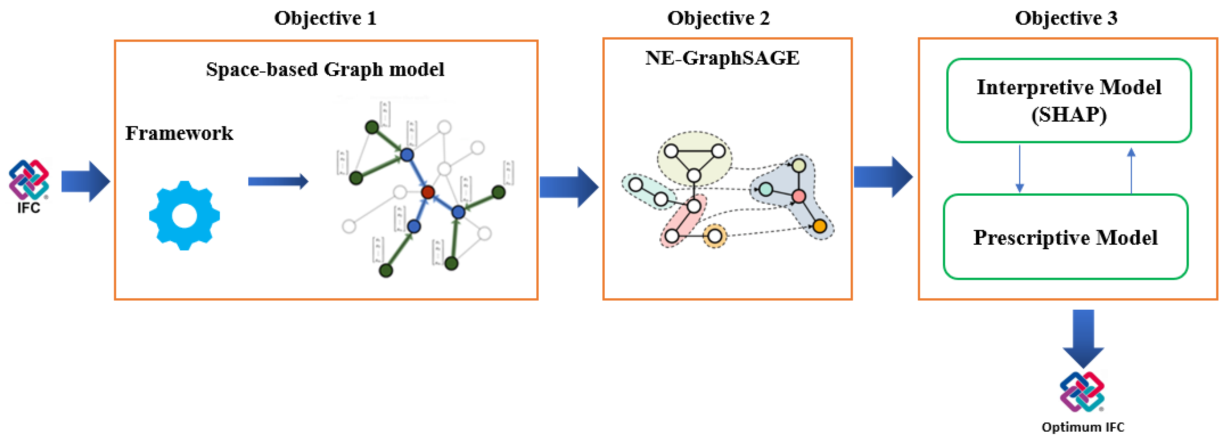


Figure 1.1. Three main steps in the research methodology.

2) The second part of the methodology employed the output of part 1. It proposed an edge node graph-based classification model to recognize the critical spaces in terms of energy consumption in the building. The proposed classification model tackles the issue in current research, which neglects the interconnected relationship of building elements. Also, current data-driven models lack interpretability and are often treated as black boxes. As a result, the models cannot be fully trusted for engineering without reasoning the underlying mechanisms behind the estimation. Besides that, the research proposed an explainable model by adopting the (SHapley Additive exPlanations (SHAP) with the proposed Node Edge-GraphSAGE classification model (NE-GraphSAGE) to make more transparency behind the data-driven model (Figure 1.1 – part 3). Preliminary results demonstrate the potential to improve pre- and post-construction steps by identifying critical spaces in buildings and identifying the parameters that affected the efficiency of the spaces with low energy consumption.

3) The third proposed method used the output of the second part (critical spaces) and the important parameters to design a prescriptive model for Architecture, Engineering, and Construction (AEC) engineers.

Nonetheless, these data-driven models often need more transparency and design optimization capabilities, rendering them insufficient for critical applications in the AEC industry. This research addresses the gap by integrating building energy simulation with optimization techniques, using BIM data and Genetic Algorithm (GA) to develop a prescriptive and explainable model for indoor building design. The research focuses on space-based indoor design, leveraging BIM interoperability to recommend optimal solutions (Figure 1.1 - Part 3). The proposed model employs the value engineering method to balance energy consumption, functionality, and cost, providing engineers and designers with insights to optimize building performance effectively. This approach enhances energy efficiency and offers tangible design optimization solutions, bridging the gap between energy prediction and practical application in the AEC industry.

1.5. CONTRIBUTION

To overcome the identified gaps and problem statement and fulfill the targeted objectives, this dissertation represents three main contributions:

- **Compatibility of 3D model (BIM) with BECE-based data-driven models** by introducing a space-based topological framework for assessing building energy consumption through BIM. The research novelty proposes an object-oriented data model that generates a space-based graph for building that is compatible with graph-based classification and prediction models encompassing 3D topology and relevant attribute information.

- **Improving the accuracy of space classification for BECE analysis by introducing NE-GraphSAGE and Making the model interpretable:** The novelty of this research lies in the proposal of a graph-based classification model that ranks spaces within a building in terms of energy consumption, utilizing a graph-based data structure derived from BIM. This model incorporates essential building parameters such as 3D geometry, material, and 3D topology information to identify critical spaces and suggests optimized, energy-efficient BIMs. Also, the result of the classification model is interpreted by the SHAP model to make it more transparent for engineers to analyze the impacted parameters in the classification result.
- **Designing a prescriptive model using a Genetic algorithm to help engineers optimize decision-making:** A prescriptive model is designed to propose optimal solutions based on building codes, employing a value-engineering method to assess the cost-effectiveness of these solutions. The proposed model in this contribution helps engineers make data-driven decisions from the model's output by exploring various solutions and scenarios based on diverse datasets.

1.6. DISSERTATION OUTLINE

This dissertation adopts a paper-based structure, comprising key chapters encompassing an introduction, literature review, specific objectives, methods, results, and discussions. Chapter 1 sets the stage with an overview of the research background, motivation, and primary contributions. Chapter 2 delves into the intricacies of the research problem statement. In Chapter 3, an innovative process is introduced, automating the generation of a space-based graph by extracting topological and semantical information from IFC for Building Energy Consumption Estimation (BECE) analysis. This process is validated using case study data from Autodesk buildings, corroborated with ground truth data.

Moving forward, Chapter 4 explores the potential of novel graph-based learning algorithms, specifically GraphSAGE, to leverage enriched semantic, geometry, and space topology information from BIM data for identifying critical spaces in buildings from an energy consumption perspective. This chapter introduces transparency in data-driven models through an explainable GraphSAGE model, employing SHAP and the proposed Node-Edge GraphSAGE (NE-GraphSAGE) prediction model. Chapter 5 outlines a method that integrates building energy simulation with optimization techniques, utilizing BIM data and Genetic Algorithm (GA) to develop a prescriptive and explainable model for indoor building design. Focusing on space-based indoor design, this model leverages BIM interoperability to recommend optimal solutions, balancing energy consumption, functionality, and cost using the value engineering method. Finally, Chapter 6 consolidates significant findings, discusses model limitations, underscores novel contributions, and offers recommendations for future research. The dissertation ended with a Bibliography and Appendices.

CHAPTER 2: LITERATURE REVIEW

2.1. PREFACE

A comprehensive and critical literature review has been conducted, and the current state of the art has been identified and outlined in this chapter as a groundwork for further investigation. Through an exhaustive examination of existing scholarship, this review aims to contextualize the research problem, highlight key theories, methodologies, and findings, and identify gaps in the current knowledge. By synthesizing diverse perspectives and empirical evidence, this literature review provides an understanding of the complexities inherent in the research. It serves as a solid foundation upon which subsequent chapters will build, offering insights that contribute to the advancement of the BECE. The research employed a systematic literature review on building energy consumption, which is structured to cover various aspects, including data-driven models and physical models, as well as their applications across commercial, industrial, and residential buildings. The review begins with a comprehensive search strategy, utilizing keywords such as "building energy consumption," "data-driven models," "physical models," "commercial buildings," "industrial buildings," "residential buildings," and "Building Information Modeling (BIM)." This is followed by a thorough screening process to select relevant studies based on predefined inclusion and exclusion criteria. Subsequently, the selected studies are subjected to detailed data extraction and synthesis, where key findings, methodologies, and technological advancements are systematically analyzed. Special emphasis is placed on understanding how data-driven and physical models are implemented and the specific benefits and challenges associated with each model type. The role of BIM in enhancing energy efficiency through integrated and optimized energy performance is also critically evaluated.

The review identifies significant trends and emerging technologies in building energy modeling, as well as common challenges and limitations reported in the literature. By synthesizing these insights, the review narrows down to identify research gaps, which are discussed in this chapter. These gaps highlight areas for future research, guiding researchers and practitioners toward developing more effective and efficient energy modeling techniques to enhance building performance across various sectors.

2.2. BUILDING INFORMATION MODELLING (BIM)

Building Information Modeling (BIM) has revolutionized the landscape of building construction, leading to a new era of efficiency, collaboration, and precision [1]. At its core, BIM is a digital representation of the physical and functional characteristics of a building, providing a comprehensive and dynamic 3D model that integrates various data sources. This paradigm shift in the Construction, Engineering, and Architecture (AEC) industry has fundamentally altered how architects, engineers, contractors, and other stakeholders approach the entire lifecycle of a building project [29].

In the context of building construction, BIM's impact is particularly pronounced. The technology allows for the creation of intelligent models that visualize the structure and incorporate detailed information about materials, components, and systems [29], [30]. This depth of information enables stakeholders to conduct detailed analyses and simulations, fostering a proactive approach to addressing potential issues before they manifest on the construction site [31]–[33]. From clash detection that identifies and resolves spatial conflicts to quantity takeoffs for precise material estimates, BIM streamlines processes and enhances decision-making, ultimately leading to more cost-effective and sustainable construction practices [29], [33], [34].

In North America, the adoption of BIM in building construction has been catalyzed by industry standards that recognize its transformative potential. Organizations such as the National Institute of Building Sciences (NIBS) and the National BIM Standard-United States® (NBIMS-US™) have played pivotal roles in developing and promoting BIM standards [35]. These standards provide a framework for consistent data exchange, collaboration protocols, and model organization. Adhering to these standards ensures interoperability among various stakeholders, allowing for seamless communication and integration of BIM data throughout the project lifecycle [8]. As the industry continues to evolve, incorporating BIM in building construction, guided by these standards, not only enhances project efficiency but also contributes to the advancement of construction practices and the realization of more resilient and sustainable built environments [8], [36].

2.3. BIM USE CASES

BIM is an enriched data infrastructure that involves creating and managing digital representations of a building or infrastructure's physical and functional characteristics. Here are some critical BIM technical utilization and their descriptions:

- **Parametric Modeling**: Parametric modeling allows users to create intelligent 3D models with elements that have specific parameters and relationships. Changes made to one element automatically update related elements, ensuring consistency throughout the model [32], [37].
- **Interoperability:** BIM supports interoperability, enabling the exchange of information between different platforms and disciplines. This facilitates collaboration among various stakeholders in building design, construction, and operation [38].
- **Data Integration:** BIM allows the integration of various data types, such as geometric, spatial, and non-graphical information. This integration enhances the accuracy and completeness of the model, supporting better decision-making at every stage of a project [39].
- **Collaboration Tools:** BIM platforms provide tools for real-time collaboration among project team members. This includes features like cloud-based sharing, version control, and multiple-user access to the model, promoting efficient teamwork [40].
- nD BIM: nD BIM integrates time, cost, sustainability, and other dimensions beyond traditional 3D modeling, enhancing project management and decision-making processes in construction projects. For example, 4D BIM incorporates time, enabling the visualization of construction sequences and project timelines. 5D BIM adds cost information, allowing stakeholders to visualize and analyze the cost implications of design decisions and construction schedules [31], [41].
- **Quantity Takeoff:** BIM facilitates automated quantity takeoff, allowing users to extract accurate and detailed information about the quantities of materials in the model. This supports cost estimation and procurement processes.
- **Asset Management:** BIM extends beyond the construction phase, offering features for managing building assets throughout their life cycle. This includes maintenance schedules, facility management, and documentation of changes over time [42].

- **Energy Analysis:** BIM often includes energy analysis capabilities, allowing users to assess and optimize the energy performance of a building. This supports sustainable design practices and compliance with energy efficiency standards [25], [43].
- **Visualization and Rendering**: BIM offers advanced visualization and rendering features, allowing users to create realistic 3D visualizations of the building, which aids in communication with clients, stakeholders, and the public [44].
- **Geospatial Integration**: BIM can be integrated with geographic information systems (GIS), enabling the incorporation of location-based applications. This is particularly valuable for projects that require land topography, environmental conditions, and infrastructure consideration [45], [46].

2.4. BUILDING ENERGY MODELLING (BEM)

Building Energy Modeling (BEM) is a computational process that simulates and analyzes building energy consumption estimation (BECE) and energy performance [1]. It plays a vital role in designing and operating energy-efficient buildings by providing insights into the building's energy consumption, thermal comfort, and environmental impact [1], [47]. BEM involves the creation of a virtual representation of a building, incorporating details such as its geometry, construction materials, HVAC (Heating, Ventilation, and Air Conditioning) systems, and occupancy patterns [48]. This virtual model allows engineers, architects, and energy professionals to assess the building's energy usage under different scenarios and optimize its design for energy efficiency [1], [49].

BEM utilizes various estimation techniques to predict the building's energy consumption. These estimations consider factors such as weather conditions, solar radiation, internal heat gains, and the performance of building systems [15], [49]. The results obtained from BEM can guide decision-making processes related to building design, retrofitting, and operational strategies [1], [15]. As energy efficiency becomes a key focus in the construction industry, BEM has gained prominence as an indispensable process for achieving sustainable and green building standards [10], [50], [51].

In the context of BEM, physical and data-driven models are used to simulate and estimate the energy performance of buildings. These two models offer different approaches to capturing and predicting the complex interactions within a building's environment [10], [51].

2.4.1. Physical Models

Physical models, also known as engineering methods or white-box models, utilize thermodynamic principles for comprehensive energy modeling and analysis [14]. They are based on the fundamental principles of physics and engineering that go into a building's energy transfer, airflow, and other thermodynamic procedures. These models simulate the behavior of individual building components and systems to predict overall energy consumption. In this research, we employed a data-driven model. Key characteristics of physical models in BEM include [1], [14], [52]:

- **Physics-Based Simulation**: Physical models use mathematical equations to simulate the physical processes occurring within a building, considering factors such as thermal conductivity, heat transfer, and system efficiencies [10], [52].
- **Detailed Representation**: These models provide a detailed representation of the building's geometry, construction materials, and HVAC systems. They aim to capture the building's structure and systems' intricacies accurately [52].

- **Engineering Principles**: Physical models rely on established engineering principles, making them suitable for understanding the underlying physics of energy transfer and thermal comfort in buildings [14].
- **Challenges**: Challenges associated with physical models include the need for accurate input parameters, simplifications made during the modeling process, and potential difficulties in representing dynamic and complex interactions within the building. Indeed, they suffer from two limitations: 1. The number of input parameters needs to be considered; therefore, they are limited to employing datasets from new technologies, such as sensor-based datasets or 3D information[52]. 2. They are not intended to incorporate geometry information, a crucial parameter for estimating energy consumption, which can result in inaccurate energy modeling [53].

2.4.2. Data-Driven Models

Data-driven models (a black box) in BEM leverage machine learning algorithms and statistical techniques to learn patterns and relationships from empirical data. Instead of relying on explicit knowledge of the physical principles governing a building's behavior, these models use historical data to make predictions or simulations [20]. Key characteristics of data-driven models in BEM include [10], [54], [55]:

- **Learning from Data:** Data-driven models learn from datasets, incorporating information about past building performance, occupancy patterns, weather conditions, building properties, and other relevant factors [54].
- **Flexibility and Adaptability**: Data-driven models are often more flexible and adaptable to changing conditions and input features. They can capture complex relationships and

nonlinearities that may be challenging for traditional physics-based models to represent accurately [24].

- **Prediction and Optimization**: These models can predict energy consumption, optimize control strategies, and identify opportunities for energy efficiency improvements based on observed patterns in the data [56].
- **Challenges**: Challenges associated with data-driven models include the need for extensive and representative datasets, potential biases in the data, and the "black-box" nature of some complex machine learning algorithms, making it challenging to interpret the underlying relationships [26].

Our research focuses on the second type (data-driven) of modeling, and physical modeling is beyond the scope of this research. Data-driven approaches, like machine learning algorithms, predict building energy consumption based on a set of features as their input parameters. Therefore, the following section discusses the major features (input features) of building energy consumption based on previous research.

2.4.2.1. Identifying and analyzing the main features of building energy consumption

The current state of the art shows different parameters for estimating building energy consumption. We can summarize them as outdoor weather conditions, indoor environmental conditions, building geometric information, building material information, time of the day, number of occupancies, occupant energy use behavior, and mechanical and electrical information [14], [54], [56]–[61]. Outdoor environment features include outdoor temperature, relative humidity, solar radiation, wind speed, wind direction, pressure, rainfall amount, and evaporation [10], [62]. Indoor environmental features include space temperature, relative humidity, indoor material isolation, HVAC system, and lighting level [63]. Building characteristic features include relative compactness [64], surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution, heat transfer coefficient of building walls (R-Value), roof heat transfer coefficient (R-Value), building size coefficient, the absorption coefficient for solar radiation of exterior walls, window-wall ratio (WWR), shading coefficient (SC) [54]. Time features include the type of day (e.g., weekday, weekend, holiday) and hour (e.g., daytime, nighttime). Occupant energy use behavior, and occupancy features include building use schedule, heat gain through lights and people, water temperature, and the number of occupants [26], [54], [65].

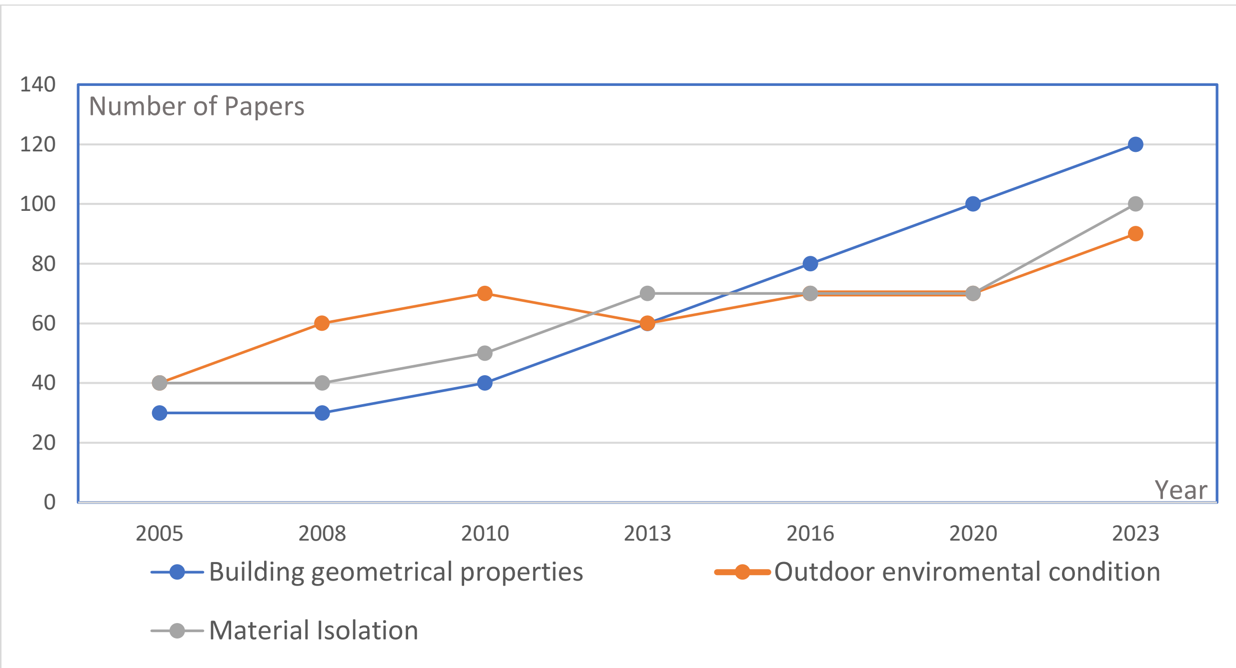


Figure 2.1.Trend of employing the building energy consumption features from 2005 to 2023 [26], [54], [65]

Figure 2.1 summarizes the diversity and trend of using features, which is listed in Appendix E, from 2005 to 2023 based on previous studies for building energy consumption prediction. There is a slight steady outdoor environmental information (orange line) trend as one of the important energy consumption parameters over fifteen years. The reason for the steady trend can be the availability of devices to measure and record weather temperature (thermometer) and humidity (psychrometer) data over time (such as the sensor) and the availability of open datasets [67]. Also, using detailed building material information (gray line), including wall, window, roof, and floor heating transfer resistance gradually increased from 2005 to 2023 because of the availability of detailed building information in building documents (specification document) and BIM in time of design and construction [9], [68], [69]. On the other hand, the number of researchers who use detailed building geometry information (blue line) has an upward trend, especially after 2010, because of the availability of data on BIM and 3D city models to enhance the detailed building geometry information [1][70]. The trend represents the demand for BIM and 3D city data (outdoor information), which increased and led to data-driven models to predict building energy consumption using detailed geometry and semantic information. Also, employing more detailed information regarding geometry and semantic information improves the prediction models and decreases the use of high-volume historical data (such as outdoor weather data) [14]. Also, with the advent of the BIM data model (IFC), the importance of the other type of feature is enhanced [16]. The relationship between the entities in 3D space can be essential information rarely considered in previous research because of a lack of entity-relationship information in traditional data models. For example, Eicker et al. show the impact of buildings in the neighborhood on each other [13]. They investigate the effect of urban compactness, which is calculated by the distance of buildings in the neighborhoods, and the building shading effect on the neighborhood, which is calculated by the building roof structure, distance, and angle of building in the neighborhood. Their simulation includes neighborhood shading calculations for each segment of the façades and roofs, allowing a detailed quantification of the building energy consumption [13].

2.4.2.2. Variety in Data Structure in Different Data Sources

Most data-driven models, like machine learning algorithms, predict building energy consumption based on a set of features as input parameters [24], [71]. Also, new data models like

IFC are increasingly pushing detailed information into data-driven approaches, generating and using more features, resulting in more accuracy. Although building energy estimation models need detailed geometry information, they need to be connected to other data sources such as outdoor weather information, indoor environmental information, and tenants' behavior [72]. Therefore, purely geometric information in the model does not satisfy the required data for prediction analysis [9]. Hence, connecting different data sources to the IFC model to predict energy consumption is necessary. Since there are various features with different natures and data structures from different data sources, we need to deal with the challenge with different structures (Figure 2.2). According to the current studies, we can summarize that the most common structured or non-structured data such as PDF format as non-structured data (65%), tabular data structure like CSV format (85%), the relational data structure in the standard databases (33%), shapefiles as a spatial-relational data structure (25%), object-oriented data structure like BIM (10%). The high percentage of users using PDF and CSV implies that most information related to buildings and energy consumption is stored in these data structures [73], [74]. The low percentage of BIM use indicates less knowledge of using these data structures and the complexity of the BIM [9]. Therefore, inconsistency in the data structure corresponding to different data sources is another challenge when using BIM in BEM.

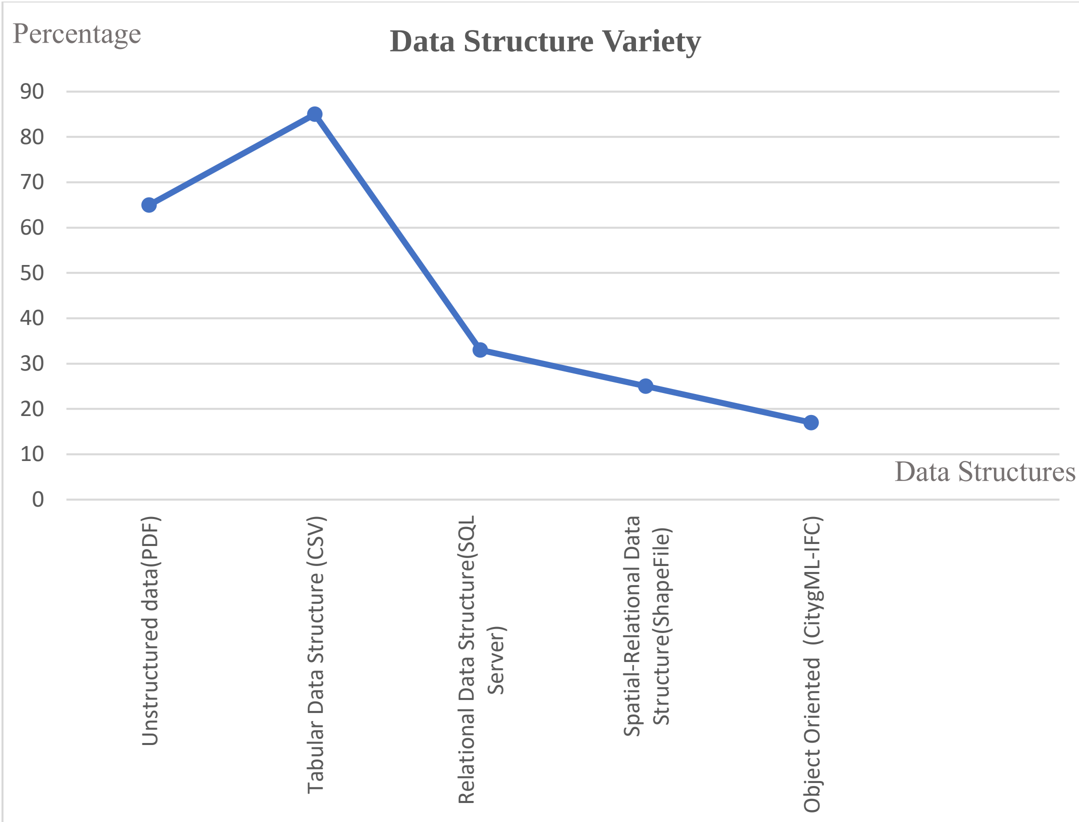


Figure 2.2. Variety of Data Structure in Different Data Sources

2.4.3. Data-driven models for BECE

The distribution of the studies by different types of machine learning algorithms is illustrated in Figure 2.3 based on the Web of Science database [9], [10], [14], [22], [23], [25], [35], [52], [54]–[57], [63], [67]–[86], [87]–[93], [94]–[103], [104]–[108]. 34% and 24% of the studies utilized ANN and SVM, respectively, and 8% DNN (Deep Neural Network) to train their models. Only 14% of the studies applied decision trees. On the other hand, 20% of the studies utilized other statistical algorithms. The study demonstrates that ANN and SVM perform better than other models in accuracy. Also, DNN performs well but is not as promising as others because it requires an enormous amount of data for training [14], [36], [37]. There are two main reasons for this result.

First, SVM and ANN are popular and efficient nonlinear problem-solving techniques suitable for energy efficiency estimation as nonlinear phenomena [38].

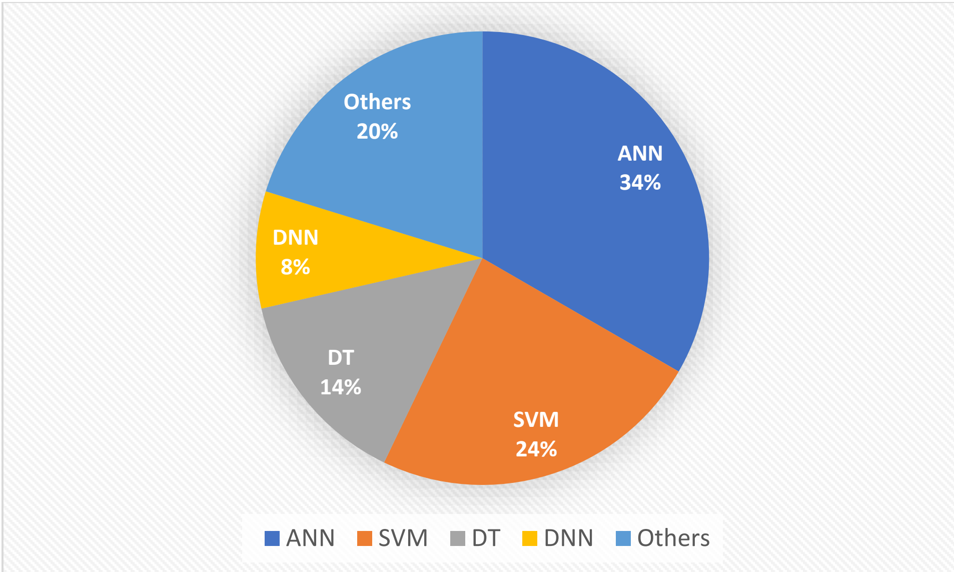


Figure 2.3. The usage percentage of data-driven algorithms was analyzed in 50 research papers

The second reason indicates that these data-driven models are more effective in high-dimensional and varied features. Despite the advantage of both algorithms in modeling nonlinearity, ANN has an advantage over SVM in terms of applying each feature's influence in the learning process. In ANN algorithms, the weight of each feature will be determined after several iterations; therefore, the more important features have more influence on the result of prediction, but SVM considers the equal importance for all input features [22]. The detailed comparative information about ANN, SVM, and statistical models is listed in Figure 2.3 [9], [10], [14], [22], [23], [25], [35], [52], [54]–[57], [63], [67]–[86], [87]–[93], [94]–[103], [104]–[108].

2.5. BIM FOR BUILDING ENERGY MODELING

In the conventional design process (physical and data-driven models), Building Energy Modeling (BEM) is usually introduced only in the final design phase due to the limited input variables available in earlier stages. However, this sequential approach often results in conflicts between building design choices and BEM analysis, especially when modifications to the design are no longer feasible [22], [116]. Consequently, the simulation results obtained after substantial investments in time and resources may prove ineffective if they cannot influence the design decisions due to late integration into the process. This highlights a potential inefficiency in the traditional workflow, emphasizing the importance of integrating BEM considerations earlier in the design stages to facilitate more informed and impactful decision-making [117].

BEM based on BIM addresses the limitations inherent in traditional BEM approaches and facilitates the seamless integration of Building Energy Modeling with digital design. A recent advancement in this field is the emergence of Building Information Modeling-based Building Energy Modeling (BIM-based BEM). This approach leverages the foundation of BIM, utilizing pre-designed BIM models to generate input for BEM models. The result is a more efficient, accurate, and consistent process that offers advantages such as cost-effectiveness, time-saving benefits, and practicality [1].

According to Bazjanac et al. [118], incorporating specific guidelines within BIM-based BEM is a preventive measure against arbitrary modifications to the data. This ensures the accuracy and consistency of the energy model across various users.

In summary, integrating BIM and BEM enhances the accuracy and efficiency of building energy analysis by providing a comprehensive and detailed digital representation of the building and its components throughout its lifecycle. This integration supports a more holistic approach to sustainable building design and operation.

2.6. CHALLENGES IN IMPLEMENTING BIM AND BEM

While integrating Building Information Modeling (BIM) and Building Energy Modeling (BEM) offers significant benefits, some challenges exist in their seamless collaboration.

2.6.1. Compatibility

In a BIM-based BEM, compatibility between the BIM data model and BEM models, primarily data-driven models such as machine learning algorithms, is essential, enabling two separate models to communicate and exchange data. The benefit of smooth data transfer lessens the likelihood of duplicate data generation or missing data in analytical models. It ensures the integration of sustainable features at the early design stage [24], [49]. Current states of the arts demonstrate the role of detailed geometrical and semantical information in predicting building energy consumption in both physical and data-driven approaches. The literature review also illustrates the value of 3D information in providing detailed building information useful for data-driven approaches. Data-driven approaches such as BEM models need semantic and detailed geometric information to predict energy consumption accurately. However, there is a challenge to integrate BIM data, which has an object-oriented data structure, into different data sources with various data structures. Nowadays, most of the research connects the other data sources to BIM information by converting IFC to a tabular (flat) data model [119], [120]. Also, machine learning algorithms are compatible with tabular data structures [121], against the fact that objects in the BIM contain many types of entity-relationship information. They include topological information such as inside, above, or touching information of building entities. This conversion deducts or eliminates the relationship between the entities, which are primary information, to accurate prediction methods [43]. Therefore, utilizing an efficient data structure to have the following three characteristics is necessary [9]:

- Capability to connect to different data sources.
- The capability of storing geometric, semantic, and topology information and
- Compatible with the machine learning algorithm.

As discussed, BIM includes the indoor environment's primary semantic, geometry, and topology information, which are of considerable interest for data-driven methods, especially by describing the indoor space concept in the BIM and defining spatial links between them, as well as the potential for BEM. Indeed, creating a space-based graph to analyze their energy transfer is possible. In this graph, the nodes denote spaces in which the property information of each space (node) is defined as vector information assigned to the node. An edge in the graph connects a pair of spaces and captures the spatial relationship of spaces if there is any energy transfer [9]. Also, graphs have started to play a central role in machine learning, incorporating real-world objects with a relationship for knowledge extraction and phenomena prediction [122]. Graph learning can be employed in BEM because of its compatibility with different data sources at the data level and compatibility with data-driven models at the algorithm level. In broad terms, graph learning pertains to the application of machine learning techniques on graphs, where methods are employed to map the graph's features into feature vectors of equivalent dimensions within an embedding space [123]. Unlike traditional approaches that project graphs into lower-dimensional spaces, graph learning models or algorithms directly transform graph data into output within the graph learning architecture. The majority of graph learning methods draw from or extend deep learning techniques, as these methods proficiently encode and represent graph data into vectors, resulting in output vectors existing within continuous space [24]. The primary objective of graph learning is to extract the desired features of a graph, enabling the representation of a graph to be readily utilized by downstream tasks such as node classification and link prediction without necessitating an explicit embedding process [122]. Consequently, graph learning emerges as a more potent and meaningful approach to graph analysis.

Pezeshki et al. [124] demonstrate that the absence of compatibility between BIM and BEM models poses challenges in realizing sustainable and energy-efficient projects across their life cycle. These projects frequently encounter issues such as incomplete or inaccurate modeling of geometry information and a lack of sufficient material information.

The current compatibility challenges between BIM and BEM are recognized in the industry. However, there is an expectation that these challenges will be addressed in future developments. Integrating BIM and BEM represents a relatively new area of research, and ongoing advancements are anticipated.

2.6.2. BIM Model Complexity

BIM models can be highly detailed, containing vast information for BEM models. Translating this complexity into a format suitable for energy simulations in BEM can be challenging and time-consuming. The geometric representation of a building in BIM includes details about its shape, size, orientation, and 3D spatial relationships (3D topology) between various components. Complex architectural designs, irregular shapes, and intricate building geometries can pose challenges in accurately capturing these details. Simplifying the BIM model for BECE analysis while retaining key information is essential, which our research addresses. Establishing a data model for model simplification and automation can help streamline the process [9], [24], [125].

CHAPTER 3: AN AUTOMATED SPACE-BASED GRAPH GENERATION FRAMEWORK FOR BUILDING ENERGY CONSUMPTION ESTIMATION

3.1. PREFACE

This chapter focuses on the first sub-objective, proposing a framework to generate a space-based graph from the IFC model to decrease the original complexity of the IFC model and is compatible with graph-based machine learning algorithms. The geometry and assigned extracted attributes (features) are tested and investigated by two sample buildings as case study models. The content of this chapter was published in Buildings in 2023 as follows:

H. Kiavarz, M. Jadidi, A. Rajabifard, and G. Sohn, "An Automated Space-Based Graph Generation Framework for Building Energy Consumption Estimation," *Buildings*, vol. 13, no. 2, p. 350, Jan. 2023.

Copyright and Licensing:

Contributions:

The contributions of authors in the current chapter are as follows: **Hamid Kiavarz**: has conducted the literature review, performed the data collection and curation, developed the method, used the required software to perform the analysis and modeling, validated and visualized the results, and wrote the original manuscript of this publication. **Mojgan Jadidi** has supervised the research, provided the funding, and contributed to developing the manuscript's method, writing, and editing. **Abbas Rajabifard and Gunho Sohn** have supervised the research and contributed to editing the manuscript.

3.2. ABSTRACT

The 3D information in Building Information Modeling (BIM) has received significant interest for innovative city applications. Recently, employing Industry Foundation Classes (IFC) as an open standard data model for BIM in data-driven methods for Building Energy Consumption Estimation (BECE) has gained momentum because of the enriched geometric and semantic information. Indeed, BECE notably depends on the semantics (building elements' properties information), geometry (building elements and shapes), and topology information of the building's elements to recognize the spaces in a building with high energy demand. However, despite extensive studies on applying the IFC data in BECE analysis, employing the full potential of the BIM remains poor due to its data model complexity and incompatibility with data-driven algorithms, namely machine learning algorithms. This paper proposes a framework to extract accurate semantic, geometry, and topology information from the space level of the IFC schema by introducing new geo-computation algorithms to address these challenges. Also, we define a new topological relationship between spaces in different stories by combining common geometry area and energy resistance value. The space's adjacency weighted matrix is generated, determining the relationship between the spaces and their weighted value. Eventually, the proposed weighted space-based graph, which has a graph-based data structure to decrease the original complexity of the IFC model, will be constructed and will be compatible with graph-based machine learning algorithms for BECE-based analysis. Obtained promising results with more than 90 percent accuracy for extracting the geometry information for the convex and non-convex polyhedron spaces and 100% accuracy in detecting vertical and horizontal adjacent spaces, resulting in an accurate space-based graph with trustworthy feature values. This study confirms the proposed approach's efficiency, effectiveness, and accuracy when investigating space-based BECE analysis. The results support the feasibility of using the proposed solution to employ BIM in data-driven analysis.

3.3. INTRODUCTION

3D information has dominated the world of building modeling using Building Information Modeling (BIM) for more than a decade [1]. BIM includes the 3D representation of the physical building elements, their properties (semantics), and the relationships of the elements [2]. A series of standards are introduced to represent the BIM data model; however, the most well-known BIM open standard is the Industry Foundation Classes (IFC) released by the International Alliance for Interoperability (IAI), now known as BuildingSMART Alliance (BSA) [8]. The IFC model is an object-oriented data model incorporating 3D geometry, spatial relationships (topology), quantities, and building element properties (semantic). IFC facilitates the interoperability problems in the different disciplines and shares almost all the building components in the construction and building management domain [32], [126]. Different disciplines, such as facility management and energy efficiency studies, need accurate geometry, semantics, and topology as preliminary information to lead to reliable and accurate results. Furthermore, the 3D information in IFC models has received significant interest in today's application domain in smart cities and smart buildings [127]. Recently, IFC as a 3D data source for Building Energy Consumption Estimation (BECE) has gained momentum, according to a study by [128].

Indeed, BECE depends on the geometry, topology, and material information of the building's elements to recognize the spaces in a building with high energy demand [10], [12], which leads to evaluating different building design alternatives and operation strategies. Physical modeling and data-driven approaches are widely used in BECE [10]. The physical models (engineering methods or white-box models) rely on thermodynamic rules for detailed energy modeling and analysis. These models calculate energy consumption based on detailed building and environmental parameters such as construction details, operation schedules, and HVAC design [12]. On the other hand, data-driven methods estimate building energy consumption based on learning from historical data and building properties [10], [129]. However, two shortcomings limit the BECE physical models: first, the high number of required input physical parameters that need to be measured by deploying sensors, and they are not designed to use geometry information, one of the influential parameters to estimate energy consumption [13], [14].

On the other hand, data-driven methods can accept BECE-based analysis hyperparameters, making them flexible in using various data sources in their process [126]. Despite extensive studies on data-driven methods, employing the IFC model remains poor due to its complexity and incompatibility with data-driven methods like machine learning algorithms. However, the IFC model includes the indoor environment's primary semantic, geometry, and topology information, which are of considerable interest for data-driven methods. Especially by describing the indoor space concept (known as the IfcSpace class) in the IFC standard model and defining spatial links between them, the potential for BECE analysis in different parts of the building is opened. Indeed, creating a space-based graph to analyze their energy transfer is possible. In this graph, the nodes denote spaces in which the property information of each space (node) is defined as vector information assigned to the node. An edge in the graph connects a pair of spaces and captures the spatial relationship of spaces if there is any energy transfer [130]. Also, graphs have been central in machine learning, incorporating real-world objects with a relationship for knowledge extraction and phenomena prediction [131]. BECE analysis based on machine learning algorithms in space scale depends on indoor geometry, semantic and relationship parameters such as relative compactness, the overall height of space, orientation, glazing area distribution, heat transfer coefficient of building walls (R-Value), roof heat transfer coefficient (R-Value, the absorption coefficient for solar radiation of exterior walls, window-wall ratio (WWR), shading coefficient (SC) [10], [54]. Unfortunately, the spaces (IfcSpace) in most existing IFC models have inaccurate geometric information and space-based topological descriptions for BECE-based analysis. Therefore, the mentioned feature information is not explicitly available in the IFC model and needs to be computed from a combination of semantic and geometry information. Furthermore, potential inconsistencies between geometric and semantic information can lead to inaccurate results when using the IFC dataset in machine learning algorithms. Therefore, a study is needed to bridge this gap, consider existing models' limitations, and propose a solution for the IFC data model's compatibility with machine learning algorithms.

This paper proposes a framework to generate an object-oriented data model as an intermediate data model to translate the IFC model to the BECE space-based graph, including BECE-related information and spatial relationships between IFC spaces. The proposed framework extracts and calculates implicit BECE-related information by introducing new objects, such as a wall's inner and outer face objects, roof, window, floor, and related geometry information. At the same time, the framework introduced accurate topological information, such as the spatial relationship between spaces horizontally and vertically. Furthermore, in contrast to other studies that consider the binary relationship between the IFC elements, the framework calculates the weighted relationship using energy transfer between the spaces. Undeniably, the advantage of using the proposed intermediate object-oriented data model is the possibility of extracting implicit BECE-related information from IFC by preserving the spatial relationship (topology information) between the spaces. Thus, the framework extracts the complex IFC data model, including the buildings' meaningful semantics for BECE, geometry, and spatial relationship, into a BECE-based graph. The main benefits of the proposed methodology are: (1) to enable building space objects in IFC to be automatically translated into a BECE-based data model; (2) to enhance the hidden values in the original IFC to make it usable in building energy estimation without human interference; and (3) to develop a more reliable and accurate BECE space-based graph from IFC models to be compatible with machine learning algorithms.

The paper's organization is as follows: A current state of the art is presented in Section 3.4, and the challenges and shortcomings are identified in Section 3.5. The proposed methodology, which allows us to consider the indoor 3D space and its topology information, is discussed in Section 3.6. Data sources, case studies, and results are in Section 3.7. Finally, the lesson learned, concluding remarks, and future research direction are described in Section 3.8.

3.4. RELATED WORK

Data-driven methods recently benefit from developments in machine learning, such as providing flexibility and reliability in modeling and forecasting building energy consumption [54], [89]. Therefore, many studies have been conducted on the semantics and geometry data exchange between BIM and BECE using standard data schemas (IFC). Briefly, three types of solutions can be distinguished [43], [132]–[134].

- Type I: Use an integration solution for integrating a BECE simulation engine into the BIM tool, such as Revit, or use the Application Programming Interface (API) in BIM-based software.
- Type II: Export the relevant information from the BIM to a file using the gbXML format and use it in BECE software for simulation [51], [135], [136].
- Type III: Export the entire BIM, such as Revit format, to the IFC format and import the IFC format into BECE tools [134], [135].

Jeong et al. (2014) applied type one in their simulation using REVIT API and the Modelica Building library. The incompatibility between the two environments and the dependence on a specific version of proprietary software constitute the main drawback of this approach. Cemesova, Hopfe, and McLeod (2015) employ the second type in their analysis and use commercial software to convert the BIM format to the gbXML data model. Although this model preserves the topology information of BIM data, it has limited geometry definition capabilities (e.g., preserving only rectangular shapes based on element centerline). Therefore, it can support convex polyhedron shapes and ignore non-convex polyhedrons, leading to a non-accurate estimation of the building energy consumption [137]. Most parts of the third type of solution focus on geometry representation despite enriched information from the IFC schema, like the face geometry properties and space topological information. All of them are considered implicit in the Type III solution.

Consequently, the IFC model's complexity and incompatibility with data-driven methods restrict its use in BECE data-driven methods. Because of the mentioned challenges of using the IFC model in BECE data-driven methods, most of the published methodologies solely focus on one specific part of BIM data (e.g., geometry or HVAC) and use traditional physical simulation engines (e.g., EnergyPlus) [18]. Therefore, utilizing an efficient data model to have the following three characteristics is necessary: 1) the capability of the data model to be connected to different data sources; 2) the capability of the data model to preserve geometric, semantic, and relationship information; and 3) compatibility of the data model with the machine learning algorithm.

3.5. CHALLENGES IN EXTRACTING THE BECE-BASED GEOMETRY AND TOPOLOGY INFORMATION FROM THE IFC MODEL

The current state of the art in estimating building energy consumption recognized the beneficial use of IFC models for extracting the building semantics, geometry information, and 3D spatial relationship between the spaces [27], [120], [138]. Indeed, efficient 3D analysis of building energy consumption relies on the indoor environment's semantic information, the geometric description of the IFC entities, and their spatial relationships. The geometry and material information of the entities such as walls, windows, roof, and floor are the main entities of each space in IFC, affecting the space's heat gain and loss. Therefore, taking the precise geometry and space's properties, such as material information of these entities for each space, is necessary. Most IFC-based building energy modeling relies on the information stored in the model. However, incomplete or erroneous implementation of the IFC and implicit topological information, typically found in many IFC export modules, cause inaccurate geometry information, such as each space's wall and window area. Also, recognizing accurate topological information, such as the spatial relationship between the spaces, is hard to extract from the IFC models. These errors generate an invalid graph in the perspective of extracting BECE-based information and wrong space spatial relationships. In this case, advanced geometric processing is necessary to obtain the desired graphs from pure geometric information. In detail, the challenges mentioned above regarding the spaces' geometry and topology information in the IFC model are categorized as two main problems: invalid geometry information in the IfcSpace object and inaccurate space adjacency (topology) information described in the following sections.

3.5.1. Illustrating Invalid Geometry Information in the IfcSpace Object

To evaluate the energy consumption of the different spaces of the buildings using data-driven approaches and the IFC model, we use the concept of space representing the concept of space [18]. Because space splits the buildings into different parts and represents an area or volume bounded actually or theoretically. Therefore, the geometric information of each space and its related entities, such as windows, walls, doors, roofs, and floors, which significantly influence the energy load, should be extracted accurately to improve a space's energy efficiency analysis [80]. The primary geometry information of these entities is listed as the Window-Wall Ratio (WWR), Exterior-Wall Ratio (EWR), Interior-Wall Ratio (IWR), Floor-Space Ratio (FRR), and Roof-Space Ratio (RRR), the total area of the space, and the total height of the space.

Unfortunately, despite the standard definitions and the implementation guidelines produced to orient toward proper IFC models, it is expected that the entity's geometry of Ifcspace objects is wrong. At the early stage of 3D indoor geometry information extraction, that is enough to rely on the query of IfcSpace objects defined in an IFC model and their geometry information to extract each space's wall, window, roof, and floor geometry properties. Figure 3.1 illustrates an example of high-quality semantic and geometral modeling of IfcSpaces and the schema of IfcSpace in which the walls are split geometrically for each space (yellow Lines). A semantic-based query using the IfcOpenShell library extracts the geometry information, such as the wall area for each space, because the walls are split geometrically based on the boundary of each space. However, most IFC models do not have the second boundary definition of the objects (split geometry) for spaces, and there are inaccurate geometry calculations [139]. Figure 3.2a represents an IFC model with the non-topological definition of walls and windows for the corresponding spaces. This figure shows a shared wall (W143) between three spaces (S202, S201, and S203). In this case, the query returns an equal value of 3.84 m^2 of the wall-143 for each space, equal to the total area of the wall-143 because it is not split based on the space's boundaries. If we rely on this type of erroneous geometry value and apply it in the BECE process, it impacts the analysis of each space's energy consumption.

Besides that, the window is a critical entity in a building with a central role in energy gain or loss. In this regard, windows are responsible for more than 10% of the building energy load and are revealed to influence space energy consumption. If a space includes an exterior low-efficient window, the space's energy loss is increased drastically. Therefore, the energy consumption of space directly depends on the window size, location, and resistance rate [140]. For example, Figure

3.2b represents a shared window (WIN923) between S155 and S156. Performing a query from the existing IFC model returns an equal area value of window (WIN923) for both spaces.

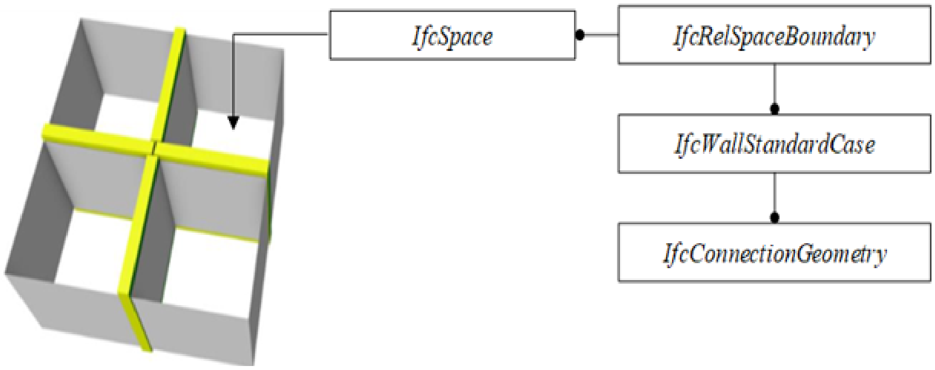


Figure 3.1. Valid IFC definition of the wall for each space

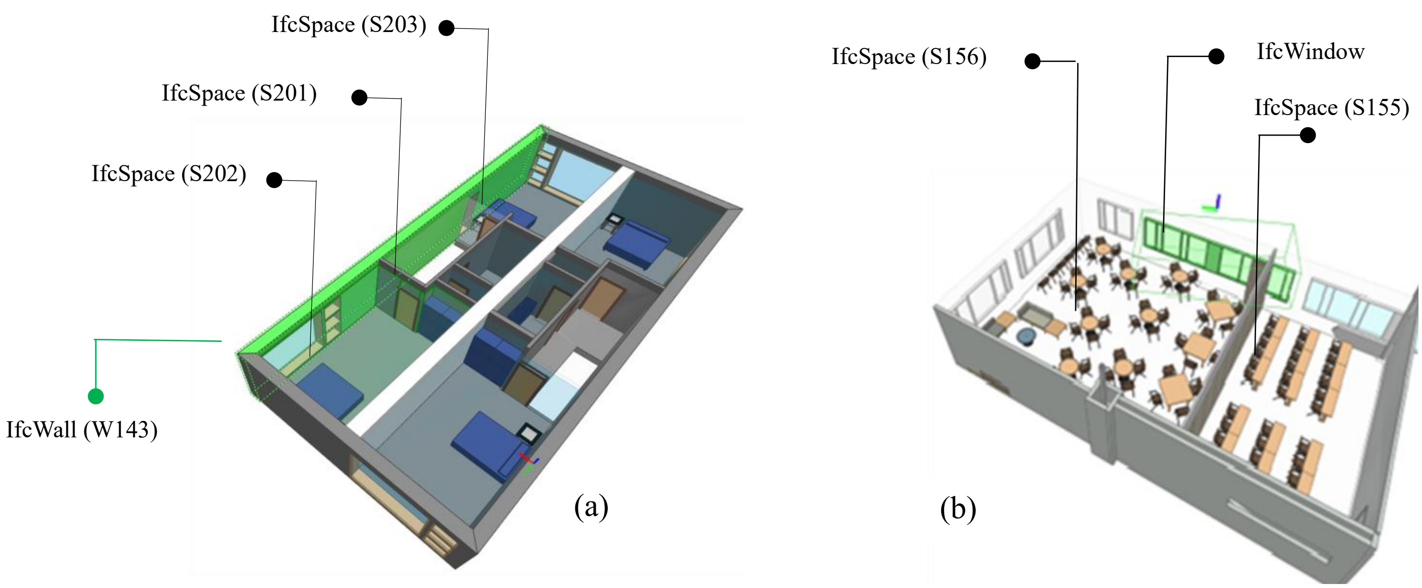


Figure 3.2. Invalid geometry of the wall and window for each space in the IFC models: (a) Illustration of the shared wall (green wall) between multiple IfcSpace objects; (b) Illustration of shared windows (green windows) between multiple Ifcspace objects

In contrast, the window size in S156 is significantly larger than in S155. Therefore, the impact of window 923 on the heating loss and gain in space 156 is more than its impact on space

155, which is not recognized if we rely on the current IFC query result. This leads to the necessity of geometric computation to extract accurate geometry information for space geometry information before using them in data-driven methods.

3.5.2. Identifying Inaccurate Space Adjacency (topology) Information in the IFC Model

Two spaces have adjacency in a 3D model when at least one shared wall, window, roof, or floor between them exists. In the domain of BECE, the energy transferred between the adjacent spaces affects the energy demand of each space. For example, if a space has a shared wall or floor (like the basement) with a cold space (with low energy efficiency), the heating loss rate increases significantly from the warm to the cold area.

Figure 3.3. The IFC model (LOD04) includes space and wall objects

Therefore, extracting accurate adjacent information is essential. Nevertheless, the IFC semantic-based query cannot return the valid adjacent information because of each space's invalid (non-split geometry) wall, roof, window, and floor geometry. Figure 3.3 illustrates the instance of this error, which results in the wrong adjacency list.

Figure 3.4 describes a semantic-based query using the current information in the IFC model. It employs the *BoundedBy* property of space objects to retrieve the *IfcWall* and *IfcCurtainWall* objects bounded to each space. By iterating through all space around the target space (B202 in Figure 3.3), the neighbor space objects are added to the adjacent list and considered adjacent. The result shows the wrong adjacent spaces for the target space because there is a continued wall 143 between the target space, spaces A202 and B203. Indeed, no geometrical relationship between them causes no energy transfer; however, the IFC semantic query considers their adjacent spaces. Therefore, retrieving the topological information based on the semantic query from the IFC model is inaccurate and causes erroneous results when generating the BECE space-based graph.

On the other hand, calculating the quantified value of space adjacency (weighted adjacency) is another parameter ignored in recent studies. Most researchers rely on binary space adjacency to find if two spaces are navigable by recognizing any opening object, such as a window or door between two spaces, such as indoor navigation applications. However, in BECE analysis, the weight of this relationship is important because the amount of transferred energy from one space to another is directly related to the surface area of the shared wall, window, roof floor, and material information. For instance, space A201 in Figure 3.3 is a cold space on this floor, significantly impacting the heating transfer to its adjacent spaces. In principle, the heating transformation from space A204 to space A201 (from a warm to a cold area) is more than the heating transformation from A205 to A201 because of the differential in the shared wall area. Unfortunately, this information is not explicitly available in the IFC model. They have to be derived from the available entities by advanced geometric processing and material isolation information to obtain the desired weighted graphs. Hence, developing a framework to translate the IFC model into a weighted space-based graph with accurate feature values compatible with BECE-based data-driven methods is necessary.

```
Adjacent_list = [ ]

spaceB202_walls = [ ]

foreach wall in space_B202.BoundedBy:

        // include ifcwall and ifcCurtainWall

        spaceB202_walls.append(wall)

foreach floor_space in spaces_in_floor:
```

Result:

- Space B201
- Space B205
- Space A203
- Space B203
- Space A202

Figure 3.4. IFC semantic-based query to extract space adjacency with the concept of the shared wall

3.6. METHODOLOGY

This research proposed a framework to generate a BECE-based Weighted space-based Graph (BECE-WG) to solve the compatibility issue between the IFC data model and graph-based machine learning algorithms. The graph represents a collection of interlinked IFC entities and their properties.

Indeed, BECE-WG organized the IFC dataset into connected graph structures, and thus, multi-source and heterogeneous data can be interlinked and integrated. This study adopts this concept to convert 3D building data models (IFC) to a graph to provide a data structure linked to different datasets compatible with machine learning algorithms. The proposed graph contains space (room) object information as a node and their relationships with neighbor spaces as the edge. The BECE-based information of each space is extracted and calculated geometry and semantically assigned as a feature vector to each node in the graph. Machine learning tasks such as supervised and unsupervised classification, node clustering, graph clustering, and node label prediction can employ the generated BECE-WG for building energy analysis. The framework has four main components: 1) BECE-based Object-Oriented data model, 2) Extract 3D Geometry Features, 3) Extract Space Adjancacy Information, and 4) Generate BECE Weighted Space-based Graph. Each component includes some detailed steps, which are represented in Figure 3.5.

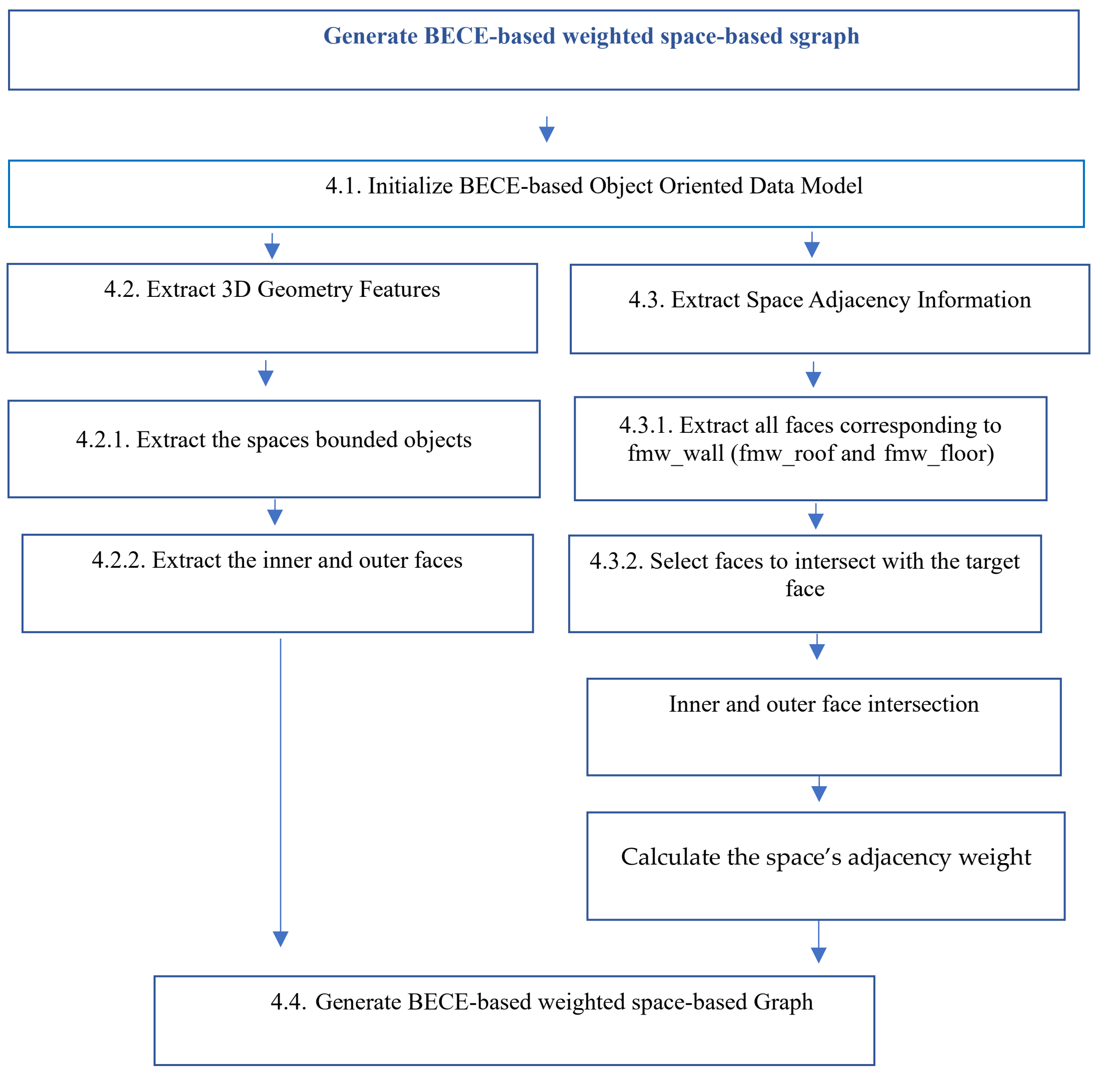


Figure 3.5.The workflow of generating a BECE-weighted space-based graph

3.6.1. BECE-based Object-Oriented Data Model

An object-oriented data model is proposed by introducing BECE-based building objects with the superclass of building and story classes. Story class includes space objects and child objects such as the wall, window, floor, roof, and face objects representing three-dimensional objects. This research uses fmw as a prefix for the proposed framework object (fmw-object), and IFC (IFC-object) is employed as a prefix for standard IFC objects. Figure 3.6 represents the proposed BECE-based data model, including the objects, properties, functionalities, and relationships. The proposed objects in the BECE framework included three main parts: properties, child objects, and functions. The property section encompasses the unique ID, energy transfer properties such as material information, and thermal Resistance Value (R-value). Also, it includes accurate geometry information such as area and volume for wall, window, floor, and roof objects for each space. The second section contains the child objects, which define the relationship between the framework objects. This section includes a single object or a list of objects. The relationship between the building entities can be defined by assigning the child object to the parent object, improving the query process, and extracting information from the proposed data model. For example, we proposed an *fmw-face* object as a child object of the *fmw-wall* to extract accurate geometry information on both sides (face of the wall) for each space. By adding the *fmw-space* object as a child object in the *fmw-face* object, we can find the corresponding space (grand-parent) of the specific face with a simple query process. However, this process leads to a complicated query in the IFC model to access the higher hierarchical objects (e.g., parent of parent objects).

The last section of the proposed data model for objects is their functionality. The functions associated with each object calculate geometry properties (e.g., area, volume, etc.), semantic property values (e.g., the type of materials) called ***Extract_3DGeometry_Features***, and identify the topological relationships between spaces, called ***Find_Adjacent_spaces***. However, as discussed in section 3.5.1, the geometry features cannot be extracted precisely by the IFC semantic-based query. Therefore, we proposed the ***Extract_3DGeometry_Features*** (section 3.6.2) function for space objects to apply the face partitioning concept to extract accurate face geometry information of wall, window, roof, and floor objects in each space, as explained in section 3.6.2 in detail.

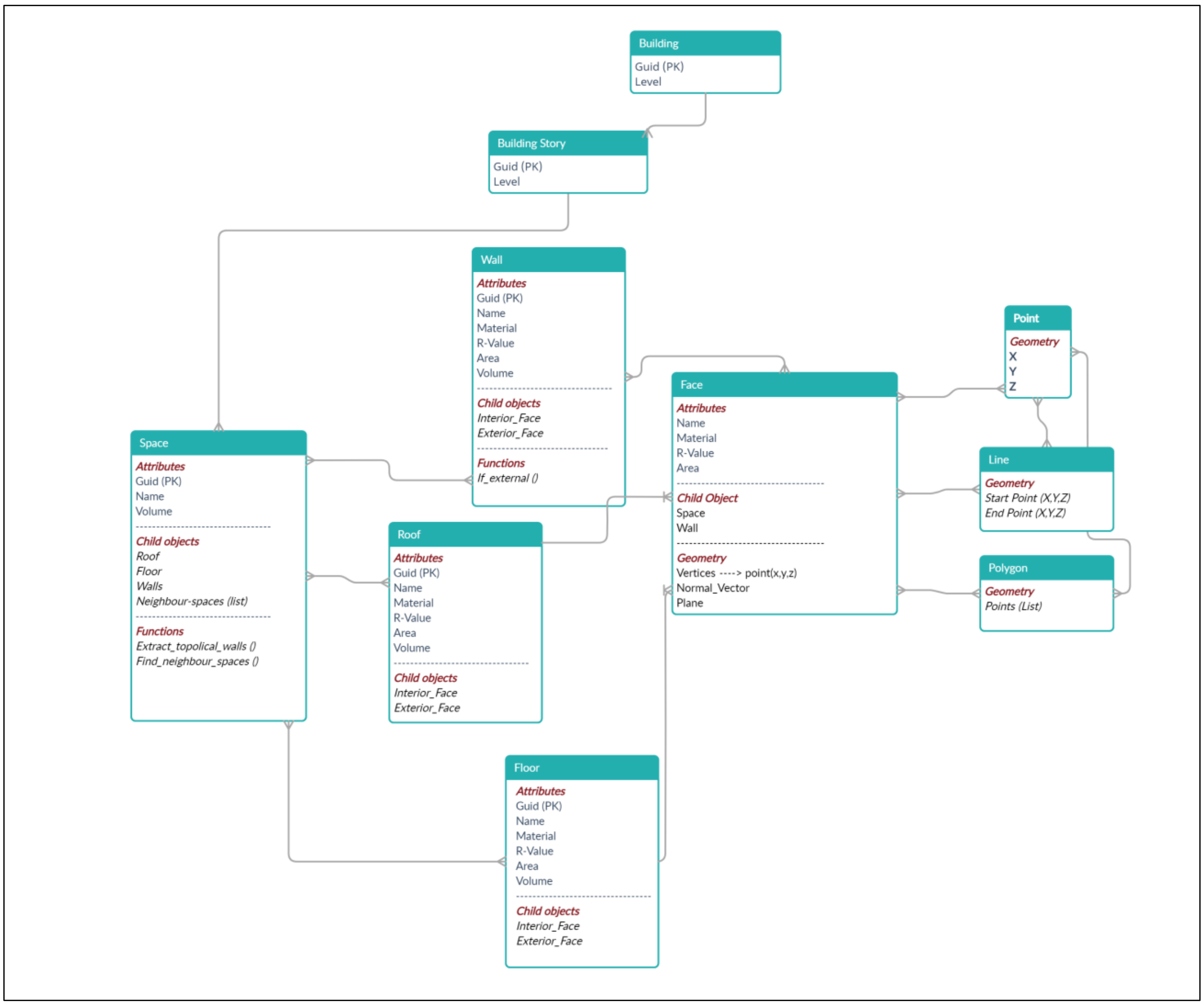


Figure 3.6. Proposed object-oriented BECE framework from IFC schema

The proposed data model tackles the space adjacent challenges in the IFC model (section 3.5.2) and improves the space-based relationship information between building objects in the IFC model. The relationships are defined based on the collection of child objects associated with each *fmw-object*. For example, an *fmw-space* object includes a collection of *fmw-wall*, and each *fmw-wall* object includes a collection of *fmw-face* as child objects. Therefore, by initializing the *fmw-space* object and calling *the Find_Adjecent_Spaces* function, adjacent spaces for the target space are calculated and populate the *Neighbor-space* list of the *fmw-space* object. Therefore, we can apply a standard object-oriented query to access all wall, face, and adjacent spaces objects corresponding to each space.

The pseudocode code in Figure 3.7 describes *fmw-space* object initialization and calculates the space properties and adjacent list by calling both ***Extract_3DGeometry_Features*** and ***Find_Adjecent_spaces*** functions. Section 3.6.3 discusses *Find_Adjecent_spaces* in detail.

```
Ifc_file = ifcopenshel.open('input IFC file')

Ifc-space = ifc_file['Guid'] --- > a single ifcSpace from IFC file

fmw-space = new fmw-Space(Ifc-space) --- > create fmw-space object
from proposed framework

fmw-space. Extract_3DGeometry_Features()

fmw-space. Find_Adjecent_spaces()
```

Figure 3.7. A pseudocode description of initializing of fmw-space object

3.6.2. Extract 3D Geometry Features.

The main goal of this section is to extract accurate geometry information such as Exterior Window-Wall area Ratio (EWWR), Exterior-Wall area ratio (EWR), Interior-Wall area ratio (INWR), Floor-Space area Ratio (FRR), and Roof-Space area Ratio (RRR) for building entities for each space. A closed volume characterizes a space in the IFC model of bounded walls, roof, and floor (space-bounding objects). Each bounded object includes a collection of faces, each of which is composed of a wire. The wire consists of edges and vertices. We used this granular information in IFC to create the *fmw-face* object for each space's walls, windows, roof, and floor to solve invalid space-based geometry issues, as mentioned in section 3.5.1. The geometric granularity description of faces allows us to generate an fmw-face for each space, which is considered an inner face for the target space. The steps of creating face geometry consist of (1) Extracting space's bounded objects (walls, windows, roof, and floor); (2) Extract the inner and outer faces of the bounded objects for each space; and (3) Define a new wall (fmw-wall) object and face (fmw-face) object for each space in the building. These steps extract the accurate geometry information for convex and non-convex spaces (non-convex hydron). Many previous studies focused on extracting the geometry information from the IFC. However, their solution is limited to extracting the geometry information for convex spaces as a boundary box (bbox). In reality, the IFC model includes convex and non-convex polyhedrons. Therefore, we dealt with this challenge by proposing a solution to extract the accurate geometry of bounded objects for convex and non-convex polyhedron spaces.

3.6.2.1. Extract the space's bounded objects.

In this method, we expect the IFC model to accurately include the wall-required information, such as *IfcSpace*, *IfcRelSpaceBoundary*, *IfcConnectionGeometry,* and

IfcWallStandardCase. Each space's bounding walls can be listed using the *BounedBy* function, which returns *ifcRelSpaceBoundary* elements, including the walls and *ifcCurtainWalls*. Figure 3.8 shows the semantic query to extract the bounding walls of ifcSpace and store them as a list of walls for a specific space. The ifc_wall_list includes all walls and curtain walls corresponding to ifcSpace. For instance, if we consider space-B202 as the target space, the query result in Figure 3.8 lists two exterior walls: wall-534, wall-590, and four interior walls: wall-960, wall-921, wall-301, and wall-239.

```
for rel_spaceboundary in ifcSpace.BoundedBy:
        related_element = rel_spaceboundary.RelatedBuildingElement
        if related_element.is_a('ifcWall') or is_a('IfcCurtainWall'):
                ifc_wall_list.append(related_element)
```

Figure 3.8. Extracting the space's bounded walls using the semantic query

3.6.2.2. Extract the inner and outer faces of the bounded objects

The reason for extracting the walls' inner face is to calculate the accurate geometry information (wall area for space). A shared wall between two spaces can have different areas based on the face geometry on each side. Figure 3.9 represents the 2D perspective of space-202 and the two inner and outer faces, planes, and normal vectors of each face for wall-235. The inner and outer face extraction is performed in two main steps:

Step I: In this step, we need to determine which face of the wall towards the inner of the space becomes part of the space's bounding volume. The IFC model has several shapes: Solids, Shells, Faces, Wires, Edges, Vertices, and compounds. We can iterate over the faces that bound the solid. Then, we can obtain a reference to the underlying surface of that face. Since the IFC model consists of planar geometry, we know that the surface of the face conforms to a plane, and then we calculate the plane for each face by knowing a point that belongs to the face ($\boldsymbol{p}_0 = (x_0, y_{0,} z_0)$) and a vector perpendicular (a,b,c) to the plane n. Eq 3.1 illustrates the plane question for all faces of the wall for space:

$$a(x - x_0) + b(y - y_0) + c(z - yz_0) = 0 \qquad \text{(Eq.3.1)}$$

By iterating through ifc_wall_list (as presented in Section 4.2.1), all faces for the walls for a specific space are listed, and their planes are calculated by selecting a node on the surface and its normal vector (Eq 3.1) and storing them as a child object of the frm_wall object (frm_face_list). Figure 3.10a illustrates the face representation, and Figure 3.9b explains the framework's fwm_wall object hierarchy. For example, the sample wall in Figure 3.10a illustrates six faces. To extract two prominent faces (inner and outer faces), the method iterates through the faces and calculates the area of each one.

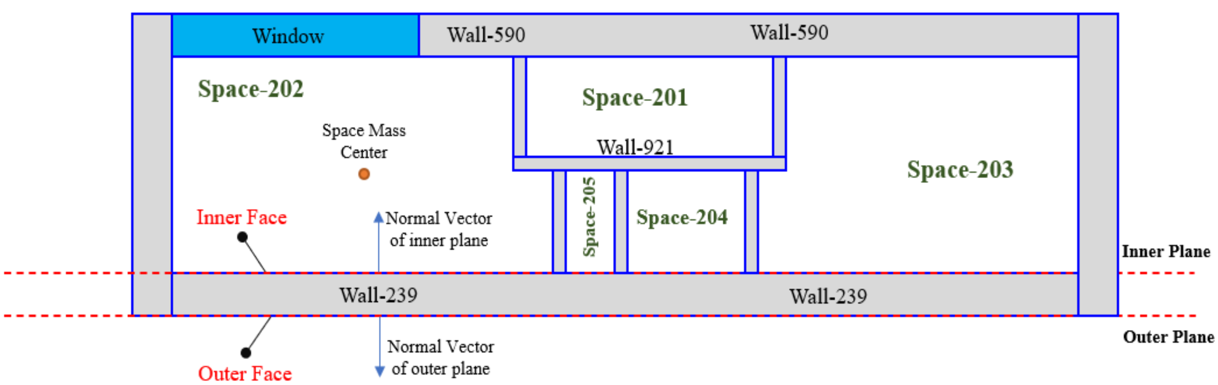
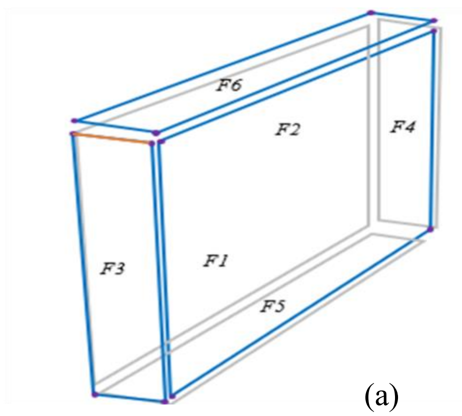
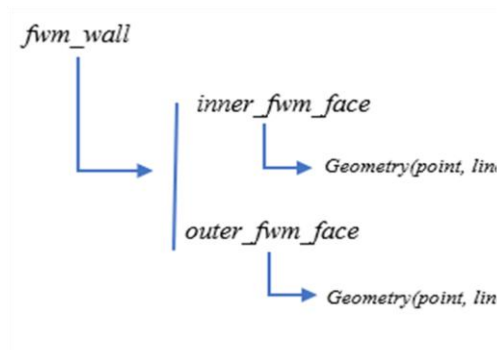


Figure 3.9.2D perspective of space-202 and inner and outer of Wall-239

The faces are sorted in ascending order based on their area value, and the two first faces are chosen as the main face objects. To find out the inner and outer face objects between these two faces, we applied the distance from plane method to calculate the distance of space's center point (Figure 3.9, space-202) from the plane of these two faces. The face with the minimum distance from the space's center point is considered the inner face, and the second is assigned as the outer face object. This method has a high accuracy rate in distinguishing between the inner and outer faces of convex polyhedron spaces. However, this method has incorrect results returned for some walls in non-convex polyhedron spaces. Figure 3.11 represents an example of this non-convex polyhedron space. The outer face of wall 921 is closer to the space's center point than the inner face. Therefore, the algorithm assigns the outer face (red dash-line) as the inner face object, which is incorrect.



(a)



(b)

Figure 3.10. (a). ifcWall representation with faces, faces' lines, and points ; (b). the proposed fwm_wall and fwm_face object-oriented data model

Consequently, we need additional conditions to increase the accuracy of the results and correct the wrong faces assignment. The next step describes this process. Also, after finding the first step of the inner and outer faces, we can explore which portion (polygon surface) of the plane belongs to the target space.

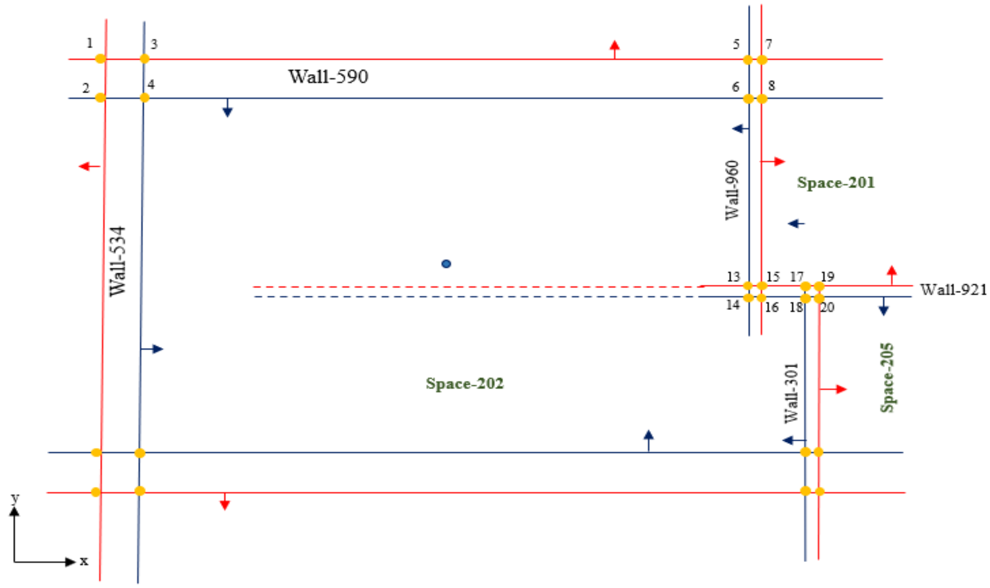


Figure 3.11. The 2D perspective of the face intersection of space-202

Therefore, there is no intersection for those planes whose normal vector's dot product is zero, and the algorithm considers them parallel faces. By looping through the frm_face_list, all pairwise planes corresponding to each face are considered The vector equation of the intersection line. The cross-product of the normal vectors of the given planes calculates Eq.3.2. To do that, a point (x_0, y_0, z_0) on the line of intersection is necessary. Therefore, the equations of the given planes are employed as a system of linear equations Eq.3.3, and the z = floor is set as the z-value in this equation. Then, calculate the intersected line equation and store the intersected line in an array list (line_arraylist).

$$v = |plane1 \times plane\ 2| = \begin{vmatrix} i & j & k \\ a_{p1} & b_{p1} & c_{p1} \\ a_{p2} & b_{p2} & c_{p2} \end{vmatrix} = (a, b, c) \qquad \text{(Eq.3.2)}$$

$$L = \frac{x - x_0}{a} = \frac{y - y_0}{b} = \frac{z - x_0}{c} = t \qquad \text{(Eq.3. 3)}$$

By the intersection of walls, windows, roof, and floor with the target face, four lines are extracted to delineate the polygon surface of each inner face. Each inner space's polygon geometry is determined using this result, and geometry information such as area, perimeter, and space boundary (wire) for each wall can be calculated. The intersection method helps solve the issue regarding inaccurate geometry information when there is a continuous multi-space shared wall in the IFC model mentioned in section 3.2. The face intersection method can reach the accurate geometry of faces and obtain the accurate geometry of each wall, roof, and floor for the target space.

Step II: At this step, post-processing of inner and outer face extraction is applied to rectify the possible error in recognizing walls' inner and outer faces. Therefore, we use the advantage of wall partitioning (fmw_face object) results to investigate the potential error for some walls in a non-convex polyhedron space. We apply the ray-tracing intersection method for all recognized inner faces in the first step to solve the mentioned issue. By iterating through to all recognized inner faces from step 1, we can make a ray-crossing line for each face using the face center point and its plane's normal vector. Then, intersect the created ray-cross (line) with all inner faces and find the intersection point (x,y,z). If the intersected point is at least inside one of the inner faces' polygon, the target face is an inner face. It is considered an outer face if it does not meet this condition. Figure 3.12 illustrates a sample of applying the ray-crossing algorithm to rectify the extracted inner and outer face for wall-921. The intersection of the red-face ray-crossing of wall-921 with the inner face of wall-590 generates a 3D point outside of all highlighted inner faces of space-202. Therefore, the red face, considered the inner face in step one, is assigned as the outer face because it does not meet the condition. Moreover, the blue face of wall-921 is assigned to the inner face object for space-202. Finally, applying these three steps to all IFC model spaces, WWR, EWR, INW, FRR, and RRR are calculated accurately using the rich geometry computation in the proposed framework, and these calculation results are stored as the space objects' attributes.

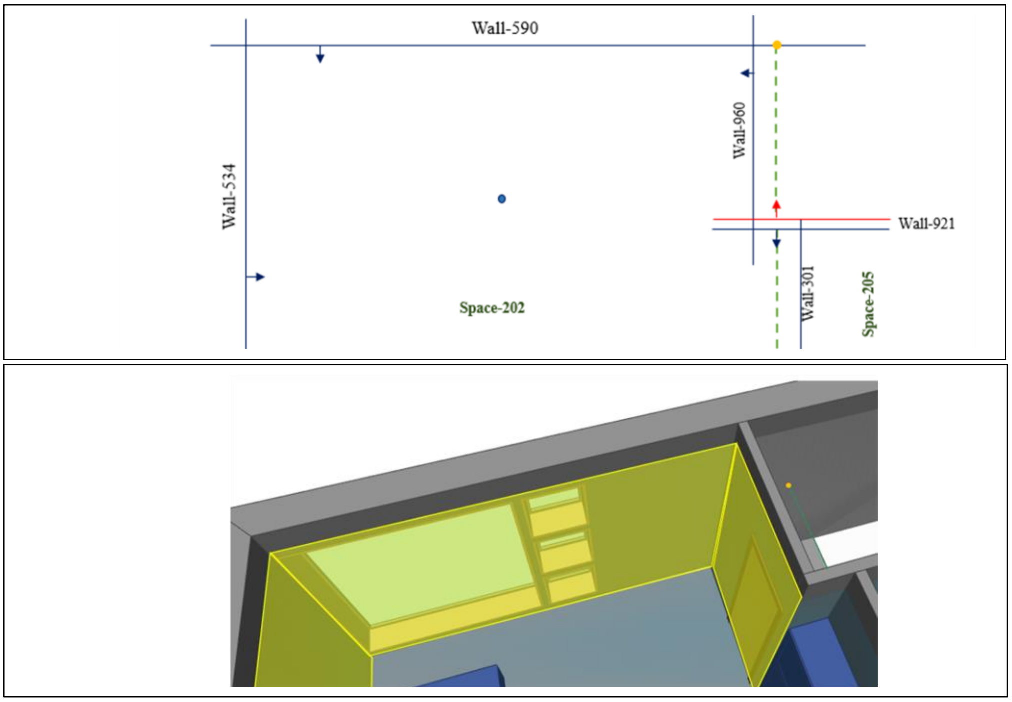


Figure 3.12. 2D and 3D perspectives of ray-crossing line to find inner and outer faces

3.6.3. Extract Space Adjacency Information

The adjacent space analysis is a critical spatial analysis in BECE to measure the effect of the neighbor spaces on each space's energy transfer. Therefore, in the BECE application, it is necessary to investigate the adjacent spaces affected by different adjacent degrees when a space consumes high energy. However, recent research has not explored the algorithm to extract the accurate space adjacent and the quantified weight of adjacency from the IFC model [70], [141]. Section 3.5.2 discussed the problem of the semantic query from the original IFC model in extracting space-adjacent information. It implies that the current methods of extracting space adjacents are inaccurate because they assume they are adjacent if two walls corresponding to two spaces have a shared edge in the floor plan. IfcSpace can acquire the adjacent wall information;

however, multiple IfcSpace objects can share one wall, which means it is not feasible for the adjacency relations of the IfcSpace to be available by sharing the same wall. In this research, we develop our space's adjacent extraction method on the inter-relationships among IfcSpaces to extract the quantified space adjacent extraction. We use this method as a node relationship information (edge) to generate the final graph. As discussed in section 3.6.1, the object in the proposed framework includes some functions to extract the information unavailable in the IFC model, such as finding adjacent spaces for each space. The proposed method encompasses three steps: First, we extract all faces related to each wall in a space using a query from framework wall objects, not the IFC objects. Second, select the faces whose normal vector of planes has the opposite direction from the target face. Finally, the polygon intersection method detects if two faces are intersected.

3.6.3.1. Extract all faces corresponding to fwm_wall

The proposed framework wall object (fmw_wall) lists fmw_face child objects. First, the face objects in this list are calculated and populated based on the proposed method in section .3.6.2 Next, we define and create the face object by extracting the fmw_face object and assigning it as a child object for each wall in the IFC model. This object is not defined explicitly in the IFC model, but the proposed framework makes it available for BECE analysis. Figure 3.13 describes the query syntax of extracting all faces (both sides) for wall-239. Lines 1 to 3 of the syntax code show how to apply a simple query to access eight fmw_face objects for wall-239.

3.6.3.2. Select the faces to intersect with the target face.

By accessing all faces and their polygon geometry, the algorithm filters them to detect any intersection between the target_face (red face in Figure 3.13 - Side1). Therefore, the dot product of the plane's normal vector of the target face and the other faces is calculated, and those with a dot product value of more than zero are selected for the intersection test.

3.6.3.3. Inner and outer face intersection

This step applies a pairwise intersection between the outer faces extracted from the previous section and the target face. The research assumes all faces are a convex 2D polygon; accordingly, we apply the 2D polygon intersection method to detect if two faces are interesting. However, we need two polygons on the same plane to apply the intersection method. Therefore, we project the outer face polygon on the target face's plane and then apply an intersection between the two polygons. Figure 3.14 represents the projected polygon schema (dashed-red line) of the outer face of the target plane. The projected polygon intersects with the target face's polygon to find the area of intersection (Figure 3.11, lines 8, 9, and 10). Figure 3.14 explains the calculate_polygon_intersection_area method, which uses the shapely open-source library to calculate two polygons' intersected areas.

Consequently, just one face from wall 239 intersects with the target face, and the parent object of that face (fmw_space) is recognized as one of the adjacent spaces. Using the proposed method in section 3.6.3, we can extract the 3D adjacent spaces for each space and the area value of adjacency. Finally, this method is applied on the floor and roof to find the adjacent spaces on different floor levels and calculate the shared area value.

```
1-  fmw_wall_236 = framework.WALL('wall-239')
2-  targer_face = framework.SPACE('space-202')
3-   for fmw_face in fmw_wall_236.fmw_face_list:
4-    if fmw_face != target_face:
5-     If  dot(target_face.plane.normal_v
    fmw_face.plane.normal_v) <> 0:
6-       projected_polygon =
7-      framework.project_polygon(fmw_face.polygon,
    targer_face.plane)
8-          int_area =
    framework.calculate_polygon_intersection_area
                 (projected_polygon, targer_face.polygon)
9-        If int_area > 0:
10-        target_space.adjacent_fmw_spaces.append
```

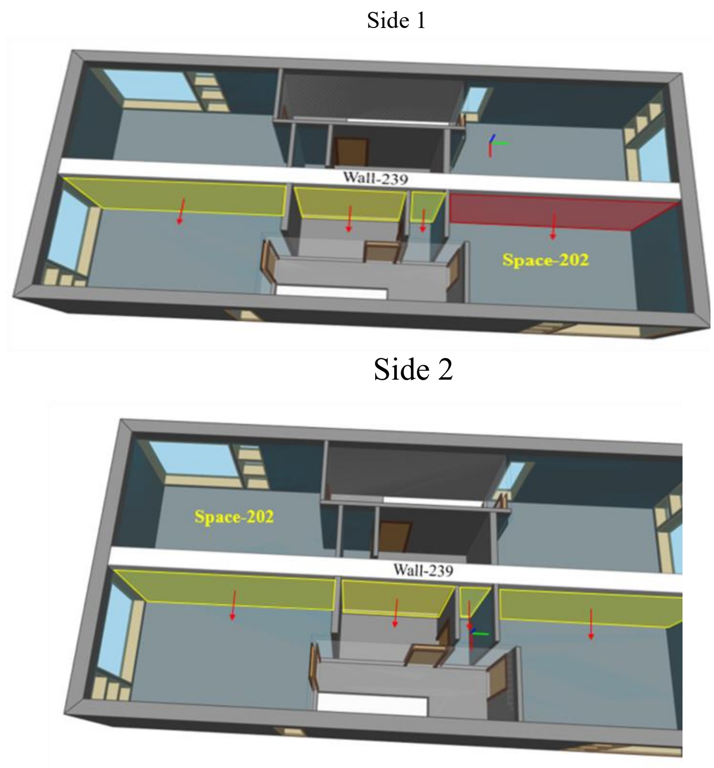


Figure 3.13. Find the adjacent spaces for space 202.

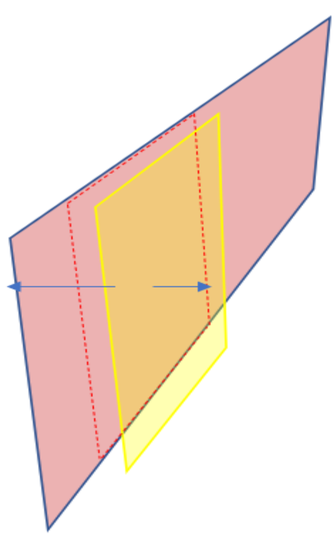


Figure 3.14. 3D schema of peojection of face's polygon onto the target plane

```
for rel_spaceboundary in ifcSpace.BoundedBy:
from shapely.geometry import
p = Polygon([[(xp1 , yp1). (xp2 , yp2), (xp3 , yp3), (xp4 , yp4)]])
q = Polygon([[(xq1 , yq1). (xq2 , yq2), (xq3 , yq3), (xq4 , yq4)]])
if p.intersects(q) :
        int_area = p.intersects(q).area
```

Figure 3.15. Extracting the space's bounded walls using semantic query

3.6.3.4. Calculate the spaces' adjacency weight

The edges in our proposed graph are representative of the adjacent space relationship. The BECE-based graph is a weighted graph in which weights reflect the relative importance of the spaces' relationships. In other words, some links between nodes have lower weights (low importance), and others have higher weights (high importance). For example, suppose two spaces have more energy transfer in our application. In that case, the relationship between those spaces is more important than two other spaces with a low rate of energy transfer. The energy transfers between two spaces through the walls, roofs, floors, and windows. Each of these building elements comprises several layers with different materials. The energy transfer of each building element is determined by the thermal resistance value (R-value). Indeed, the R-value measures how well a two-dimensional barrier, such as a layer of insulation, a window, or a wall or ceiling, resists the conductive flow of heat [142]. Therefore, we introduced a wight for the space relationship by calculating the barrier's R and shared area values (section 3.6.3.3). Eq. 3.4 describes the proposed weight value of edges in a BECB space-based Graph. The $Weight_{n1,n2}$ shows the importance value of two space's relationship. Share area is the common area value of spaces n_1 and n_2. The $\sum_{l=1}^{m} R - Value_l$ is the sum of R-Value for n number of barriers between two spaces.

$$Weight_{n1,n2} = \frac{Shared\ area}{\sum_{l=1}^{n} R{-}Value_l} \qquad (\text{Eq } 3.4)$$

3.6.4. Generate BECE-based Weighted Space-based Graph

After creating fmw_face objects and extracting the BECE-based attributes and relationship information, a new method is developed to transfer the *fmw-face* objects to a weighted BECE space-based graph. This method constructs the adjacency matrix following and the weighted space-based graph based on the *fmw-fac*e object list from sections 3.5 and 3.6. It includes three main steps. First, the process starts with the Normalization Function (NF) to normalize the feature values. Since each parameter has a different scale, the normalization must create common scale feature values for machine learning algorithms. Therefore, we applied the min-max normalization method to normalize the input feature values. The second step generates the adjacency matrix using the list of adjacent objects (*fmw_space*). Indeed, the adjacency matrix represents the graph as a square matrix with the size of n * n (n: number of nodes). In the last step of this process, the function is designed to iterate through the spaces to find the space's adjacency and assigned weight value. In this step, we have used the NetworkX library in Python [143] to construct the space-based graph using the adjacency matrix and embed six feature values as each node's attributes. The generated graph is homogeneous because all nodes have equivalent types (space) and similar edges to the space's neighborhood. Therefore, the BECE space-based graph is a compatible data model for future space-based energy consumption analysis using a graph-based machine learning algorithm (Graph-ML).

3.7. EXPERIMENTS AND RESULTS

In order to test and validate our approach, we apply the proposed methodology to two test case studies and two multi-level IFC models of ERDC[1]: the Duplex Apartment Model and the Autodesk Trapleo building. These test cases are chosen because multi-level building information is available in the models. Also, the ifcSpace objects are defined in these two IFC models. We hypothesized that the framework's extracted semantic and space relationship information are

[1] . http://openifcmodel.cs.auckland.ac.nz/

identical due to the manual measurement of geometry information and visual space relationship extraction.

The entire BECE framework discussed in the previous sections has been implemented by Python, which several open-source libraries support, such as IfcOpenShell and NetworkX [144]. Finally, we notably used Revit and BIMVision software for visual representation and ground truth measurements while simultaneously supporting the original IFC models' geometry, topology, and semantic information.

3.7.1. Case Study I: ERDC-Duplex Model

The first IFC model as a test case is a two-story building with eighteen spaces on two levels. This model includes the convex and non-convex spaces, and all the walls, roofs, and floors are perpendicular. Also, eight continuous exterior and two continuous interior walls are available in this model, which describes the first mentioned issue in this research. The multi-space shared wall issue in this model opens this opportunity to test our methodology from the perspective of accurate geometry extraction and recognizing the correct space adjacent. The proposed BECE framework generates an accurate graph from the IFC model for test case 1. Figure 3.16 represents the steps of this process for the IFC model.

Figure 3.16a represents the IFC model of this test case. Using the proposed method, in the first step (Figure 3.16b), the framework objects (parent objects) such as fmw_wall, fwm_roof, fmw_floor, and fmw_window are instantiated for each space. Next, the face extraction method is applied to extract both faces of the wall, window, roof, and floor objects and store them as child objects for each space parent object (Figure 3.16c). Finally, the Extract Space Adjacency method is employed to find the adjacent and calculate the weight value for all spaces (space) in two building levels (Figure 3.16d).

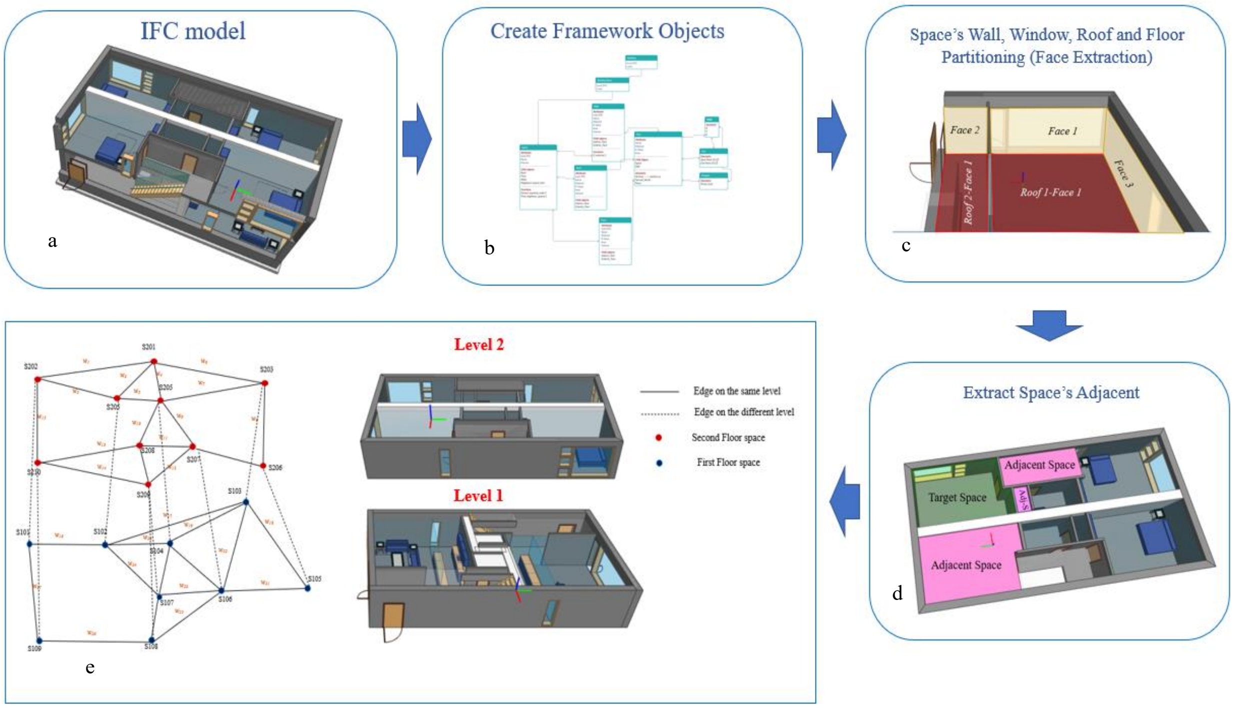
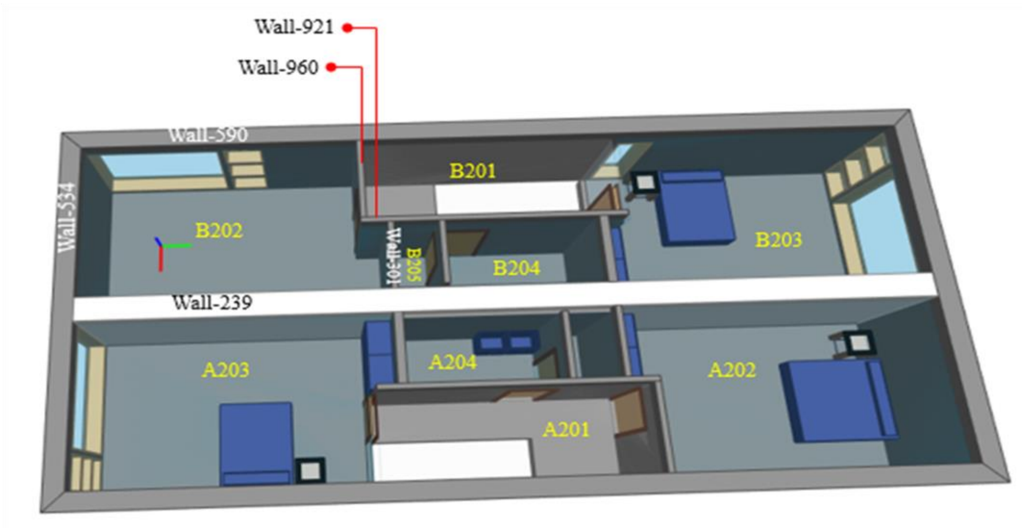


Figure 3.16. The steps of generating a BECE Space-based Graph from the IFC model in test case 1

Figure 3.17 illustrates geometry and adjacent information for space 202 extracted from the IFC model and proposed framework. The accuracy of extracted information is evaluated by the ground truth values measured using BIMVision software. Figure 3.17 and Figure 3.17b show the wall, roof, and floor area information for space B202. Comparing these two tables' values declares the wrong area value for continued walls of 239, 534, and 590 when geometry information is extracted from the IFC model. However, the framework has accurate extracted information.



(a) Geometry Features Extracted from IFC Model

Wall-239 - Int	**49.20 m²**
Wall-534 - Ext	22.95 m²
Wall-590 - Ext	49.20 m²
Wall-960 - Int	5.69 m²
Wall-921 - Int	15.73 m²
Wall-301 - Int	4.64 m²
Roof	135.15 m²
Floor	23.17 m²

(b) Adjacents Spaces Extracted from IFC Model

Space B202	A203	1
	B201	1
	B205	1
	B203	1
	A202	1
	Outside-Wall	1
	Outside-Roof	1

(c) Geometry Features Extracted from Framework

Items	Results
Wall-239 - Int	18.12 m²
Wall-534 - Ext	10.75 m²
Wall-590 - Ext	16.56 m²
Wall-960 - Int	5.69 m²
Wall-921 - Int	1.57 m²
Wall-301 - Int	4.64 m²
Roof-102	17.50 m²
Roof-103	4.58 m²
Floor	22.04 m²

(d) Adjacents Spaces Extracted from Framwork

Target Space	Adjectes Spaces	Weight
Space B202	A203	0.906
	B201	0.114
	B205	0.092
	A102	0.72
	B103	0.16
	Outside-Wall	22.61
	Outside-Roof	3.06

Figure 3.17. The extracted information from the IFC model and framework

3.7.2. Case Study II: Trapelo Building

As a test case, the second IFC model is a commercial three-story building with 130 spaces (Figure 3.18.a). This model includes convex and non-convex spaces, but several walls have a non-perpendicular intersection, unlike the case test one (Figure 3.18b). More than 40 continuous exterior walls and 25 shared windows are available in this model.

(a)

(b)

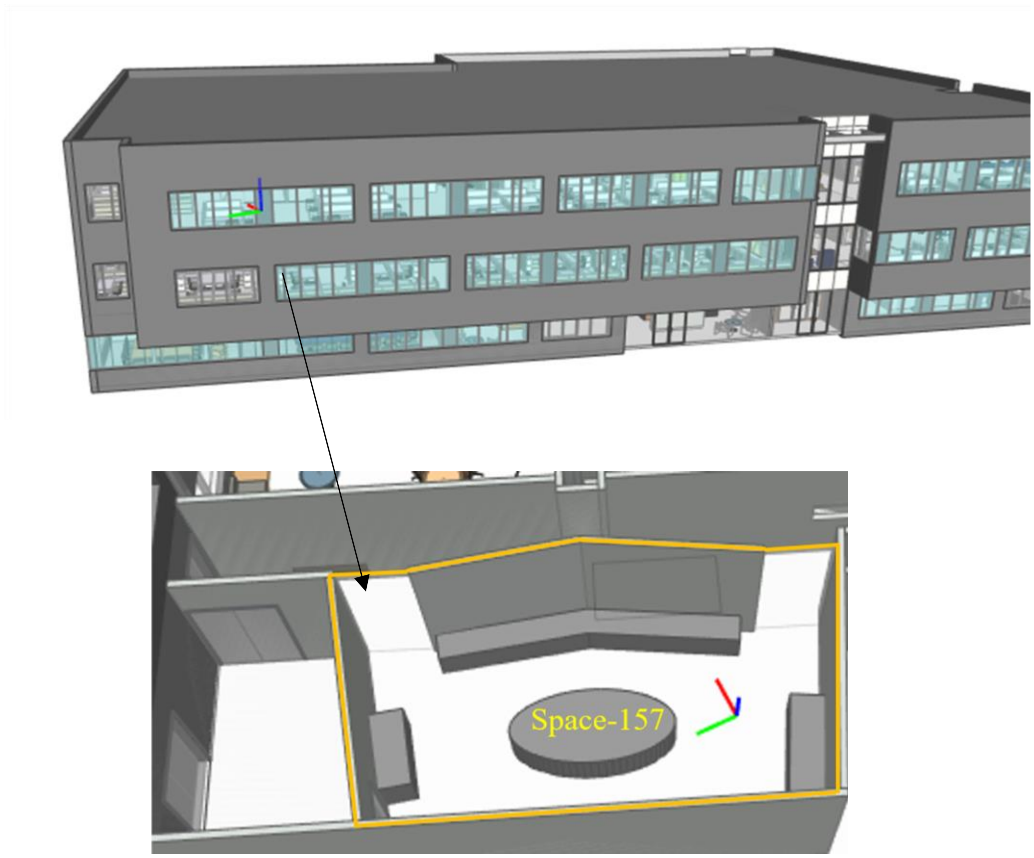


Figure 3.18. (a) The 3D IFC Model of Test Case; (2) the TheNon-Perpendicular Space (Space-157)

This test case is a more complex model than the first one because of the number of levels, geometry, and space relationships. Therefore, it could be a suitable model to evaluate the proposed methodology to extract accurate geometry and semantics. The four steps of generating a framework data model, extracting BECE-based geometry and semantic information, extracting space's adjacent information, and generating the BECE-based weighted space-based graph are applied to this test case. The proposed methods for these two test cases are evaluated by validation parameters introduced in the following section.

3.7.3. Experimental Result and Discussion

We verified the validity of the proposed algorithm in extracting geometry information using the Root Mean Error (RMSE), a standard way to measure a model's error (or level of accuracy) to evaluate the model's estimated values for quantitative data. Indeed, the geometry (area) of walls, windows, roofs, and floors for all spaces in the two test cases are extracted by BIMVision software, which is considered a ground truth value. Then, the RMSE value is calculated using Eq. 5 for the extracted geometry information of building elements for estimated values from the IFC model and framework objects. $RMSE_e$ shows the accuracy of the estimated area value for the building elements (Wall, Window, Roof, and Floor). The a_i is the calculated area of elements from models, and ag represents the area value based on the ground truth measurement.

$$RMSE = SQRT(\sum_{i=1}^{N} \frac{(a_i - a_g)^2}{N}) \qquad \text{(Eq.3. 4)}$$

Table 3.1 and Table 3.2 represent the root mean square error of geometry information extracted from the IFC model and calculated by the Test Case I and II frameworks. A comparative analysis of the errors between the geometry estimated values from the IFC model and the BECE-framework model is used for two test cases. Table 3.1- test case I and Table 3.2.- test case II) demonstrate the high accuracy in calculating space's geometry information using the proposed methodology. Also, the error in the area value extracted from the IFC model is incremented by increasing the number of spaces. For example, the error of extracted wall geometry information from the IFC model is increased between case studies I and II because of the higher number of non-perpendicular spaces in case study II. Also, in test case II, the number of continuous walls (shared between multiple spaces) and windows is higher than in test case I, which causes a decrease in the accuracy of geometry information. Thus, it demonstrates that BECE analysis cannot account for space's geometry information in the IFC model because the inaccurate extracted information is causing the wrong BECE data-driven analysis.

Table 3.1.The error of geometry information extracted from the IFC model of Test Case I with 18 spaces

Validation Parameters	RMSE of IFC model (m2)	RMSE of BECE-Framework (m2)
Wall area estimation	2.1	0.10
Roof area estimation	1.9	0.09
Window area estimation	1.78	0.058
Floor area estimation	1.89	0.026

Table 3.2. The error of geometry information extracted from the IFC model of the Test caseII with 130 spaces

Validation Parameters	RMSE of IFC model (m2)	RMSE of BECE-Framework (m2)
Wall area estimation	3.9	0.11
Roof area estimation	1.4	0.12
Window area estimation	3.1	0.08
Floor area estimation	3.2	0.03

Another aspect being considered is the high possibility of the failure of the space relationship (adjacent information). This is primarily due to non-topological information about the IFC model's wall, window, roof, and floor elements. Indeed, the IFC model's incorrect information of space adjacent leads to a wrong BECE space-based graph. Table 3.3 describes the percentage of the total number of correctly detected adjacent spaces to the total number of spaces for each floor. The low percentage of adjacent information from the IFC model illustrates the importance of recognizing and extracting the valid adjacent information in the IFC model. The proposed methodology for extracting adjacent information shows high accuracy for vertical and horizontal adjacent information. However, although the framework has a high performance in finding the adjacent spaces, it could not extract 100% adjacent spaces for case study 1 because there is more than one wall (two walls – IfcWallStandardCase and Party Wall) between two spaces. Therefore, this is a limitation of the proposed methodology. Another limitation to be considered is that the primary assumption of this research is that the ifcSpace object is defined in the original IFC model. However, the methodology failed for those IFC models without the ifcSpace object. Also, the proposed framework does not support spaces with complex geometry, such as star-shaped or architectural roofing spaces. Therefore, this part of the solution remains an unsolved challenge for future attempts.

Table 3.3. The percentage of correct extracted adjacent spaces for two test cases

Name	**Query from IFC(%)**	**BECE-Framework (%)**
Test Case 1	70	98
Test Case 2 (First Level)	65	100
Test Case 2 (Second Level)	60	100
Test Case 2 (Third Level)	60	100
Test Case 2 (All Levels)	70	100

3.8. CONCLUSION AND FUTURE WORK

This paper introduced a BECE-based framework to fill the data-driven model gap, such as machine learning algorithms for building energy consumption estimation (BECE) using the IFC model. This research proposed a framework to generate a BECE-based weighted space-based graph from the IFC model. While the existing approaches rely on the semantic and geometry information in the IFC model, which causes several erroneous analysis results, the proposed methodology defined classes and objects in the new data model comprise the explicit semantic, geometry, and space adjacent information that is not available in the native IFC models. In addition, some functionalities are also designed and implemented to apply geometry computation to extract accurate semantics, geometry, and weighted relationships for IfcSpaces in the IFC model. The low percentage of adjacent information from the IFC model illustrates the importance of recognizing and extracting the valid adjacent information in the IFC model. The proposed methodology for extracting adjacent information shows high accuracy for vertical and horizontal adjacent information. Also, the result demonstrates the high accuracy of semantic and geometry information for the spaces in the model.

Furthermore, an accurate space relationship algorithm is proposed to extract space-adjacent information with a weighted value. Moreover, the proposed methodology is not limited to extracting the information for spaces with convex polyhedron geometry; it is designed to apply geometry computation for convex and non-convex polyhedron spaces. To this end, an accurate BECE-based weighted graph is generated. Thus, the proposed graph encompasses accurate semantic, geometry, and space-adjacent information. Also, the space-based graph does not have IFC model complexity and is compatible with data-driven algorithms.

Nevertheless, the results depend on the quality of the definition of spaces in the input IFC model on which the process relies. Stability and topology validity cannot be ensured if the input spaces are not appropriately defined in the model. Another aspect that needs to be considered in future work is the adjacent information between the spaces, which can be calculated more accurately with more than one wall, roof, or floor. Our proposed methodology assumes a single object between the spaces and calculates the weight of a space's relationship based on area and material. However, additional attempts must be considered and investigated to find the objects' number and material information between the two spaces. Consequently, the proposed data model has more accurate geometry and semantic information for spaces, more realistic objects close to real-world entities, and a more readable structure for different applications.

CHAPTER 4: A GRAPH-BASED EXPLANATORY MODEL FOR SPACE-BASED ENERGY EFFICIENCY ANALYSIS BASED ON BIM DATA

4.1. PREFACE

This chapter focuses on the second sub-objective, proposing a graph-based classification model to consider 3D geometry, semantic, and 3D topology information for space classification problems regarding energy consumption in the IFC model. Additionally, an interpretable model is introduced by incorporating the SHAP model with the proposed classification model, enhancing transparency in data-driven models.

H. Kiavarz, M. Jadidi, and P. Esmaili, "A graph-based explanatory model for room-based energy efficiency analysis based on BIM data," *Frontiers in Built Environment*, vol. 9, Sep. 2023.

Copyright and Licensing:

Contributions:

The contributions of authors in the current chapter are as follows: **Hamid Kiavarz** has conducted the literature review, performed the data collection and curation, developed the method, used the required software to perform the analysis and modeling, validated and visualized the results and wrote the original manuscript of this publication. **Payam Esmaili** contributed to the case study energy simulation and performed it using RETScreen software. **Mojgan Jadidi** has supervised the research, provided the funding, and contributed to developing the manuscript's method, writing, and editing.

4.2. ABSTRACT

In recent years, the growing interest in building energy consumption and estimation has led to a wealth of energy data and Building Information Modeling (BIM), providing ample opportunities for data-driven algorithms to be widely applied in the building industry. However, despite promising accuracy in data-driven models for building energy estimation, they only consider building elements and their attributes independently and neglect the interconnected relationship of building elements. Also, Current data-driven models lack interpretability and are often treated as black boxes. As a result, the models cannot be fully trusted for engineering without reasoning the underlying mechanisms behind the estimation. This paper emphasizes the potential

of graph-based learning algorithms, specifically GraphSAGE, in utilizing the enhanced semantic, geometry, and space topology information derived from BIM data. The aim is to identify critical spaces within the building based on their energy consumption characteristics. Besides that, the paper proposed a GraphSAGE explainable model by adopting the SHAP with the proposed NE-GraphSAGE prediction model to make the data-driven models more transparent. Preliminary results demonstrate the potential to improve pre-and post-construction steps by identifying critical spaces in buildings and the parameters that affect the efficiency of the spaces with low energy consumption.

4.3. INTRODUCTION

The global energy consumption rate has experienced a significant increase over the past decade, reflecting the growing demand for energy resources. This is not good news, leading to sustainability and fighting climate change. The building sector accounts for over 40% of global energy consumption, as reported in 2022 [145]. Energy consumption in buildings is projected to increase by an average of 1.3% per year from 2018 to 2050 in OECD countries, including the USA, Canada, Europe, and Australia. Non-OECD countries, such as the Middle East, China, and Russia, are expected to experience a higher annual increase of over 2% in energy consumption. On the other hand, climate change is an emerging issue that needs to be considered and makes our buildings more resilient and sustainable to heat loss.

To address these concerns, recent research on new building development projects has focused on seeking novel ways to design and retrofit the buildings more energy-efficiently. To do so, the interest in using newly available datasets such as Building Information Models (BIM) and applying data-driven solutions emerge as the efficient and suitable option for the Building Energy Consumption Estimation (BECE) analysis rather than employing classical physics-based model

[57] to improve the estimation and prediction of building energy consumption in different stages of design, development, and retrofitting.

Data-driven-based BECE requires detailed static and dynamic data to simulate building energy consumption, eventually leading to prediction learning from historical/available data. Data-driven energy consumption prediction has been a prominent area of research in recent years. Numerous review studies have been conducted to analyze and evaluate existing data-driven approaches. These studies contribute to advancing knowledge and informing the building industry, for example, categorizing building energy consumption prediction methods into several categories, including elaborate engineering methods, simplified engineering methods, statistical methods, Artificial Neural Network (ANN)-based methods, Support Vector Machine (SVM)-based methods, and grey models. They conducted a comparative analysis considering various factors such as model complexity, ease of use, running speed, required inputs, and accuracy of the methods. Their analysis provided insights into the strengths and weaknesses of different approaches for energy consumption prediction in buildings. Ahmad et al.[146] specifically examined Artificial Neural Network (ANN)-based, Support Vector Machine (SVM)-based, and hybrid methods for building energy consumption prediction. They discussed the principles, advantages, and disadvantages of these methods, providing valuable insights for researchers and practitioners. Fumo [129] focused on summarizing the classification of building energy consumption prediction methods proposed in various studies. They emphasized the importance of model calibration and verification, which are essential steps in the modeling process. Their review highlighted the significance of accurate and reliable predictions in the field, and they conducted a comprehensive review encompassing various aspects of building energy modeling and prediction. They examined state-of-the-art studies on indoor building space energy modeling, prediction, and critical component modeling, such as photovoltaic power generation. Additionally, they covered topics such as building energy modeling for demand response, agent-based building energy modeling, and system identification for building energy modeling. This inclusive review provided a broad perspective on building energy modeling and prediction facets, contributing to a comprehensive understanding of the field. Li et al. [147] reviewed building energy benchmarking methods and presented a flowchart to guide users in selecting the appropriate prediction method. Their work aimed to assist users in making informed decisions based on their specific needs. Chalal et al. [53] focused on energy consumption prediction at building and urban scales. They classified and discussed the available methods within each scale, providing insights into the different approaches and their applicability in different contexts. Their study contributed to understanding energy consumption prediction in the broader context of buildings and urban environments. Wang and Srinivasan [148] comprehensively compared single AI-based and ensemble methods (e.g., ANN and SVM) methods. They examined these approaches' principles, applications, advantages, and disadvantages. The analysis of past research revealed that 34% of the studies utilized ANN, 24% utilized SVM, and 8% utilized Deep Neural Networks (DNN) for training their models. Only 14% of the studies applied decision trees, while 20% utilized other statistical algorithms. The study highlighted that ANN showed better accuracy than other models, and although DNN performed well, it required actual training data [13], [57], [149].

Despite the promising accuracy of data-driven models, they often overlook the interconnected relationships between building elements, such as the topology of the spaces inside the building. This limitation leads to inaccuracies in building energy consumption estimation (BECE) models. In reality, energy transfer occurs between adjacent spaces with walls, windows, roofs, or floors. For instance, if a space shares a wall with a poorly insulated, cold space, the heating loss rate increases significantly from the warm space to the cold space. Therefore, it is crucial to consider these interconnected relationships in order to achieve more accurate BECE predictions [150]. Therefore, the spatial relationship between the spaces is vital in BECE-based analysis, which recent studies ignore. Hence, due to the complexity of topological information in the BIM data (e.g., IFC format). A space-based graph is developed to capture the spatial relationships and energy transfers between spaces. This graph represents the building's spaces, with each space (node) having semantic information assigned to it in the form of vector data. The edges in the graph connect pairs of spaces and capture the spatial relationships where energy transfer occurs. By modeling the building as a graph, we can analyze the energy flow and identify critical spaces more effectively.

Graphs have proven to be a powerful tool in machine learning, allowing for the incorporation of real-world objects and their relationships. Graph-based models facilitate knowledge extraction and prediction of various phenomena. In the context of building energy consumption analysis, leveraging graph-based approaches enables a more comprehensive understanding of the interconnectedness and energy dynamics within a building [9]. This paper proposes a graph-based classification algorithm called Node-Edge GraphSAGE (NE-GraphSAGE) to consider space information and its adjacency in learning. A GNN-based approach that includes a new aggregator function to leverage nodes (space) and edge (topology) features instead of only node features in traditional BECE data-driven models to overcome the mentioned limitation. NE-GraphSAGE is a machine learning model for space-based BECE classification that applies space properties and topology information in the learning process. To the best of our knowledge, our approach represents the first successful and extensively evaluated implementation of GraphSAGE for BECE space-based classification, providing a practical solution in this domain.

Also, accurate models such as ANN and GNN lack transparency for BECE analysis. Due to their lack of interpretability, data-driven models are often considered black boxes, limiting their trustworthiness and applicability in critical applications that require a clear understanding of the underlying mechanisms behind predictions. In order to trustfully deploy GNN models, it is necessary to provide both accurate predictions and human-intelligible explanations, especially for the architecture, engineering, and construction (AEC) industry. These facts raise the need to develop an explanatory model for BECE analysis to explain why energy consumption prediction results.

Recent researchers have often faced a trade-off between accuracy and interpretability when selecting a model. Some advanced models, while accurate, can be challenging to interpret, while simpler models like logistic regression or decision trees provide more uncomplicated explanations but may sacrifice some accuracy. However, basic models like logistic regression or decision trees have limitations in terms of their predictive power. To improve accuracy, more complex models may utilize a large number of decision trees, often in the form of ensemble methods, and combine their results with other models. On the other end of the complexity spectrum, deep learning models, including graph neural networks (GNNs), consist of multiple interconnected layers that capture higher data abstraction levels. These complex models offer greater flexibility and can achieve high levels of accuracy that simple models cannot match. However, the trade-off is that the inner workings of these models may become less interpretable, making it difficult to understand the rationale behind their predictions. Even the data scientists who designed and trained the complex model can no longer explain the result, such as a space classification problem in a building assigned to an energy-efficient or inefficient class. Understanding GNN predictions is important and valuable for multiple reasons: (i) it enhances trust in the model, (ii) it provides transparency for building designers and decision-makers, and (iii) it enables the identification and correction of model errors and patterns before real-world deployment. Additionally, understanding the network characteristics, including topology information, empowers designers to gain insights and make informed decisions.

While few models explain graph-based neural networks, GNNExplainer is a recent interpretable method designed for GNNs. It learns a mask on edges and features to create a subgraph summarizing the connections and features influencing a node's prediction. However, GNNExplainer only interprets node features, not edge topology, which is crucial for our research.

This paper proposes a GraphSAGE explainable model by integrating the SHAP method [151] with the NE-GraphSAGE prediction model to address the issue of interpretability. The SHAP method interprets the results by assigning contribution values to node and edge features in the space classification task. SHAP values are a computational approximation of Shapley values, known for their properties of additivity and consistency. By leveraging the descriptive nature of SHAP, the proposed method offers a promising approach that combines model complexity and accuracy with intuitive explanations for each predicted space efficiency class. This allows for a more comprehensive understanding of the model's behavior and predictions.

4.4. BACKGROUND AND RELATED WORK

4.4.1. Graph Representation of BIM

Graph representation of BIM is an emerging research area in the construction industry [9], [29], [150]. It improves interoperability between different disciplines and facilitates accessing architectural, structural, and mechanical design knowledge from the BIM data. Despite extensive studies for extracting semantic, geometry, and topology information from BIM digital data models such as IFC, creating a data model that preserves all three types of information remains poor due to their complexity and incompatibility with other data models in the architecture, Engineering, and Construction (AEC) industry. The semantic, geometry, and topology information are preserved by mapping the IFC data model to the graph domain. In the context of the IFC standard model, the concept of indoor spaces and their spatial relationships has been described, enabling the creation of a graph directly from BIM for various indoor applications [9]. Researchers have proposed different approaches to represent BIM models as graphs. Combining shared geometry areas and energy resistance values in one approach introduced a new topological relationship between spaces in different stories [9]. This resulted in generating a weighted adjacency matrix for spaces, which captures the relationship between spaces and their corresponding weighted values. The proposed weighted space-based graph effectively reduces the complexity of the original IFC model. It is compatible with graph-based machine learning algorithms for analyzing Building Energy Consumption Estimation (BECE). Khalili & Chua [152] proposed a graph data model derived from BIM, where nodes and edges were created based on objects and their topological relationships. The model included attributes such as material type, geometry, functionality, and more associated with the nodes and edges through a semantic data table. This approach provided a flexible mapping of Industry Foundation Classes (IFC) data to the graph domain, enabling effective representation and analysis of BIM information within a graph structure. Pauwels & Terkaj [151] developed a tool called ifcOWL, which is an EXPRESS-to-OWL conversion tool. This tool transforms models based on IFC standards into a widely used Web Ontology Language (OWL) ontology format. By utilizing ifcOWL, models can be represented and processed using OWL-based tools and frameworks, enabling interoperability and semantic reasoning capabilities for IFC-based data. Similarly, Simeone Cursi [153] employed semantic web technology to facilitate the mapping, comparison, and transfer of data within the BIM environment. They developed an ontology called the BIM Semantic Bridge, a knowledge base for organizing and integrating BIM data. By utilizing semantic web technologies, the BIM Semantic Bridge enables enhanced data interoperability, semantic reasoning, and knowledge sharing among stakeholders involved in the BIM process. In contrast, Ismail et al.[147] developed a methodology to convert the Industry Foundation Classes (IFC) schema and individual building models into meta and object instance graphs, respectively. However, the object instance graphs did not include part information such as geometry. The conversion process involved transforming IFC files into a CSV format and importing them into Neo4j, a graph database management system. Neo4j provides a transactional backend for applications and complies with ACID (Atomicity, Consistency, Isolation, Durability) principles.. [154]. Finally, the set of tools was integrated and placed on loud, called IfcWebServer [155]. Therefore, mapping the BIM model as object-oriented data, including 3D geometry, semantic, and 3D spatial relationships (topology) to the graph data model, is complex. Previous research studies show the necessity of an intermediate object-oriented data model for this conversion.

4.4.2. Graph Neural Network (GNN)

In recent years, a large amount of data has become available due to developments in the AEC industry. Compared to traditional data exchanging, transferring, and storing techniques, modern digital data types such as BIM can be more reliable and suitable for further design, development, maintenance, and retrofitting. Therefore, after the proper data collection and engineering processes, the data are hugely beneficial and considerably impact reliable design and development. Therefore, the ability to leverage statistical models and machine learning algorithms for data-driven solutions is one of the essential mechanisms in this industry. Furthermore, rather than traditional data storage, which stores just the geometry and attributes of building entities, BIM includes topology (relationship) information, especially for indoor space-based analysis.

Therefore, this opens the potential for including topology information in learning algorithms. Instead of directly processing BIM models, mapping the BIM to graph data models leads to feasible solutions for applying graph-based learning algorithms such as GNN in AEC applications. GNN, a type of Neural Network, is designed to process graph structures directly. One common application of GNN is node classification, where each node in the graph is assigned a label, and the goal is to predict the labels of unlabeled nodes without having access to ground truth information [156]. The other applications of GNN are edge prediction for graph classification. GNN is an efficient method for applications in which the neighborhoods have an essential effect on each other, such as optimizing indoor navigation and predicting space-based (space-based) energy consumption. The reason is that Graph Neural Networks (GNNs), such as Graph Attention Networks (GAT) [24] and GraphSAGE [157], generate new embeddings by taking into account both the node itself and its neighbors. These GNNs utilize the concept of embedding, a technique for transforming properties into low-dimensional, dense, and continuous vectors [158]. In other words, embedding involves generating a vector representation based on the features and attributes of a graph while attempting to preserve as much graph information as possible.

Jin et al. [150] present a graph-based unsupervised method to obtain functional knowledge from building space structures. They used the space properties and their boundary relationships for space clustering. Wang et al. In [150], novel GNN algorithms were designed for the semantic enrichment of the BIM model. They improve the GNNs method for both node and edge features. The research shows a promising avenue for typical space classification tasks by adopting graphs and the GNNs algorithm. Collins et al. [125] developed a novel GNN in the AEC industry. They apply the Graph Convolutional Network (GCM) algorithm to classify the point cloud objects and enrich IFC models by semantic segmentation. Although there have been research efforts to utilize graph-based learning algorithms in the Architecture, Engineering, and Construction (AEC) industry, the application of Graph Neural Networks (GNNs) remains relatively limited. The potential of GNNs in the AEC domain has not been extensively explored and studied in detail. While there are a few existing studies on the use of GNNs in the AEC industry, more comprehensive research is needed to fully understand and harness the capabilities of GNNs for various applications in this field.

4.4.3. Explainable data-driven models in the AEC industry

Data-driven models created by machine learning have gained significant attention in the AEC (Architecture, Engineering, and Construction) engineering fields, offering potential assistance in design, development, retrofitting, and maintenance. These models promise to improve building performance and sustainability [159]. However, their limited generalization and black-box nature hinder their exploitability and reusability, limiting their practical application and understanding. Therefore, transparency is important, given the current design concerns surrounding data-driven models. In particular, data-driven solutions are becoming more prevalent in the AEC industry, and their decisions can bear significant consequences. Explainable data-driven methods could help eliminate the ambiguity of the results and engineering bias in decision-making processes.

Furthermore, this enables engineers to understand the models to effectively manage the benefits of data-driven methods while maintaining high prediction accuracy. By decomposition of machine learning algorithms, the contribution of each building's elements can be considered in explainable models. It also can be used as an innovative tool for engineers to improve the design before applying the model, reducing the cost of the building development process.

4.5. METHODOLOGY

This section introduces the process of the BECE space-based graph construction, followed by a description of the proposed GNN method. Subsequently, the experimental results and comparative analysis with the other classification methods are discussed in greater detail. Finally, an explainer method using SHAP values adapted from Wang et al. in [160] is introduced to interpret the results of the proposed GNN classification method.

4.5.1. Space-based Graph Representation

This research employs IFC as a well-known representation of BIM data format to generate a new graph called a space-based 3D graph to store the space geometry, semantic, and topology information in a graph data model. In this graph, the node represents space (spaces). The geometry and properties of each space (node) are defined as vector information assigned to each node, such as Window-Wall area Ratio (WWR), Exterior-Wall to total wall area Ratio (EWR), Interior-Wall total wall area Ratio (IWR), Floor to total Space area Ratio (FRR), Roof to total Space area Ratio (RRR), Exterior Wall R-Value, Exterior Window R-Value (Resistance Value), Roof R-Value, and Floor R-Value. The edges in the graph connect a pair of spaces and capture the spatial relationship if there is any energy transfer. Energy transfer between adjacent spaces impacts the energy demand of each space by influencing heat exchange and overall thermal dynamics. For example, if a space has a shared wall, floor, or roof with a cold space (with a low energy efficiency space), the heating loss rate increases significantly from the warm area to the cold area. Therefore, the edges in this graph represent the energy transfer relationship between spaces.

Furthermore, the rate of energy transfer between two spaces is related to some parameters, such as the common area and the material Resistance Value (R-value - Resistance to heat flow through a given thickness of material [161]) of the common entities between the spaces. Accordingly, a vector with the dimension of 1×2 is assigned to each vector, which comprises the entity's common area and resistance value between two spaces. The space-based graph is a 3D graph representing the relationship between the spaces in the same story and the other stories. Figure 4.1 represents a sample of spaces, the space-based graph, and the feature vectors assigned to nodes and edges.

4.5.2. Energy Use Intensity (EUI) Calculation

This research uses the ground-truth value as the reference value for training the proposed classification model and evaluating its accuracy. To simulate the energy in the case study BIM data, we used RETScreen software, an energy modeling package developed by the Canadian government. Energy modeling, or energy system modeling, involves the creation of computer models to analyze energy systems and their components. It enables the evaluation and simulation of energy-related scenarios and the assessment of system performance and efficiency. This research uses energy modeling inputs such as building thermal envelope characteristics, ventilation loads, equipment efficiencies, lighting power densities, and plug loads from ASHRAE 62.1 and ASHRAE 90.1 standards for the same climate space [162]. The Energy Utilization Intensity (EUI) was determined and calculated for each space to evaluate more critical and energy-intense spaces. Energy Use Intensity (EUI) has been defined as the measurement of a space's annual energy consumption relative to the area (kWh/m^2). If the EUI for space A is higher than space B, it indicates that space A consumes more energy per square meter annually. This difference in EUI suggests that various parameters contribute to the higher energy consumption in space A. The main effective parameters which are considered in our research are summarized below:

Exterior Wall Thermal Resistance. Thermal resistance across the wall is the sum of the resistances of the individual layers. Also known as R-Value, the higher the R-Value, the more resistance the energy transfer. Windows Thermal Conductivity. The overall coefficient of heat transfer quantifies the thermal conductivity of a window, representing the amount of heat gained or lost between the indoor and outdoor environments based on the temperature difference. Also known as the U-factor, the lower value represents the window's thermal performance better.

Space's Location (Space Location): Spaces with exterior walls and windows or adjacent to an interior space with different temperature requirements have higher energy consumption due to heat transfer than internal spaces with isometric walls. Window-to-Wall Ratio. Basically, spaces with more windows in their exterior walls require more heat in winter and cooling in summer to offset the heat transfer. Spaces with different space types might have different lighting requirements. For instance, offices require more lighting than storage or rest space types. Space type or space functionality is essential for energy consumption in each space. For example, ASHRAE 62.1 mandates mechanical designers to provide fresh air to different spaces based on the occupancy density presented in the standard. For instance, meeting spaces have an occupancy intensity of 50 people/100 m2 compared to 5 people/100 m^2 for offices. This means that meeting spaces require more energy than offices to provide fresh air. Also, designers need to provide more cooling in more occupied spaces because of the internal heat gain of 130 W/person. Total energy consumption is calculated based on the mentioned parameters, and the EUI value is calculated by dividing the total energy consumption of a building by its corresponding floor area. This ratio provides a standardized energy efficiency measure and allows for comparison between different buildings or spaces(Appendix D).

4.5.3. NE-GraphSAGE algorithm for Space Classification

This research aims to address the space classification task related to energy consumption by employing a graph-based learning algorithm that incorporates the spatial relationships (topology) between spaces. The aim is to identify and classify critical spaces within the building based on their energy consumption patterns. The space classification task is the type of GNN node classification in BECE analysis. For BECE-based classification, we applied an inductive GNN method based on GraphSAGE. In this research, the choice of an inductive method over a transductive method is motivated by the need to build a generic model capable of predicting new nodes (spaces) based on observed training data. Transductive learning, like GCN, constructs a model specific to the training and testing data it has already encountered, which is unsuitable for our research application. In retrofitting processes, where buildings undergo structural changes and new spaces are added or demolished, it is crucial to have a model that can accommodate new data and spaces. Thus, an inductive approach allows us to design and train a model for such scenarios.

GraphSAGE is an inductive learning algorithm leveraging node features (e.g., node profile information, node degrees) to generate node embeddings [157]. The main idea of GraphSAGE is to consider features from the local neighbors of a node. Specifically, the GraphSAGE forward propagation algorithm feeds the neighbor features into an aggregator function (e.g., mean, pooling). Then, its output updates the node in the next layer (or depth). Therefore, it considers node features in the process of embedding and training. In BECE-based classification, both the node features (attributes of spaces) and edge features (relations between spaces) play an essential role in identifying critical spaces with inefficiency areas; thus, designing a GNN algorithm that can process both nodes and edge features is desired. However, the GraphSAGE algorithm is limited to node-level features, which is insufficient for this task. Therefore, this research improves the GraphSAGE algorithm by introducing the Node-Edge GraphSAGE (NE-GraphSAGE) method, which involves node and edge features in the training and classification process.

NE-GraphSAGE aims to learn a representation for every node and corresponding edge based on some combination of its neighboring nodes, parametrized by h for the node and e for the edge. Recall that every node can have its feature vector parameterized by X. Each edge has its feature vector parameterized by Y. Let us assume that all the feature vectors for every node and edge are the same size. One layer of NE-GraphSAGE can be run for k iterations. Parameter k controls the neighborhood depth. If k is 1, only the adjacent nodes are involved in the learning process. If k is 2, the nodes at walk depth two are considered.

$$\boldsymbol{h_v^{k-1} = h_v^0 = X_v}$$

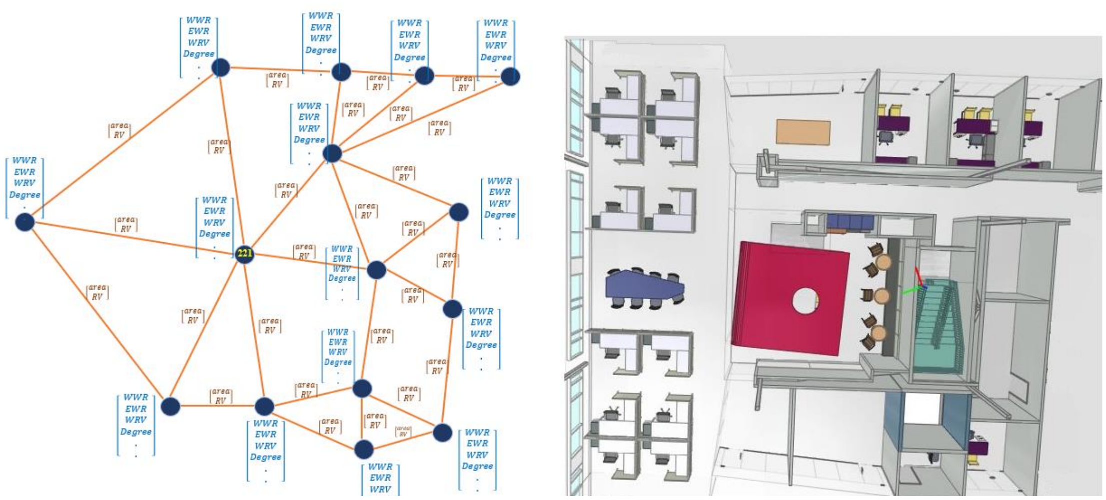


Figure 4.1. Sample building layout and corresponding graph representation

The following statements represent the definitions of the notations:

$\boldsymbol{X_v} = Node\ features\ for\ a\ node\ v$

$\boldsymbol{Y_{uv}} = edge\ features\ for\ \ node\ u\ and\ v$

$\boldsymbol{h_v^k} = embedding\ features\ node\ v\ in\ neighbour\ depth\ k$

$\boldsymbol{e_{uv}^k} = embedding\ features\ edge\ uv\ in\ neighbour\ depth\ k$

$\boldsymbol{N_v} = neighbour\ set\ of\ Node\ v$

$\boldsymbol{Z_v} = target\ node\ embedding\ represenation$

Remark that having k = 2 means nodes at neighborhood depth two can affect each other through the node in the middle; therefore, there is a node (space) representation h for every node at every k iteration. The value of k is determined experimentally using multiple neighborhoods. We can construct a computation graph for each node that represents the k-hope neighborhood graph of a target node. The computation graph represents the neighbor nodes of each space and the topology information between them. Figure 4.1 illustrates an example of a computation graph of target node 3 with a neighborhood depth of two. The NE-GraphSAGE algorithm follows a three-step process, starting with an initialization step. This step sets all the initial node embedding vectors to their respective feature vectors. The algorithm then iterates over the steps, with k representing the iteration number.

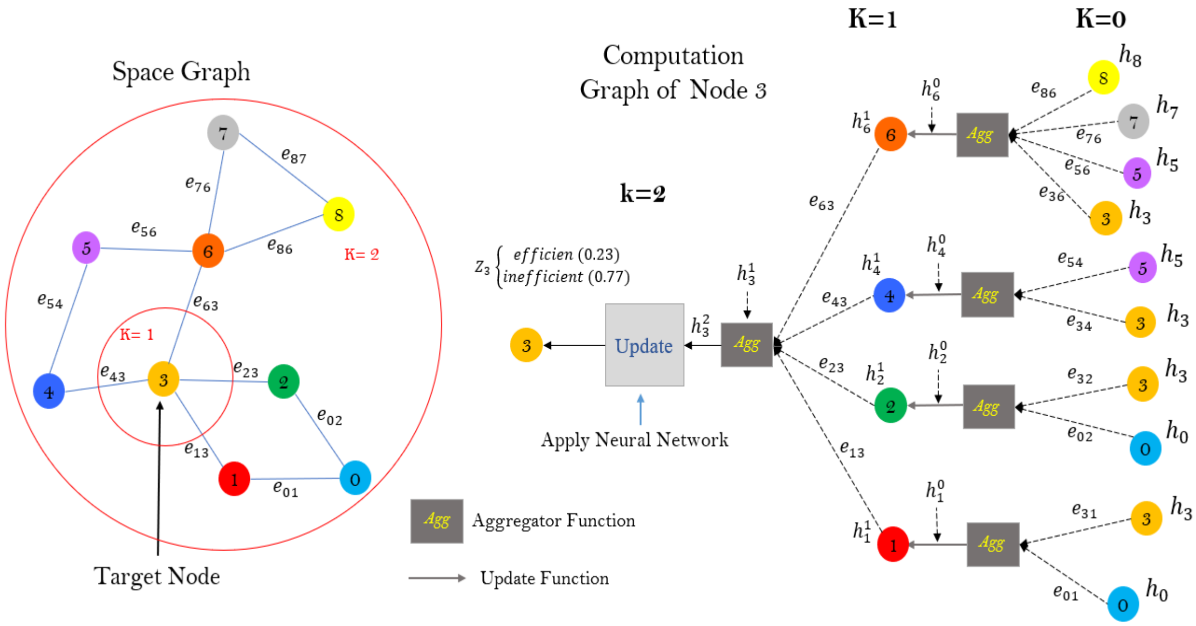


Figure 4.2. Computational Graph of Space 3 with the neighbor depth of 2

4.5.3.1. Information Aggregation

Once the neighborhood has been defined, the next step is to establish an information-sharing procedure among neighboring nodes. To achieve this, a computational graph is created for each node, which enables the calculation of new embedding (feature) values for the target node. In this process, the computational graph considers the features of the target node as well as the features of its neighboring nodes. By incorporating information from the neighboring nodes, the graph can update and refine the embedding of the target node, capturing its surrounding context's collective knowledge and characteristics.

This information-sharing procedure plays a crucial role in graph-based learning algorithms, such as GraphSAGE or Graph Attention Networks (GAT), as it allows for the propagation of information across the graph and facilitates the generation of more accurate and meaningful embeddings for each node. In this research, we design aggregation functions or aggregators that accept the features from neighborhood nodes and edges as input and aggregate the neighbors' attributes (features) to create a neighborhood embedding for the target node. We first initialize all node and edge embeddings to node features as the node and edge attributes to learn embeddings with aggregators. Then, for each neighborhood depth until K, we create a node and edge embedding with the aggregator function. This aggregation method helps us employ the effect of the neighbor spaces on energy consumption in the learning algorithm process (NE-GraphSAGE). Also, employing the edge's feature in the learning process, the topology information is applied based on the common area and R-value between neighbor and target spaces. Different aggregation functions are the LSTM (Long Short-Term Memory) aggregator, Pooling aggregator, and Mean aggregator [157]. We have chosen the Mean aggregator for this research because of its simplicity in implementation. Eq 4.1. demonstrates the proposed aggregation function in which h^{k-1} shows the feature values of the neighbor nodes, e^{k-1} demonstrates the feature values of edges between neighbor nodes and the target node. In the aggregation process, the mean of neighbor nodes and edge features are calculated, and then two result vectors are concatenated in-depth k-1.

$h_{N(v)}^{k}$ is the aggregated node and edge representation of node v considering its neighboring nodes and edges in depth k.

$$h_{N(v)}^{k} \leftarrow Concat(Mean_{N(v)}(h_{u}^{k-1}), Mean_{N(v)}(e_{uv}^{k-1}) \qquad \text{(Eq 4.1)}$$

Finally, we calculate the embedding features of node 3 by aggregating the feature values of its neighbor nodes described in the next section.

4.5.3.2. Feature Updating

After obtaining an aggregated representation for node v based on its neighbors, update the features of node v using a combination of its previous representation (${h_v}^{k-1}$), the aggregated representation (${h_{N(v)}}^{k-1}$) includes the edge representation (e_{uv}^{k-1}). This research proposed a new updating and aggregation function to consider both node and edge features during learning. Equation 2 demonstrates the updating function by concatenating the aggregated representation in depth k-1 and target. Node 3 is an example (Figure 4.2) representing the proposed algorithms with neighbor nodes of $\mathrm{h}_6^1, \mathrm{h}_4^1, \mathrm{h}_2^1, \mathrm{h}_1^1$ and the feature between the target node and neighbors ($\mathrm{e}_{63}, \mathrm{e}_{43}$, $\mathrm{e}_{23}, \mathrm{e}_{13}$). By aggregating neighbor nodes and edge features, the vector feature of h_3^1 is feeding into the update function in layer 2 (k=2). In the next step, the vector feature of Node 3 in Layer 2 (h_3^2) and aggregated vector feature of node 3 from layer 1 (h_3^1) are concatenated. The non-linearity (σ) is applied to the result as an activation function. This research applies the Sigmoid activation function [163] because its output is between 0 and 1, which is suitable for our space binary classification output. The algorithm is implemented in Python script with PyTorch, NumPy, DGL, Panda, and sklearn libraries in the Google colab environment. Then, the concatenated vector is passed to the neural network layer as the last step. Figure 4.3. shows the pseudocode of the embedding generation of each node in the depth walk k.

4.5.3.3. NE-GraphSAGE Training

In this step, a Multi-Layer Perceptron neural network (MLP) is designed as a learnable updating process (Figure 4.4) to update the target node embedding to determine the probability of a critical space of spaces discussed in the next section. In order to train the neural network and optimize its weights, a differentiable loss function is required. In this study, the Squared Error Loss (SE) function [164] is chosen to calculate the distance between the actual value of the node class (0 or 1) and the predicted values. The Squared Error Loss function measures the square of the difference between the predicted and actual values, penalizing more significant errors more significantly. The neural network aims to minimize the discrepancy between its predictions and the ground truth labels by minimizing this loss function during training. By utilizing the Squared

Error Loss function, the neural network can effectively quantify the error in its predictions and adjust its weights to minimize it, leading to improved accuracy and convergence during training. Indeed, compared to other node classification benchmark datasets, the space-based graph in commercial buildings typically does not have many nodes or node features. Commercial building graphs are often smaller in scale and may have fewer nodes and edges than datasets used in other domains. While this may limit the direct applicability of specific graph-based learning algorithms designed for larger graphs, it also presents an opportunity to explore tailored approaches and optimizations specifically suited for the characteristics of space-based graphs in commercial buildings.

Input: Space-based Graph $G(v, e)$; $input\ features\ \{X_v, Y_{uv}, \forall v \in V\}$; $walk\ depth\ K$; $non-linearity\ \sigma$; $aggregator\ function\ AGG_k\ , \forall k \in \{1, \dots. K\}$; $neighborhood\ function\ N\ \rightarrow v\ \rightarrow 2^k$

Output: Vector represenation $\boldsymbol{Z_v}$ **for all** $\boldsymbol{v} \in \boldsymbol{V}$

$h_v^0 \leftarrow\ x_v, \forall v\ \in V;$

For $k = 1 \dots K$ ***do***
  for $\boldsymbol{v} \in \boldsymbol{V}$ ***do***

$h_{N(v)}^k \leftarrow [Concat(Mean_{N(v)}(h_u^{k-1}), Mean_{N(v)}(e_{uv}^{k-1})]$

$h_v^k \leftarrow\ \sigma(Concat(h_v^{k-1}), h_{N(v)}^{k-1})$

  end

Figure 4.3. The pseudocode of the embedding generation of each node in the depth walk k.

Unlike other approaches, the space-based graph in this research represents the entire building as a single 3D graph, with each space being represented as a small node and the spatial relationships between spaces as edges. This compact graph representation allows for a holistic analysis of energy consumption patterns within the building, comprehensively considering the interconnectivity of spaces. Therefore, we considered all nodes (spaces) in one patch for the learning process. Instead of involving each node's feature in the learning step, the embedded vector of the updated feature vector (section 4.5.3.2) is considered input of the neural network to use the advantage of neighborhood and topology information in the learning process. A two-layer MLP neural network method is designed in which the weighted matric is shared between all computation graphs in the last depth (k). Figure 4.5 represents a sample computation graph for Nodes 4 and 5 in which the neural network parameters are shared.

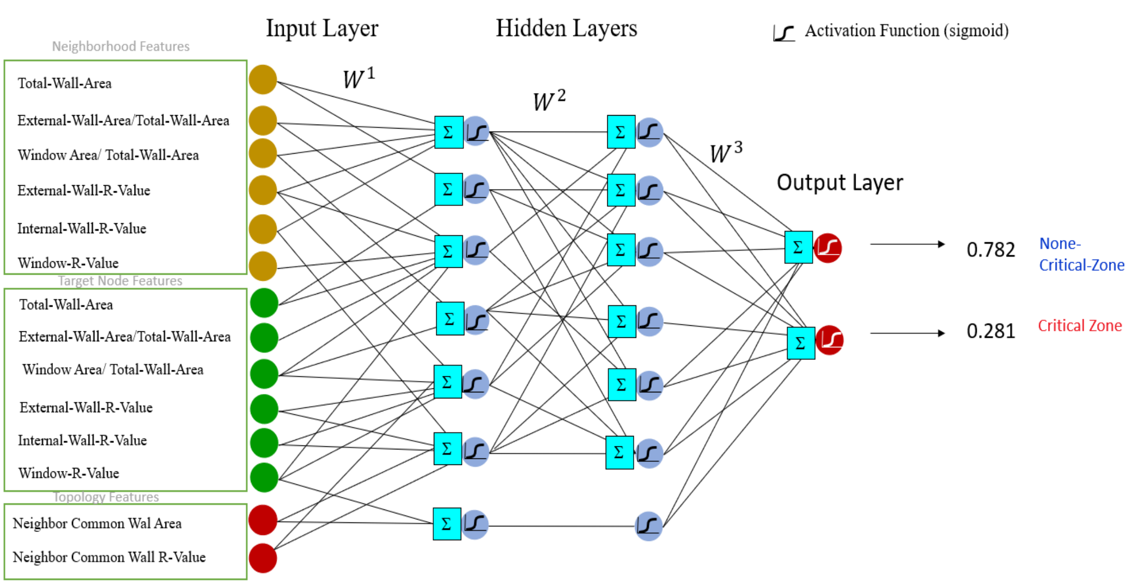


Figure 4.4. MLP neural network design for space classification

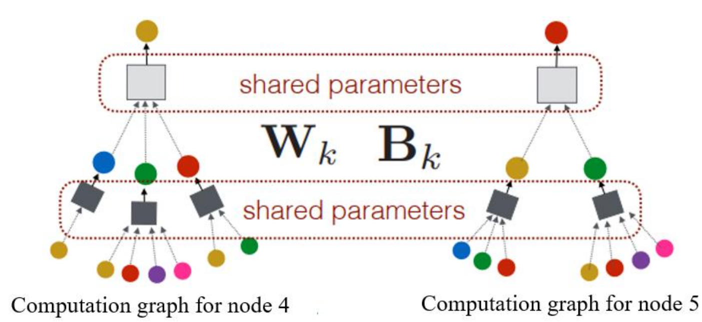


Figure 4.5. Sample computation graphs and shared weighted matrices

Therefore, the generated computation graph of each building's space is applied in a batch learning process. Indeed, we employed the Batch Gradient Descent (BGD) method for the learning process to update the shared weighted matric [165]. We calculate the average gradient across all training examples to update the parameters and use that mean gradient for parameter updates. Therefore, there is just one gradient descent step in one epoch (Figure 4.6). The distance between the ground truth and the predicted output value is measured by Mean Squared Error (MSE), the function. . MSE (Eq 4.2) is calculated for each node iteration (epoch). At the end of each epoch, the neural network's weighed matrices are adjusted by the backpropagation process [166]. Then, the learning process is continued iteratively to catch the minimum MSE value.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{Y_i})^2 \qquad \text{(Eq 4.2)}$$

To recognize which parameter(s) have more impact on the result of the space classification as a critical space or non-critical space, we need a method to measure and score the contribution of each feature (space properties or neighbor's relationship) for each node (space). For this purpose, we combined the accurate NE-GraphSAGE method and the explainable ML method SHAP (Shapley Additive exPlanations) to study the critical spaces of buildings.

In a predictive model, Shapley values (Eq 4.3) represent the contribution of each input feature to the prediction of each space. Shapley's values consider the outcomes of all possible combinations of features in a space or edge to determine the importance of each feature. It considers each feature's contributions in different feature combinations and evaluates their overall impact on the prediction. The set corresponds to each possible combination of space's features, the neighbor embedder's features, and the edge's features contribute to calculating the Sharpley value.

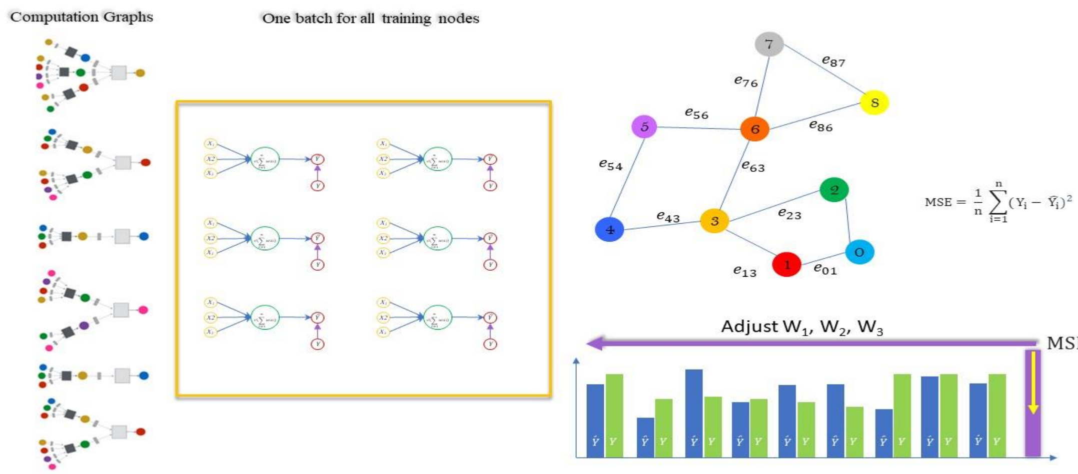


Figure 4.6. Batch gradient descent of the learning process

In this way, each feature with a higher Sharpley value contributes more to the final result. We interpret the features of all spaces marked as critical Spaces in the classification model to demonstrate more insight into architecture and engineering for accurate decision-making.

$$\emptyset_i(p) = \sum_{S\subseteq N/i} \frac{|S|!(n-|S|-1)!}{N!}(p(S \cup i) - p(S)) \qquad \text{(Eq 4.3)}$$

4.5.3.4. Model Interpretation

In addition to accurately applying EN-GraphSAGE for the space classification model in two critical and non-critical spaces, we must understand and interpret how and why our models make their predictions. The knowledge gained from the prediction and interpretation models can help planners and engineers develop more effective strategies and manage the demand for energy in different spaces in buildings. To better understand the classification result, we employ a method to investigate the role of each input parameter.

The Sharpley values are calculated for features belonging to the spaces marked as critical to explaining the role of each feature in the final output. NE-GraphSAGE can explain the model's output by calculating the probability of each space (space). If the probability is closed 1, it can be considered a critical space. If it is close to 0, it will be marked as non-critical space. This is called global interpretability. Next, we explain why each space receives the prediction according to its specific predictor (features) values. This is called local interpretability. The Sharply values can show both. For example, Figure 4.7 represents the force plot of a test node (Space 154) in the first story marked as a critical space.

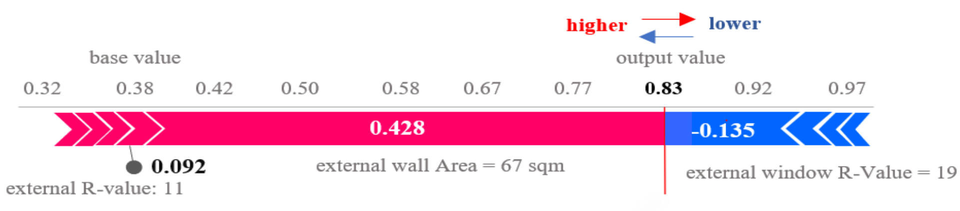


Figure 4.7. SHAPE force plot

The Sharpley value for Space 154 is calculated for all features. The Features with high Shapley values have a more significant impact on model output (0.83 in Figure 4.7), and features with low Shapley values have less impact on the prediction. This is because the features with the positive Shapley value force the prediction to the output class, and the features with a negative Sharpley value push the output to the other class. In this example, the External Wall Area (EWA) and External Wall R-Value (EWR) have the most positive Sharpley values of 0.428 and 0.092, which impacts the energy consumption of Space 154 and push it into the critical space class. The emergence of this insight provides valuable information for architects and engineers, enabling them to focus on optimizing the wall geometry design and enhancing the insulation materials used in Space 154 to reduce its energy intensity. Additionally, the negative Shapley value associated with the External Window R-Value (EWINR) suggests that the window material insulation is already in good condition and does not contribute significantly to the space's energy consumption.

4.6. EXPERIMENTAL RESULTS AND DISCUSSION

We apply the proposed methodology to a three-story IFC model collected from open IFC model datasets. The results of the proposed classification model and the interpretation models used to explain the classification results are presented in the following sections.

4.6.1. NE-GraphSAGE Classification Model

The BIM data as the case study is downloaded from the Open IFC Model Repository in IFC format. It includes 130 spaces with different activities, including an office space, conference space, lounge, kitchen lobby, and washroom. The Energy Use Intensity (EUI) (Ma and Cheng 2016) is considered an energy consumption index in this research. First, a ground-truth value is calculated for all spaces for three stories using RETScreen software to perform energy modeling for multiple spaces in this paper. The Energy Use Intensity (EUI) measures annual energy consumption per square foot. It is obtained by dividing the total energy consumed by the building in a year by its total floor area. Next, the total energy is estimated for each space using RETScreen software by considering space geometry information, material information, heating and cooling energy to keep the comfortable temperature, thermal envelope characteristics, ventilation loads, equipment efficiencies, lighting power densities, and plug loads were taken from ASHRAE 62.1 and ASHRAE 90.1 40 standards for building climate spaces.

Then, Z-score 41. is calculated for EUI values to find an outlier. The spaces with outliers EUI values are considered critical spaces. In the last step, EN-GraphSAGE is employed to classify the nodes on the space-based graph and classify the spaces in critical and non-critical spaces. The best accuracy achieved during the range of epochs from 1000 to 5000 was 91.26%, occurring at epoch 3500. Therefore, we pick up the weight matrices in epoch 3500. To assess the accuracy and correctness of each classifier, we utilized two commonly used metrics: accuracy and F1 score.

These metrics provide a quantitative measure of the classification performance. We evaluated the performance of four well-known classification methods: NE-GraphSAGE, GraphSAGE, ANN, and SVM. Accuracy is a reliable criterion for assessing overall correctness, as it represents the fraction of correct predictions over the total number of samples, regardless of their respective classes. It is calculated using Eq 4.4, where the numerator represents the number of correct predictions and the denominator represents the total number of samples. The F1 score is another important metric that considers both precision and recall. It provides a balanced measure of the classifier's performance by considering positive and negative class predictions. However, the specific calculation for the F1 score is not mentioned in the given context. By evaluating the accuracy and F1 score of the four classification methods, we can gain insights into their performance and determine their effectiveness in the context of the problem being addressed. Accuracy may be biased towards predominant classes, so we used the F1 score as an evaluation metric to assess balanced predictions. The F1 score calculates each class's harmonic mean of precision and recall and then takes the arithmetic mean across all classes (Equation 4.5).

$$Accuracy = \frac{\sum_i TP_i}{N_n}$$ TP = true positive, N_n= number of nodes (Eq 4. 4)

The accuracy and F1 are calculated for four graph-based (NE-GraphSAGE, GraphSAGE) and none graph-based (ANN and SVM) methods to compare the result of the proposed method with other classification methods (Table 4.1).

$$F1 = \frac{\sum_{i=1}^{n} \frac{2 \times TP_i}{2 \times TP_i + FP_i + FN_i}}{n} \qquad \text{(Eq 4.5)}$$

TP = true positive, FP = false positive, FN = false negative, n = number of classes

Table 4.1. Classification Models Accuracy and F1Score

Model	Accuracy	F1
GraphSAGE	79.5	0.73
NE-GraphSAGE (2 layers)	86.6	0.82
NE-GraphSAGE (3 layers)	91.2	0.88
NE-GraphSAGE (4 layers)	86.9	0.83
ANN	72.6	0.67
SVM	71.8	0.66

We prove that the proposed classification method (NE-GRAPHSAGE (3 layers) provides significant gains (12% on average) compared to other classification methods, especially non-graph-based methods, which have the highest accuracy and F1 from the experiment results, arriving at 91.2% and 0.88, respectively. On the other hand, NE-GraphSAGE (2 layers) and NE-GraphSAGE (4 layers) have slightly lower accuracy, and F1 has less balanced prediction compared to NE-GraphSAGE (layers3). We also trained GraphSAGE with three layers for a comprehensive comparison. We used the same dataset and fine-tuned GraphSAGE to obtain its best results. The difference between the two algorithms is that NE-GraphSAGE can consider both node and edge features during updating embedding, while GraphSAGE only learns from node features. As a result, NE-GraphSAGE (3 layers) improved the accuracy by 12% and the F1 score by 0.15 compared with GraphSAGE. Also, the result recognized the effect of edge features (topology features) in finding the critical spaces. For example, six interior spaces are wrongly marked as non-critical spaces by non-graph-based methods, but two graph-based methods classified these spaces correctly as critical spaces. This is because these spaces are in the neighborhood of critical spaces with a low R-value of shared walls on the first and third floors. However, they are classified as non-critical spaces for those models where the topology information is ignored in the learning process.

Consequently, NE-GraphSAGE (3 layers) generally has the highest accuracy and most balanced prediction performance. Therefore, NE-GraphSAGE (3 layers) is chosen as the space classification algorithm in this research compared to other algorithms. Hence, in the remainder of the paper, the term NE-GraphSAGE refers to the 3-layer variant of NE-GraphSAGE.

4.6.2. Interpreting EN-GraphSAGE Classification using SHAP values

The SHAP method is employed on the result of test data to assign a Sharpley value for features in each space considered a critical space. The impact of features on the prediction of critical space probability for each space is measured by evaluating the deviation from the mean predicted value when different combinations of features are utilized. In addition, this paper uses a forced plot to find the importance of features, their interpretation, and the parameter(s) that cause high energy consumption in these spaces. We calculated the Sharply value of features for those spaces classified as critical spaces. As the proposed classification model considered the space's properties and adjacency information, the Sharpley values can be calculated for both feature sets. Therefore, the SHAP method in this research can pinpoint which parameters of the space and its adjacent spaces are most impactful for high energy consumption. For every space, the features are ordered by their Sharpley value. Some of those features negatively affect energy consumption, and these form the basis for expert recommendations in process design or retrofitting to increase energy efficiency. Figure 4.8 represents the force plot for Space-153, which shows the Sharpley value normalized between 0 and 1.

Considering all features, the Sharpley value for Space 153 is estimated at 0.71. The neighbor common R-value is the primary parameter pushing this space as a critical class. Although the boarding space has no external wall or window, it is pinned as the critical space. It is in the neighborhood of critical space 154, and a high energy rate is transferred between these two spaces.

Therefore, this space recommends using different window materials (with higher R-value) between them.

Conclusion This paper introduced a new GNN-based classification model (NE-GraphSAGE) to find the critical space using a 3D building model. This model considers the space's properties and the adjacent information in training the classification model. The promising result demonstrates a better model accuracy than the non-graph base and regular GraphSAGE models. In order to address the trade-off between model accuracy and interpretability, the NE-GraphSAGE model is combined with the SHAP (SHapley Additive exPlanations) model. SHAP is a method used for model interpretation that provides explanations for individual predictions. By incorporating SHAP with the NE-GraphSAGE model, the researchers aim to achieve high accuracy and interpretability. This allows for a better understanding of the factors influencing the model's predictions, enhancing the transparency and trustworthiness of the model's results. Instead of choosing between accuracy and interpretability, the proposed methodology finally proposed a solution that lets us push the envelope regarding model complexity and accuracy while still allowing us to derive intuitive explanations for each space. Also, the proposed interpretation model can explain the impact of individual features in a space and aggregated features in 3D adjacent spaces. The result shows that combining the high-accuracy classification model (NE-GraphSAGE) and the SHAP model can help designers and engineers have accurate insight into their design and decision-making process. However, this methodology has two main limitations that need future work.

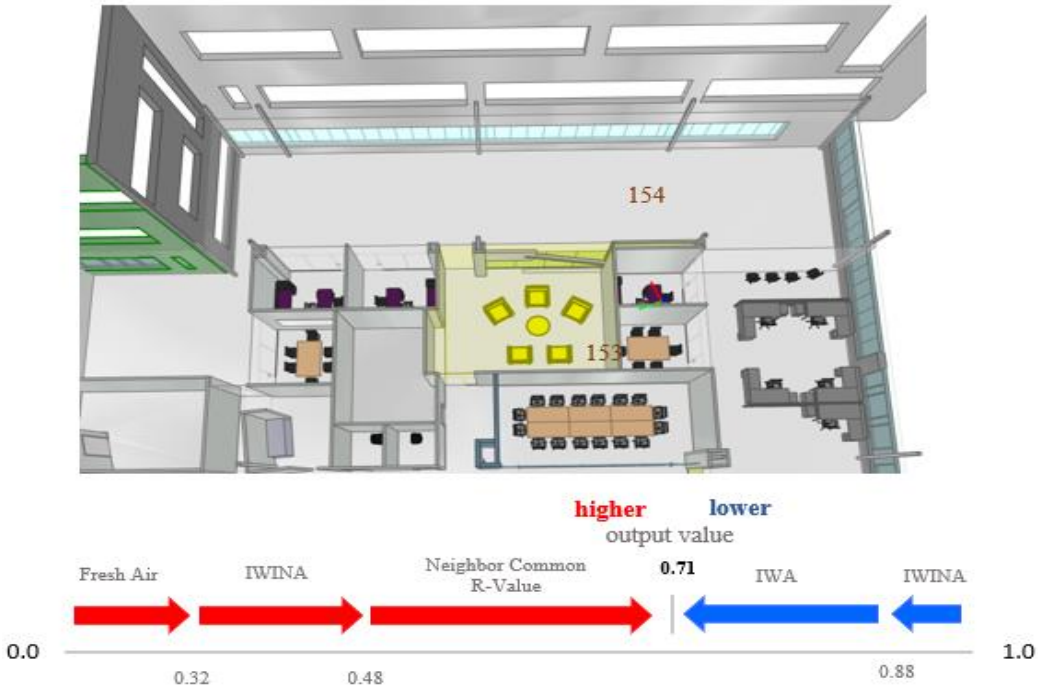


Figure 4.8. SHAP force-plot for Space 153

The first limitation addresses the adjacent interpretation. Although we can find the Sharpley value for adjacent features, we cannot find which adjacent space(s) has more impact on the target space because the adjacent feature values result from the aggregation function. The second limitation is that the knowledge expert should explain the recommendation. The proposed methodology can find and explain the reason for classification results but is limited to using the result for the recommendation and prescription. The latter is ongoing research, and the outcome of the research will be presented in the near future.

CHAPTER 5: AN EXPLAINABLE AND PRESCRIPTIVE SOLUTION FOR SPACE-BASED ENERGY CONSUMPTION OPTIMIZATION USING BIM DATA AND GENETIC ALGORITHM

5.1. PREFACE

This chapter introduces a proposed prescriptive method to address the existing gap in Building Energy Consumption Estimation (BECE) data-driven models. The research tackles this gap by integrating building energy simulation with optimization techniques, utilizing BIM data and Genetic Algorithm (GA) to formulate a prescriptive and explainable model for indoor building design. The study concentrates on space-based indoor design, using BIM interoperability to suggest optimal solutions. The model incorporates the value engineering method to balance energy consumption, functionality, and cost, offering engineers and designers valuable insights for optimizing building performance. The content of this chapter was published in Buildings in 2023 as follows:

H. Kiavarz, M. Jadidi, A. Rajabifard, and G. Sohn, "A Genetic Algorithm as a Prescriptive Solution for BIM Model to Optimum Space-based Energy Consumption," The *Journal of Building Engineering* (JOBE).

Copyright and Licensing:

Contributions:

The contributions of authors in the current chapter are as follows: **Hamid Kiavarz** has conducted the literature review, performed the data collection and curation, developed the method, used the customized coding to perform the analysis and modeling, validated and visualized the results and wrote the original manuscript of this publication. **Mojgan Jadidi** has supervised the research, provided the funding, and contributed to developing the manuscript's method, writing, and editing. **Abbas Rajabifard and Gunho Sohn** have supervised the research and contributed to editing the manuscript.

5.2. ABSTRACT:

The design and retrofitting of energy-efficient buildings pose a complex challenge, requiring an intricate balance between diverse architectural, mechanical, and electrical parameters. Integrating energy performance simulation into the design process is essential with the rise of low-energy building mandates and green construction practices. However, existing methods need help handling vast design information, exploring optimal alternatives, and employing suitable assessment techniques. Contemporary approaches, such as Building Information Models (BIM)

and data-driven solutions, offer promising avenues for energy analysis. Nonetheless, these data-driven models often need more transparency and design optimization capabilities, rendering them insufficient for critical applications in the Architecture, Engineering, Construction, and Operation (AECO) industry. This research addresses the gap by integrating building energy simulation with optimization techniques, using BIM data and Genetic Algorithm (GA) to develop a prescriptive and explainable model for indoor building design. The study focuses on space-based indoor design, leveraging BIM interoperability to recommend optimal solutions. The proposed model employs the value engineering method to balance energy consumption, functionality, and cost, providing engineers and designers with insights to optimize building performance effectively. This approach enhances energy efficiency and offers tangible design optimization solutions, bridging the gap between energy prediction and practical application in the AECO industry.

5.3. INTRODUCTION

The optimum energy consumption design in new buildings or retrofitting existing buildings is a challenging problem in many construction projects [50], [55]. With the increasing emphasis on low-energy building design, decarbonization, and green buildings, there is a growing demand for analyzing building elements and seeking optimal designs through energy performance simulation during the design process. However, the application of design analysis integration in actual building design practice faces several important challenges, including (1) the vast amount of design information and variety of expert knowledge involved, (2) the substantial time and effort required to explore optimal design alternatives, and (3) the need for different assessment methods at each stage of the building design process. Architectural, mechanical, and electrical parameters such as geometry, envelope type, floor plan, materials, HVAC type, and lighting load significantly impact building energy consumption, and these parameters are determined during the design stages

[51]. These days, in the design stage, the interest in using newly available data models such as Building Information Models (BIM) and applying data-driven solutions emerge as the efficient and suitable option for building energy analysis rather than employing classical physics-based models to improve the energy-efficient design in different stages of design, development, and retrofitting [9].

In recent years, data-driven building energy analysis has garnered significant research attention, leading to various classification studies of prediction methods, including elaborate engineering methods, statistical approaches, and artificial intelligence-based models [54]. Despite the predictive capabilities of data-driven models in Building Energy Consumption Estimation (BECE), they often lack transparency, making it challenging to trust their outcomes in critical applications within the Architecture, Engineering, Construction, and Operation (AECO) industry. While Building Information Modeling (BIM) allows for precise digital construction, designing energy-efficient buildings remains complex due to many design variables and performance objectives. Additionally, data-driven models can predict energy consumption but fail to provide inherent design optimization solutions for engineers [5].

To address this research gap, several studies emphasize the importance of integrating building energy simulation with optimization techniques [14]. This integration enables the determination of optimal building designs and the generation of recommendations to enhance energy efficiency. However, challenges arise in dealing with computational complexity and time uncertainty in simulation-based optimization methods. Notably, EnergyPlus® and TRNSYS® are commonly used for simulations, while GenOpt® and MATLAB® are preferred platforms for optimization [33], [167]. Examples include utilizing the Genetic Algorithm (GA) with EnergyPlus to minimize energy consumption in residential buildings and optimizing natural ventilation using sensitivity analyses to identify influential design parameters [167]. Another study focused on developing a methodology to optimize the building facade components for the optimal annual energy cost by configuring insulation, glazing, shading devices, and ventilation systems [16].

Yu et al. [6] employed a simulation-based approach integrating a backpropagation network and a multi-objective genetic algorithm to optimize green building energy efficiency and thermal comfort by identifying optimal design parameters. Weerasuriya et al.[7] utilized various optimization algorithms and decision-making methods to minimize building energy consumption and optimize lift-up design in hot climates, emphasizing the use of Artificial Neural Networks (ANN) for efficient [168] design solutions [8], [9]. Gou et al. [168] coupled Non-Dominated Sorting Genetic Algorithm-II (NSGA-II) with an ANN model to optimize passive parameters in residential buildings, demonstrating the significant potential for energy savings and improved thermal comfort in Shanghai residences. Bre et al. [10] proposed a multi-objective building optimization approach using NSGA-II and ANN metamodels, focusing on energy efficiency and thermal performance as key objective functions.

Existing studies emphasize reducing building energy consumption, but current data-driven models only predict without offering design optimization solutions. The need arises for methods that explain and prescribe optimized designs for energy efficiency, considering diverse variables and performance goals. Prior research often underutilizes BIM data in optimization and overlooks construction costs for energy efficiency in buildings with lower construction expenses.

This study leverages BIM Data and a Genetic Algorithm (GA) to create a prescriptive and explainable model for decision-making and optimization in indoor building design. Focused on space-based (room-based) indoor prescriptive model actions, the study aims to recommend optimal designs, utilizing BIM interoperability during the design stage. The prescriptive model assists architects and engineers in evaluating cooling, lighting, and heating loads by applying the value engineering method [169], [170]. This method identifies solutions that balance lower costs and higher energy consumption, enabling engineers and designers to analyze the energy performance and cost value of proposed scenarios, facilitating decision-making, and optimizing overall performance.

5.4. BACKGROUND RELATED WORK

5.4.1. Building Energy Consumption Data-Driven Models

The AECO (Architecture, Engineering, Construction, and Operation) industry has recently witnessed a significant data influx, primarily due to technological advancements [120], [171]. Unlike traditional data exchange, transfer, and storage methods, modern digital data formats, notably BIM, offer increased reliability and are better suited for various phases such as design, development, maintenance, and retrofitting [29], [33], [126]. Consequently, once data is collected and processed effectively, it becomes a valuable asset, significantly influencing sound design and development practices. This underscores the importance of harnessing statistical models and machine learning algorithms for data-driven solutions within the AECO sector [24].

Furthermore, in contrast to conventional data storage methods that primarily capture the geometry and attributes of building elements, BIM goes a step further by incorporating topological (relationship) information, which is particularly vital for indoor space-based analyses [9], [24]. This feature unlocks the potential for integrating topology information into learning algorithms rather than directly processing BIM models. Mapping BIM data to graph structure presents a viable approach for applying graph-based learning algorithms like Graph Neural Networks (GNNs) in AECO contexts. GNNs, a subtype of neural networks, are purpose-built for directly processing graph structures. One of the common use cases for GNNs is node classification for both types, transductive and inductive, such as Graph Sample and AgrreGatE (GraphSAGE), wherein each node within the graph is assigned a label, and the objective is to predict the labels of unlabeled nodes, all without access to ground truth information [24], [157]. This technology promises to enhance decision-making and predictive capabilities within the AECO industry [24], [156]. Another important application of Graph Neural Networks (GNNs) is edge prediction for graph classification tasks. GNNs are highly efficient in scenarios where the interactions between neighboring elements have a significant impact, such as optimizing indoor navigation and predicting space-based energy consumption [156], [172]. This efficiency stems from the ability of GNN variants like Graph Attention Networks (GAT) and GraphSAGE to generate new embeddings by considering the node itself and its neighboring nodes. These GNNs leverage the concept of embedding, a technique that transforms properties into low-dimensional, dense, and continuous vectors. In essence, embedding involves creating a vector representation based on the graph's features and attributes while striving to retain as much graph-related information as possible.

Several notable studies have explored the application of GNNs in the AEC (Architecture, Engineering, and Construction) industry. Jin et al. [131] introduced a graph-based unsupervised method to extract functional knowledge from building space structures. They utilized space properties and their boundary relationships for space clustering. Wang et al. [150] developed novel GNN algorithms to enhance the semantic enrichment of BIM models. Their approach improved GNNs for node and edge features, offering potential benefits for typical room classification tasks when adopting graphs and GNN algorithms. Collins et al. [125] innovatively applied Graph Convolutional Network (GCN) algorithms to classify point cloud objects and enrich IFC (Industry

Foundation Classes) models through semantic segmentation. While there have been noteworthy research efforts to leverage graph-based learning algorithms in the AEC industry, the full potential of Graph Neural Networks (GNNs) remains to be tapped. Further exploration and in-depth studies are necessary to understand comprehensively and harness GNNs' capabilities for diverse applications within this field.

5.4.2. Genetic Algorithm (GA)

Genetic Algorithm is a heuristic optimization method inspired by biological processes, particularly natural selection [10], [173]. It is used to search for near-optimal or optimal solutions to complex problems across various fields, including engineering and design in the AEC industry. It operates by evolving a population of potential solutions over successive generations. Each individual in the population represents a potential solution to the problem at hand and is encoded as a string of symbols, often referred to as chromosomes or genomes. The Algorithm iteratively applies selection, crossover, and mutation operators to generate new offspring individuals, mimicking the process of natural selection. Through this iterative process, individuals with better fitness, i.e., those that are more adapted to the problem, have a higher chance of being selected for reproduction, thereby driving the population towards optimal solutions. Over successive generations, the population evolves and improves, converging towards optimal or near-optimal solutions for the given problem [173], [174].

Here is a brief overview of how GAs work as heuristic optimization methods, such as :

Design Optimization: GAs can be used to optimize architectural or structural designs. By defining design parameters as genes in the genetic Algorithm, you can evolve and refine designs over generations to find the most optimal solutions. This can lead to cost-effective and efficient building designs.

Material Selection: Recommending suitable construction materials is crucial in AEC. GAs can consider various factors such as cost, durability, environmental impact, and regulatory compliance to recommend the best materials for a particular project.

Energy Efficiency and Sustainability: GAs can be used to recommend sustainable and energy-efficient design solutions. They can explore various design alternatives and configurations to identify the most environmentally friendly and energy-efficient options.

Cost Estimation: Recommending accurate cost estimates for construction projects is critical. Genetic algorithms can analyze historical cost data, project specifications, and other factors to generate more precise cost estimates, aiding in budget planning and decision-making.

Multi-Objective Optimization: Many AEC decisions involve trade-offs between multiple objectives, such as cost, time, and sustainability. Genetic algorithms excel in multi-objective optimization, helping decision-makers find balanced solutions.

To implement genetic algorithms effectively in the AEC industry, it is essential to define the problem, encode it as a genetic representation, specify fitness functions, and carefully tune algorithm parameters. Additionally, integrating GAs with other data-driven approaches and technologies, such as BIM and machine learning, can enhance AEC's recommendation capabilities and overall decision-making processes.

5.4.3. Value Engineering

Value Engineering (VE), or Value Analysis or Value Management, is a systematic and organized approach used in the Architecture, Engineering, and Construction (AEC) industry to optimize projects in terms of cost and functionality [175]. Value Engineering can be employed in our research to maximize the value of GA recommendations while minimizing costs without sacrificing quality, performance, or safety. It is a structured process focusing on finding cost-effective solutions, improving efficiency, and enhancing overall suggested solution performance. VE is introduced as a fundamental concept of solution value, defined as the ratio of two essential components: function and cost. In this context, "function" pertains to the specific solution and energy consumption based on the suggested solution. The solutions are generated from the GA model, and the building energy consumption for each solution is considered a function value in the VE process. While "cost" encompasses the total expenses associated with the solution. This ratio, function to cost, implies that the value of our recommendation method can be enhanced by enhancing its functionality or reducing its overall cost.

By applying value engineering principles to BECE analysis, project teams can make informed decisions that reduce energy consumption and operating costs and enhance the overall value and sustainability of the building design. This approach aligns economic considerations with environmental and functional objectives, leading to more efficient and sustainable building designs.

5.5. METHODOLOGY

This section delineates the proposed methodology to tackle the challenge of data-driven learning algorithms. The process encompasses three fundamental steps, expounded upon in the subsequent sections. Firstly, an introductory object-oriented framework is presented to construct a space-based graph derived from the IFC model. In the second step, the methodology employs a graph-based classification algorithm to exemplify the implementation of the proposed graph-based learning approach. This approach incorporates various facets, including space geometry, semantics, and neighborhood information. The final step illustrates developing a prescriptive model grounded in the Genetic Algorithm (GA). This model serves the purpose of suggesting optimized BIM designs tailored for BECE applications.

5.5.1. An Object-Oriented Framework to Generate Space-based Graph

The proposed framework employed the graph concept [9], [24], [157] to convert 3D data models (IFC) to a space-based graph to provide a data model that can be linked to different datasets and is compatible with machine learning algorithms. As a case study, we use an IFC file as a building 3D data model to construct a graph. The proposed graph contains space object information as a node, their relationships with neighbor spaces as an edge, and other data sources, such as entity material information with different data models. Our primary intent is to classify the indoor spaces of the building into two critical and non-critical classes from the energy consumption perspective. The classification method is applied to the proposed graph, including each space's geometry, semantic, and relational information. Therefore, the proposed framework generates the graph using the IFC file for future classification. The framework encompasses two main modules: Feature Extraction and Graph Construction, as presented in Figure 5.1, implemented by Python using the IfcOpenShelland NetworkX libraries [128]. Feature Extraction is the first module to extract all geometrical and semantical information of building spaces (space) from the IFC file. This module includes four main functions. The first function extracts each space's geometry (FG) (Volume, Area, and Perimeter) information from the IFC file. The second function (FS) extracts and calculates semantical information (Thermal Resistance Index) for each space. The third function (FNB) finds adjacent spaces for the target space by considering the shared wall. Moreover, the fourth function (Feature Generator (F_{FG})) creates a list of feature values (important parameters) for each space, along with a GUID as a unique ID.

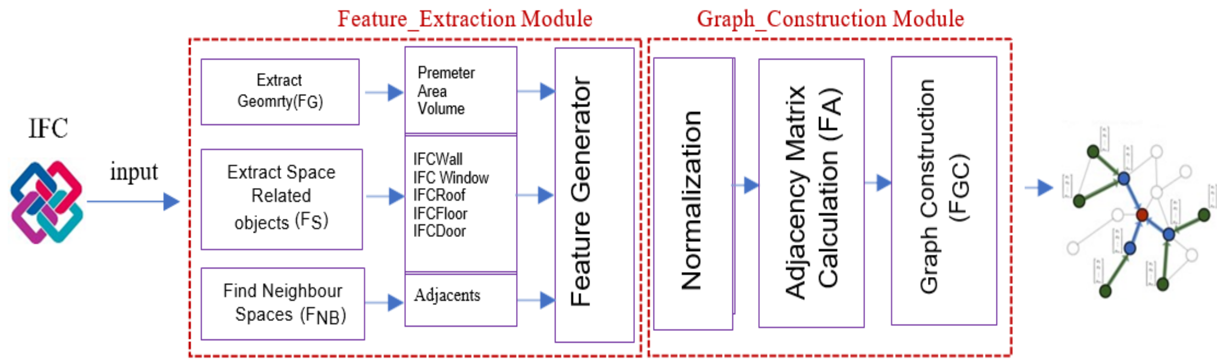


Figure 5.1. The workflow of the proposed framework

5.5.2. Finding Critical Space using GraphSAGE Algorithm

The second step involves identifying the critical spaces within the building, which are spaces or spaces with high energy consumption for heating or cooling. This is achieved through a space classification task focused on energy consumption. The goal is to pinpoint these critical spaces within the building using the property of spaces and the topology (relationship) information between them. This is possible by a graph-based learning algorithm that takes into account space-specific information such as Total Wall Area, External Wall Area, Window Area, External Wall R-Value, Internal Wall R-Value, and Window R-Value, as well as the topology connecting the spaces. To perform the space classification task for building energy analysis, we employ an inductive Graph Neural Network (GNN) method based on GraphSAGE, as proposed by Xu et al. in 2018 [157]. The choice of an inductive method is deliberate because it builds a generic model capable of predicting any new node (space) based on a set of observed training data points.

In contrast, transductive learning, such as Graph Convolutional Network (GCN), constructs a model that fits only the training and testing data it has already encountered, which is unsuitable for our research application. This is especially relevant in retrofitting scenarios where the building's structure may change, new spaces may be added, and others may be removed. Rather than following traditional machine learning classification approaches, we opt for a graph neural network (GNN) to tackle node classification problems, specifically classifying spaces (nodes) within the graph into two classes based on probability values for each space. By establishing explicit connections between spaces, our classification method no longer treats spaces in isolation, unlike traditional building energy estimation algorithms. Instead, it leverages the graph's structure, considering factors like space degrees and neighborhood information. The effectiveness of graph-based properties relies on the assumption that individual spaces are interconnected and influence one another. We proposed an Edge Node GraphSAGE (NE-GraphSAGE) model based on the GraphSAGE method, employed as a supervised classification approach trained using training nodes and edges, typically accounting for 80% of the graph's nodes [24], [176]. Once trained, this model predicts the efficiency class of the remaining nodes (spaces). Ultimately, we assess the model's accuracy by comparing the predicted efficiency class with the actual efficiency class of the testing nodes. Our model of the GraphSAGE classification task involves three key stages: context construction, information aggregation, and the learning process, guided by the loss function described below [24], [176]:

5.5.2.1. Context Construction

The Algorithm has a parameter k that controls the neighborhood depth. If k is 1, only the adjacent space is involved in learning. If K is 2, the spaces at walk depth 2 are considered. Remark that having $k = 2$ means spaces at neighborhood depth two can affect each other through the space in the middle. The value of K is determined experimentally using multiple neighborhoods.

5.5.2.2. Information Aggregation

An information-sharing procedure between neighbors is needed for this step. Therefore, in the first step, we generate a computational graph [177] for each space in the graph to calculate new embedding (feature) values for the target space. Next, aggregation functions or aggregators accept the neighborhood spaces as input and aggregate the neighbor's attributes (features) with weights to create a neighborhood embedding for the target node. To learn embeddings with aggregators, we initialize all space features' embeddings to node features as node attributes. In turn, for each neighborhood depth until k, we create a node embedding with the aggregator function for each node. Different aggregation functions are the LSTM aggregator, Pooling aggregator, and Mean aggregator[157]. The mean aggregator for our calculation is because of its simplicity in the implementation. Eq 5.1 demonstrates the mean aggregation function in which h^{k-1} shows the feature values of the neighbor spaces and $|N(v)|$ is the number of the neighborhood of space v [178].

$$AGG_{u\epsilon N(v)} = \frac{{h_u}^{k-1}}{|N(v)|} \qquad \text{(Eq 5.1)}$$

Each space has a feature vector with a size of 6 * 1 for this test case, and after aggregation, it generates an embedding feature of node 3 with the size of 6 * 1. We normalize the embeddings when each node is processed to have a unit norm. Equation 2 represents updating node embedding calculation using the neighborhood's and target node's features. In Eq 5.2, h_v^k denotes, as an embedding feature node v in walk depth k, and σ represents the activation function.

$$h_v^k \leftarrow \sigma(w.MEAN(\{h_v^{k-1}\} \cup \{h_v^{k-1}, \forall u \in N(v)\}) \qquad \text{(Eq 5. 2)}$$

We incorporate the activation function to introduce nonlinearity into our model. The Sigmoid activation function [131] is employed as a nonlinearity function due to its output range being between 0 and 1, making it ideal for computing the probability of output classes.

5.5.2.3. Training the GraphSAGE by the loss function

To train the neural network weights in GraphSAGE, we need a differentiable loss function to calculate the distance between the actual value of the node class and the predicted values. Therefore, we have applied the Squared Error Loss (SE) function [165] for each space classification (130 nodes in our dataset). Then, we split the nodes into a training set (100 nodes) and a testing set (30 nodes). The predicted process takes input features from each computation graph and calculates the probability for each space. The SE function measures the distance of actual and predicted output value for 100 nodes, which is called loss value. The mean of loss values (Mean Squared Error – MSE (loss function)) for 100 nodes is calculated for each iteration (epoch). At the end of each epoch, the backpropagation process adjusts the neural network's metrics weight. The learning process is continued iteratively to catch the best accuracy on training nodes [176].

5.5.3. Designing a Prescriptive Model based on a Genetic Algorithm

In the third step, we have developed a prescriptive model to provide recommendations for enhancing critical spaces' energy efficiency. Additionally, we have designed a node and edge-based graph neural network known as NE-GraphSAGE, which serves as the objective function for calculating the energy consumption of spaces within the building. This model is precisely engineered to incorporate node and edge features throughout the learning process, enabling accurate predictions of building energy consumption based on input parameters. Furthermore, a genetic algorithm model is developed that suggests optimal solutions concerning geometry, electrical systems, and materials to reduce energy consumption in critical spaces, which is detected in section 5.4.2. The recommendations are invaluable for engineers, aiding them in optimizing their decision-making processes and achieving energy-efficient building designs. The below describes the proposed prescriptive model.

5.5.3.1. Designing a Prescriptive Framework

This paper has introduced a prescriptive framework to suggest alternative solutions in terms of geometry, materials, mechanical systems, and electrical systems for spaces within the building identified as critical spaces. This framework comprises three primary phases: (1) Space Functionality Evaluation Phase (SFEP), (2) Cost Evaluation Phase (CEP), and (3) Prescription Evaluation Phase (PEP). The schematic representation of this framework is provided in Figure 5.2 for a comprehensive overview.

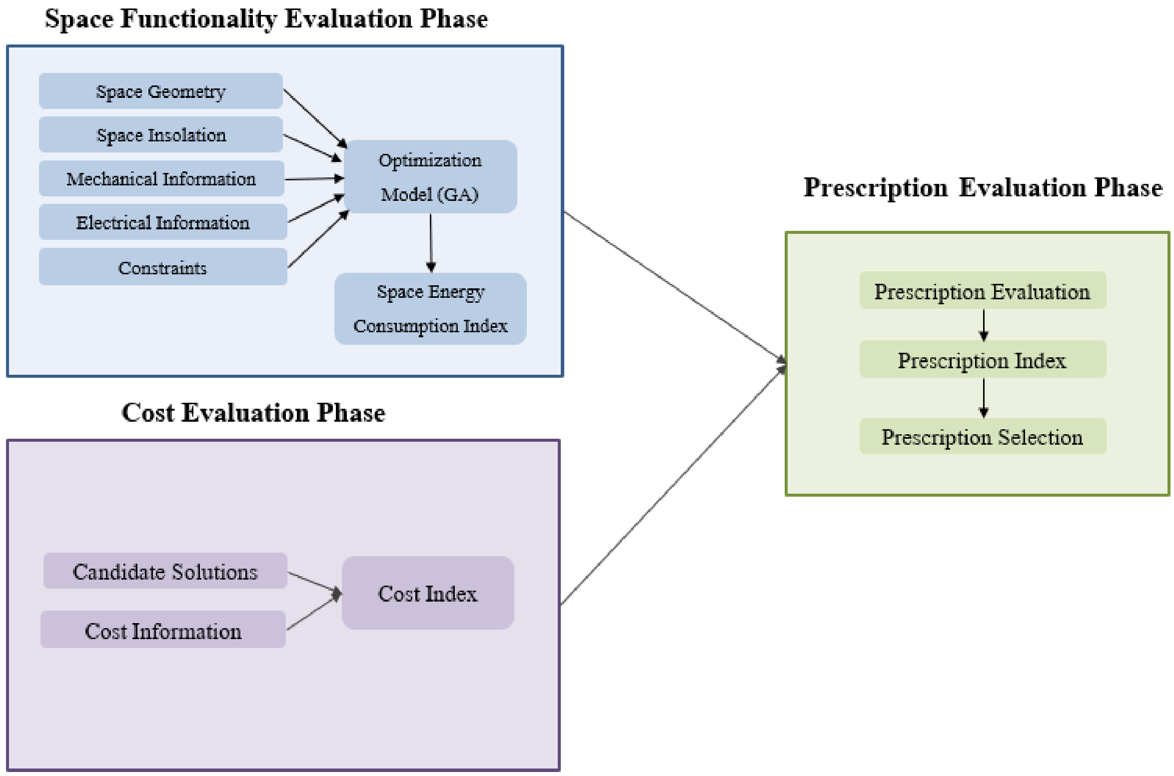
- ***Space Functionality Evaluation Phase (SFEP)***

Within the Space Functionality Evaluation Phase (SFEP), various parameters related to the building's geometry and insulation, including details about walls, windows, roofs, and floors, are utilized. These parameters encompass information such as window dimensions, external wall surface area, materials used for space walls and windows, and compliance with building code standards. These inputs are fed into a genetic algorithm chosen for its heuristic optimization capabilities. Heuristic models are particularly suitable for problems with numerous potential solutions, such as optimizing energy consumption in a building. Constraints in this phase are defined by establishing upper and lower limits for space properties. Moving to the Cost Evaluation Phase (CEP), candidate solutions are assessed based on North American building standard codes to determine their associated costs [180].

Also, a genetic algorithm (GA) is adopted as an optimization method to heuristically search for the near-optimum solution evaluated by the predefined fitness function. GA's fitness function is the energy consumption of each space calculated by the graph-based neural network (GraphSAGE) (). The output value of the neural network (energy consumption) is normalized by dividing it by the volume of each space. GA identifies and imports a set of feasible candidate solutions (i.e., completely satisfies constraints) to the CEP phase. Figure 5.4 represents the process GA algorithm in which the input parameters are determined based on the SHAP value.

The variables in Appendix E are selected to be optimized as they significantly impact the energy consumption of critical spaces. Ultimately, we define the Space Functionality Index (SFI) as the one-minute normalized value of each space's energy consumption by the building's total energy consumption (Eq 5.3). This means that spaces with less energy consumption have more functionality. One of the advantages of the proposed process is that the input parameters for each space can be variable. Therefore, we introduce block concepts for each space, including parameters involved in the optimization process and the critical spaces forming a chromosome in the GA algorithm.

Figure 5.2. Overview of Prescriptive Framework



- ***Cost Evaluation Index (CEI)***

In the Cost Evaluation Phase (CEP), the focus shifts to estimating quantities based on the actual quantities of building components such as walls, roofs, floors, and window areas. This estimation accounts for the material used across all three floors in the case study dataset. The building is meticulously dissected into measurable components, and the costs associated with each component are summed to arrive at an overall estimate. This process assumes a standardized cost for each unit of work and material purchased while excluding labor costs, which are not part of the optimization parameters.

$$SFI = 1 - normalized\ energy\ consumption\ value \qquad \text{(Eq 5.3)}$$

The genetic algorithm designed for this research, which was completed with a population size of 50 and a total of 120 generations, selected a crossover rate of 0.7 and a mutation rate of 0.15, which proves advantageous. With a population of 50 individuals, a crossover rate of 0.7 ensures that during reproduction, a substantial portion of the population undergoes genetic exchange, fostering exploration across diverse solution spaces. This encourages the algorithm to efficiently explore a wide range of potential solutions throughout 120 generations. Meanwhile, a mutation rate of 0.15 introduces necessary diversity by randomly altering a small fraction of the genetic material in each offspring, preventing premature convergence and facilitating effective exploitation of promising regions within the solution space. By carefully balancing exploration and exploitation through these parameter choices, the genetic algorithm maximizes its ability to converge toward high-quality solutions within the given computational constraints.

Table 5.1.Upper and lower limits of space properties as constrain values

Space Items	Item Variable Values [179]
WIN_Area (m2)	48 * 36 48*72 56*36 56*72 64*36 64*72
WIN_U_Value (m²·K)/W)	Single glazing: 4.8 to 5.8 Double glazing: 1.2 to 3.7 Triple glazing below 1 W/(m²K).
Wall_R (m²·K)/W)	R-15 R-19 R-30
Air_Freshing (cfm)	200-500
Roof_R (m²·K)/W)	R-15 R-50
Floor_R (m²·K)/W)	R-55 R-60
Plug Load(W/m2)	150-600
Ventilation Types	Supply and extract-only systems.
	Balanced ventilation systems.
	Heat Recovery Ventilation (HRV) - Demand-based
Lighting	luorescent/compact fluorescent (CFL)
	LED
	Occupancy or daylight-dimming sensors

Typically, this quantification is carried out through a takeoff process. BIMVision software was utilized for this purpose. The inputs for this phase comprise the candidate (elite) solutions and their respective cost information. To determine the cost index (CI) for each solution and the total project, the cost for that solution is divided by the maximum project cost among all candidate solutions.

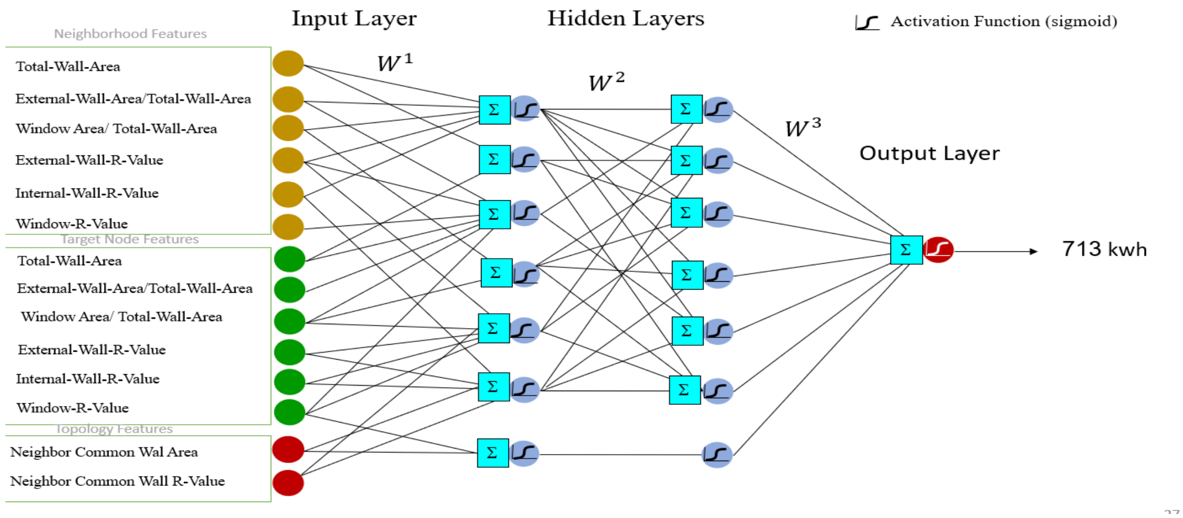


Figure 5. 3.Graph-SAGE neural network (Fitness Function) [24]

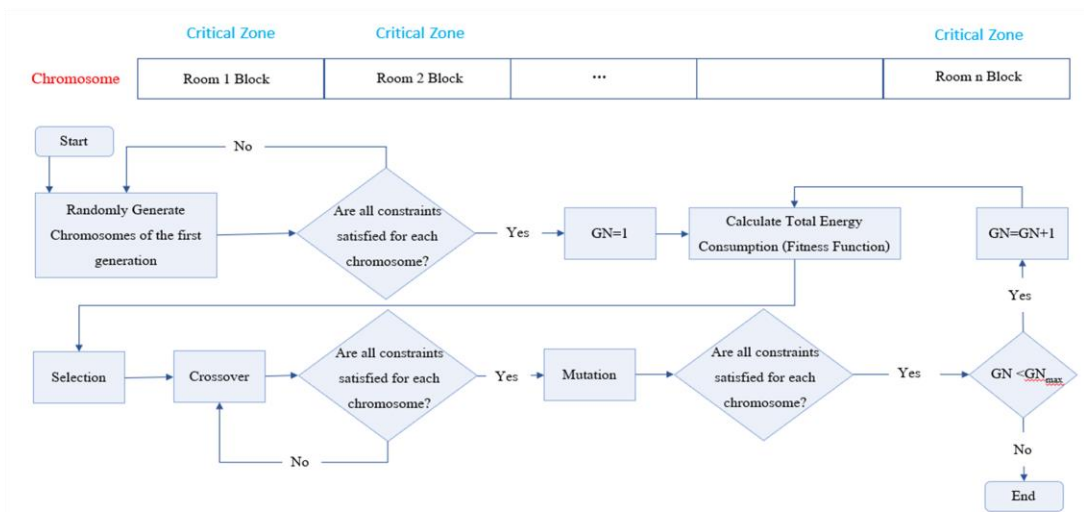


Figure 5.4.Designed GA Optimization Process.

- ***Prescription Evaluation Phase (PEP)***

The genetic algorithm designed for this research, which was completed with a population size of 50 and a total of 120 generations, selected a crossover rate of 0.7 and a mutation rate of 0.15, which proves advantageous. With a population of 50 individuals, a crossover rate of 0.7 ensures that during reproduction, a substantial portion of the population undergoes genetic exchange, fostering exploration across diverse solution spaces. This encourages the algorithm to efficiently explore a wide range of potential solutions throughout 120 generations. Meanwhile, a mutation rate of 0.15 introduces necessary diversity by randomly altering a small fraction of the genetic material in each offspring, preventing premature convergence and facilitating effective exploitation of promising regions within the solution space. By carefully balancing exploration and exploitation through these parameter choices, the genetic algorithm maximizes its ability to converge toward high-quality solutions within the given computational constraints.

$$PVI = \frac{SFI}{CI} \qquad \text{(Eq 5.4)}$$

As a result, the solutions with the highest PVI are the most desirable solutions because they have higher functionality (efficient energy consumption) and lower costs.

5.6. EXPERIMENTAL RESULT AND DISCUSSION

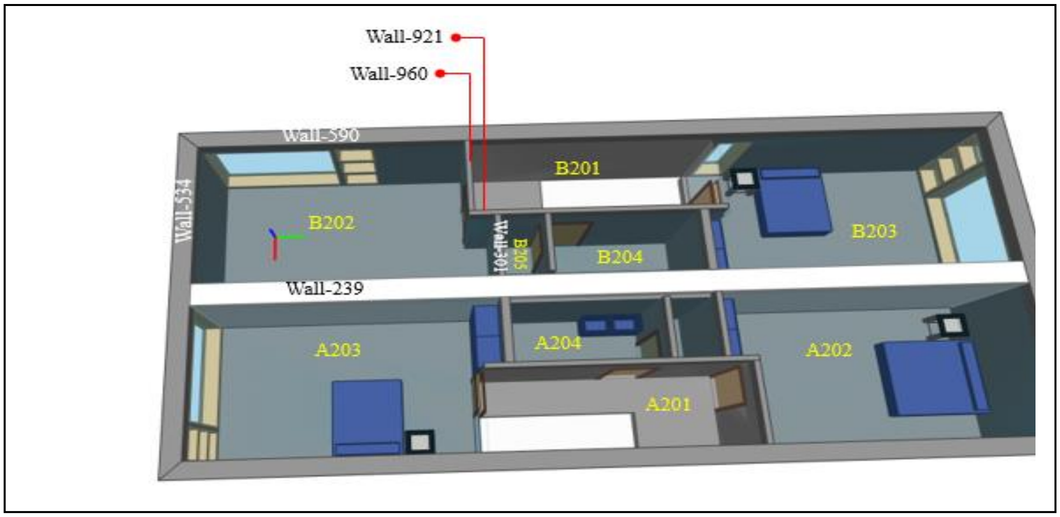
The IFC data representing the BIM as a case study model was obtained in IFC format from the Open IFC Model Repository. It comprises 130 spaces with various functions, such as offices, conference spaces, lounges, kitchens, lobbies, and washrooms. This model includes convex and non-convex spaces, with several walls having non-perpendicular intersections. It features over 40 continuous exterior walls and 25 shared windows, making it a more intricate model due to its multi-level structure, complex geometry, and space relationships. This complexity makes it an ideal candidate for evaluating the proposed methodology in accurately extracting geometry and semantics. The proposed methodology is applied to the mentioned dataset to generate a graph model discussed in section 4.5.1 [9]. Then, the proposed classification and the interpretation models to explain the classification results, which are debated in Section 4.5.2, are applied to the dataset. The detailed implications are discussed in the following sections.

5.6.1. Generate a space-based graph from the IFC dataset

The methodology involves four steps: generating a framework data model, extracting BECE-based geometry and semantic information, obtaining adjacent space information, and constructing the BECE-based weighted space-based graph. The proposed framework found that the IFC model includes convex and non-convex spaces, and all the walls, roofs, and floors are perpendicular. Also, eight continuous exterior and two continuous interior walls are available in this model, which makes it a complex model to generate a space-based graph. The multi-space shared wall issue in this model opens this opportunity to test our methodology from the perspective of accurate geometry extraction and recognizing the correct space adjacent. The proposed BECE framework generates an accurate graph from the IFC model for the test case. Figure 5.5 represents the steps of this process for the IFC model [9].

5.6.2. Apply genetic algorithm for optimized recommendations

In this phase, the Genetic Algorithm (GA) generated a range of solutions, resulting in a Functional Index (FI) that varied between 0.42 and 0.88, with an average value of 0.69. Among the 100 solutions produced, only ten demonstrated a high level of functionality, exceeding 85% (SFI > 0.85), and were therefore selected as elite solutions. The top five of these elite solutions are presented in Table 3, which lists their respective SFI, CI, and PVI values.



(a) Geometry Features Extracted from IFC Model

Items	Results
Wall-239 - Int	49.20 m²
Wall-534 - Ext	22.95 m²
Wall-590 - Ext	49.20 m²
Wall-960 - Int	5.69 m²
Wall-921 - Int	15.73 m²
Wall-301 - Int	4.64 m²
Roof	135.15 m²
Floor	23.17 m²

(b) Adjacents Spaces Extracted from IFC Model

Target Space	Adjectes Spaces
Space B202	A203
	B201
	B205
	B203
	A202

(c) Geometry Features Extracted from Framework

Items	Results
Wall-239 - Int	18.12 m²
Wall-534 - Ext	10.75 m²
Wall-590 - Ext	16.56 m²
Wall-960 - Int	5.69 m²
Wall-921 - Int	1.57 m²
Wall-301 - Int	4.64 m²
Roof-102	17.50 m²
Roof-103	4.58 m²
Floor	22.04 m²

(d) Adjacents Spaces Extracted from

Target Space	Adjectes Spaces
Space B202	A203
	B201
	B205
	A102
	B103

Figure 5.5.The extracted information from the IFC model and framework

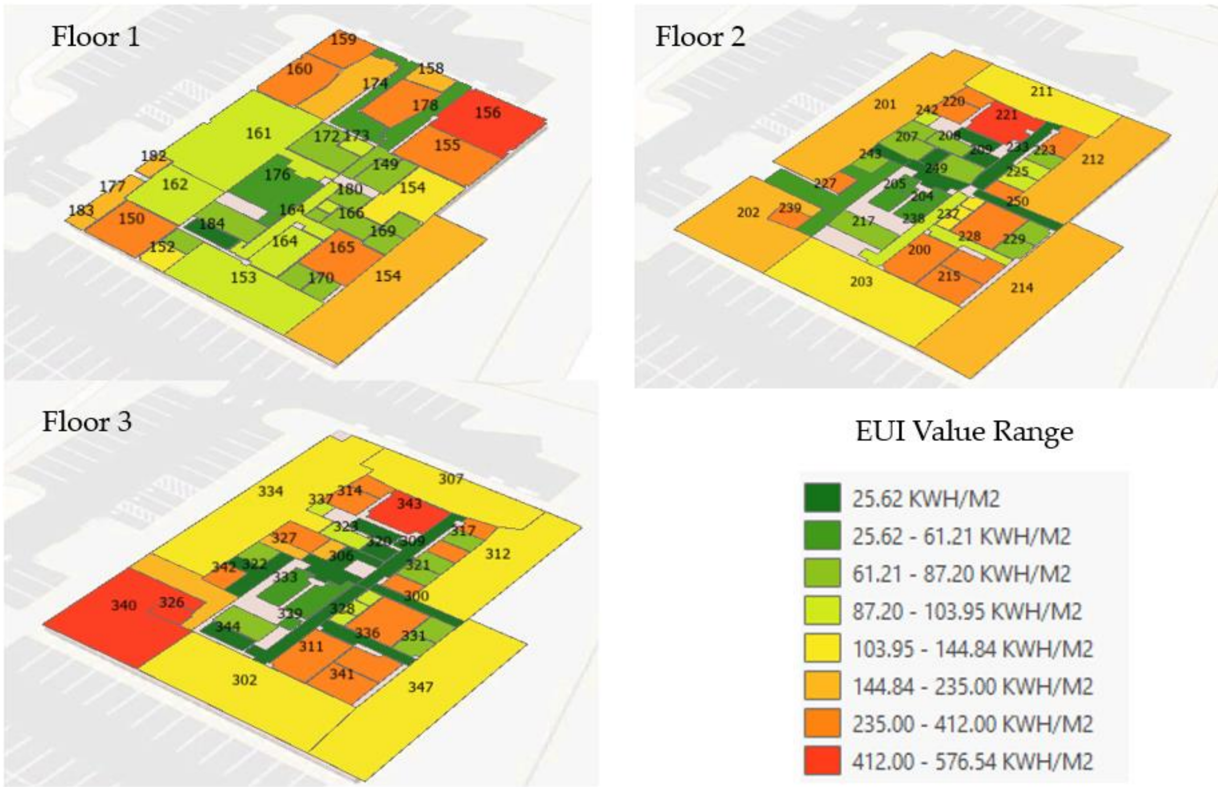


Figure 5.6. The classification result symbolized based on estimated EUI values

Table 5. 2. Different solutions recommended by the GA Algorithm

	Space -156				**Space 221**	**Space -340**	**Space - 326**
Solutions	Ventilation	Lighting	External-Wall – R-Value	Windows Area	Ventilation	WIN-U-Value	Neighbour-Wall-R-Value
1	Type3	Type3	R-30	48*36	Type2	Double glazing	R-15
2	Type3	Type3	R-30	48*72	Type3	Triple glazing	R-19
3	Type2	Type3	R-15	48*72	Type2	Single glazing	R-15
4	Type1	Type2	R-15	64*36	Type1	Single glazing	R-15
5	Type1	Type1	R-15	64*72	Type1	Single glazing	R-15

Table 4 illustrates that the solutions with the highest functionality (e.g., Solution 1) do not necessarily have the lowest cost. In this case, the optimal solution is Solution 1, which boasts the highest SFI but comes with a marginal increase in cost, approximately 1.05%, compared to the lowest-cost elite solution. Notably, some solutions exhibit nearly identical SFI values, indicating that certain variables, like soft lighting and internal wall materials, have a minimal impact on costs.

On the other hand, Solution 2 carries the highest cost, suggesting that factors such as ventilation type and high U-value windows can significantly influence project expenses. Comparing the cost values across all solutions, CI ranges from 0.92 to 0.99, indicating that project costs can vary by approximately 7% due to alterations in window and wall materials, window dimensions, and ventilation system configurations. However, some solutions yield identical CI values because changes in lighting and usage-based systems for specific spaces do not substantially affect project costs but enhance space functionality.

Table 5.3. SFI, CI, and PVI values of recommended solutions

Recommended Solutions	SFI	CI	PVI
Solution 1	0.8805	0.9482	0.9325
Solution2	0.8716	0.9632	0.9049
Solution3	0.8666	0.9012	0.9178
Solution4	0.8572	0.9542	0.8983
Solution5	0.8570	0.9543	0.8980

5.6.3. Graph-base classification model to find critical spaces

The classification model (NE-GraphSAGE) is analyzed for energy consumption using the Energy Use Intensity (EUI) index. Ground-truth values were calculated for three stories, factoring in various parameters such as space geometry, materials, and HVAC details. Outliers were detected using Z-scores, highlighting critical spaces. EN-GraphSAGE was applied to classify these spaces as critical or non-critical, achieving 91.26% accuracy at epoch 3500. Four classification methods, including NE-GraphSAGE, GraphSAGE, ANN, and SVM, were evaluated based on accuracy and F1 score. F1 score, considering both precision and recall, ensured a balanced assessment, providing valuable insights into the classifiers' effectiveness (Table 5.4) [24].

Table 5.4.The comparison analysis of classification models

Model	Accuracy	F1
GraphSAGE	79.5	0.73
NE-GraphSAGE (2 layers)	86.6	0.82
NE-GraphSAGE (3 layers)	91.2	0.88
NE-GraphSAGE (4 layers)	86.9	0.83
ANN	72.6	0.67
SVM	71.8	0.66

Our study demonstrates that NE-GraphSAGE (3 layers) outperforms other methods, showing a 12% average increase in accuracy compared to non-graph-based methods, achieving 91.2% accuracy and 0.88 F1 score. NE-GraphSAGE (2 layers) and NE-GraphSAGE (4 layers) exhibit slightly lower accuracy and less balanced predictions than NE-GraphSAGE (3 layers). GraphSAGE with three layers was also trained, but NE-GraphSAGE advantage lies in its ability to consider both node and edge features during embedding updates. This feature led to a 12% accuracy improvement and a 0.15 F1 score increase compared to GraphSAGE. The impact of edge features (topology) was evident, as non-graph-based methods misclassified critical spaces neighboring low R-value shared walls. In contrast, NE-GraphSAGE (3 layers) correctly identified these spaces as critical. Hence, NE-GraphSAGE (3 layers) was chosen as our space classification algorithm for superior performance. Figure 5 represents the classification result symbolized based on estimated EUI values.

5.6.4. Result Validation

The results from the final epoch of the Genetic Algorithm, comprising optimized solutions (50 individuals), are validated from various perspectives and summarized in Table 5.

Table 5.5. Different Types of Validation of the *Result*

Validation Type	Propose of Validation	Validation Process	Validation Result
Comparison to RETScreen simulation software in The results of the model being validated are compared to results in software cases with known results.	Validation of GA proposed solutions which produce near optimum solutions.	The GA algorithm developed in this model is tested by comparing its results to the known results of RETScreen simulation software.	The GA results are identical or very close to the known results of RETScreen energy simulation software. For instance, a space with a large size of external windows and a low rate of external walls is classified as a critical space. GA suggested increasing the R-value or changing the material by decreasing the window size, which is expected considering the defined constraints and building codes.

Constraint Value Validation: This process assesses the proposed geometry and material solutions to ensure they align with reasonable standards, filtering out implausible options. Also, parameter Analysis in which the model undergoes scrutiny to confirm that alterations in input values correspond proportionally to real-world system changes, ensuring accuracy and consistency.	Validation of GA proposed solutions that meet the building codes and the outputs are sensitive to input features with expected results.	The model was tested for extreme input feature values, such as having a very high Windows R-value and a high rate of thermal residency for walls. The models show low energy consumption by drastically increasing the cost of materials.	The outputs are tested for the tested extreme conditions. The impacts of the tested changes on project cost and model functionality
Compare the solution results with US Energy Plus Technical Reference.		Compare the solutions with technical Reference energy consumption values.	The energy consumption from GA solutions has close values to the US Energy Technical Reference (Figure 5.7)

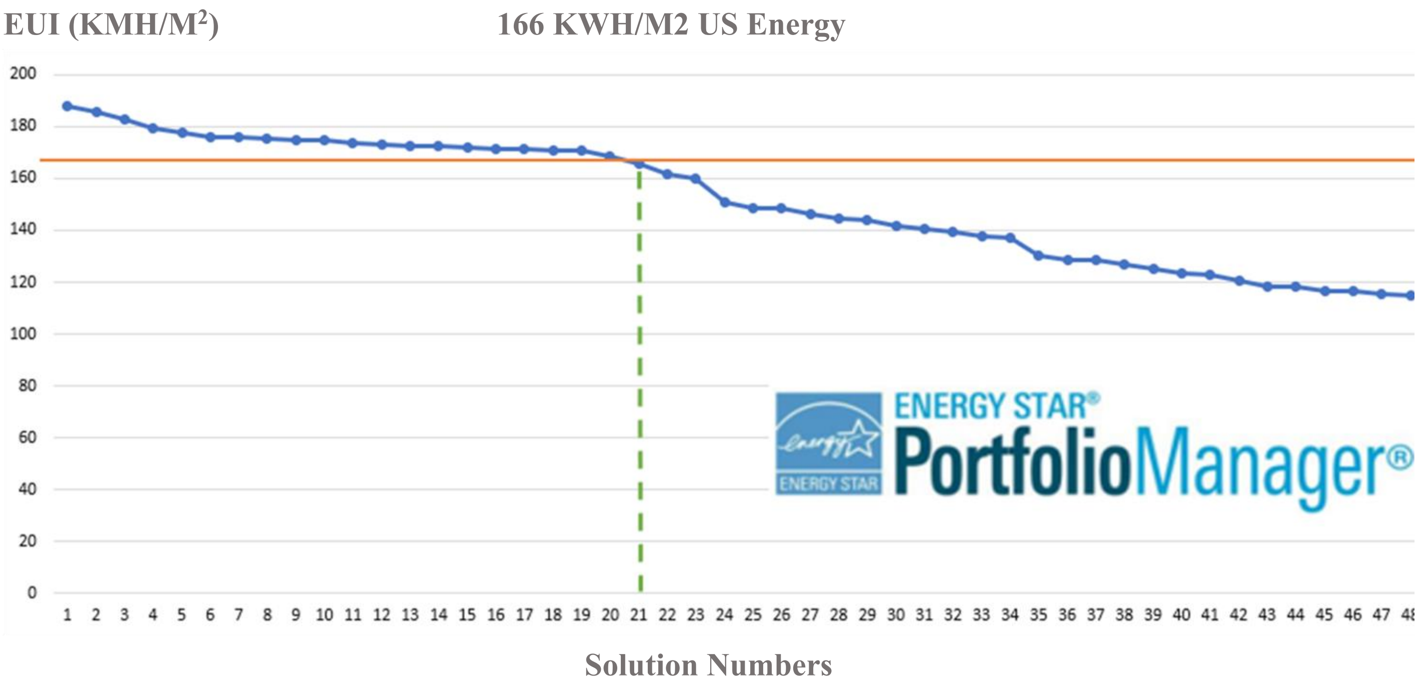


Figure 5.7. GA Solution EUI values in comparison with US Energy Technical Reference

Office energy consumption, as defined by the Energy Star technical reference with a benchmark of 116.4 EUI (Energy Use Intensity) in kBtu/ft², plays a pivotal role in shaping sustainable work environments. Adhering to the specific EUI standard set by Energy Star ensures that offices are designed and operated with energy efficiency in mind. This stringent benchmarking approach encourages businesses to adopt innovative strategies to reduce their energy footprint. Implementing high-efficiency lighting solutions, such as LED technologies, and optimizing heating, ventilation, and air conditioning (HVAC) systems are paramount. Moreover, Energy Star's emphasis on selecting energy-efficient office equipment underscores the importance of integrating advanced technologies that consume minimal energy while delivering optimal performance. By aligning office spaces with the 116.4 EUI benchmark, companies contribute significantly to environmental conservation and benefit from reduced operational costs, making it a strategic investment for a sustainable future (Figure 5.7).

As a result, the proposed prescriptive model in this research aids building engineers toward energy-efficient design by offering standardized guidelines and benchmarks in the design phase or retrofitting phase. The recommended solutions, which outline specific geometry, material, mechanical, or electrical, are aimed at minimizing energy consumption and maximizing efficiency. By adhering to prescriptive modes, engineers gain clarity and direction, ensuring that their designs incorporate proven strategies and technologies known to enhance energy performance and meet the building codes. They provide different scenarios for engineers to have optimum decision-making by choosing the specific solution to consider, energy consumption, and cost of the project. Ultimately, by providing clear and actionable guidance, prescriptive modes empower building engineers to make informed decisions that result in more sustainable and environmentally friendly buildings, thus contributing to the global effort towards mitigating climate change and promoting a greener future.

5.7. CONCLUSION

This study has devised a framework to offer solutions to improve the energy consumption of critical areas within a building. It achieves this by identifying cost-effective building materials, optimizing geometry, and enhancing mechanical and electrical systems using Building Information Modeling (BIM). The study employs Genetic Algorithms (GA) as an optimization model to generate viable candidate solutions, conducting heuristic searches to approach optimal variables. Furthermore, it integrates value engineering principles to model energy consumption functionality and assess the cost-efficiency of these candidate solutions. Within the Genetic Algorithm (GA) framework, this study incorporates building construction codes as constraints, encompassing factors like R-value, window size, air freshening, plug load, ventilation type, and lighting into the optimization process.

Additionally, value engineering principles are employed to accurately gauge the influence of building materials and geometry on project costs. This approach marks a significant improvement over existing methods, as it thoroughly tackles the shortcomings commonly associated with them. Through a successful implementation in a real-world case study, the primary contributions of this paper can be summarized as follows:

(1) Various comparisons and analyses can be undertaken using the presented data, demonstrating the proposed method's capabilities to help engineer efficient and data-driven decision-making. In addition, (2) it accounts for more variables such as construction, mechanical, and electrical variables that can influence the cost efficiency of the building design, captures their complex dependency, and determines the significance of their impacts on the project costs ; (3) it considered geometry and material variability in optimization and evaluated the impact of them on the project functionality and cost through SFI and CI, respectively; (4) it evaluated and selected the optimum solution based on both functionality and cost, which enabled the designer to evaluate the satisfaction of the subjective constraints and quantify the cost impacts of the solution; and (6) it provides different construction design scenarios, enabling the engineer to identify the most efficient plan based on the primary goal of the project. It is advisable to incorporate the project's long-term cost into the optimization process in future endeavors. This would involve estimating the extended cost implications of the elite solution and providing insights into its impact on energy efficiency and cost savings over time. Additionally, exploring alternative heuristic optimization methods alongside Genetic Algorithms (GAs) could be pursued to evaluate their suitability and effectiveness compared to GA.

CHAPTER 6: CONCLUSION

This section, divided into four parts, outlines the key findings from each chapter, highlights the unique contributions of this dissertation, and provides suggestions for future research endeavors.

6.1. GENERAL CONCLUSION

The overall objective of this dissertation is to develop a novel framework for indoor building energy consumption estimation by employing BIM as a 3D data source and a data-driven model. This objective is reached by dividing it into three sub-objectives:

i) Develop an intermediate data model to generate a space-based graph based on the IFC model to design a graph-based machine learning algorithm based on Building Energy Consumption

Estimation space-based graphs to consider detailed information about building objects and their adjacent information in the learning process.

ii) Design an optimized graph-based learning algorithm, such as Graph Neural Network (GNN), based on a weighted knowledge graph.

iii) Develop an explainable and prescriptive model to interpret the result of the graph-based prediction model so we can have a quantified-based interpretation for each critical space.

The major conclusions of each objective are provided in the following sections.

6.1.1. Develop an intermediate data model to generate a space-based graph based on the IFC model to tackle the compatibility issues between the IFC model and machine learning methods.

Chapter 3 introduces a BECE-based framework to solve the complexity of using IFC in the data-driven model, such as machine learning algorithms, for building energy consumption estimation (BECE). This chapter proposes a framework to generate a BECE-based space-based graph from the IFC model. While the existing approaches rely on the semantic and geometry information in the IFC model, which causes several erroneous analysis results, the proposed methodology defined classes and objects in the new data model comprise the explicit semantic, geometry, and space adjacent information that is not available in the native IFC models. In addition, some functionalities are also designed and implemented to apply geometrical computation to extract accurate semantics, geometry, and relationships for IfcSpaces in the IFC model. The low accuracy of adjacent information from the IFC model illustrates the importance of recognizing and extracting the valid adjacent information in the IFC model. The proposed methodology for extracting adjacent information shows high accuracy for vertical and horizontal adjacent information. Also, the result demonstrates the high accuracy of semantic and geometry information for the spaces in the model.

Furthermore, an accurate space relationship algorithm is proposed to extract space-adjacent information with a weighted value. Moreover, the proposed methodology is not limited to extracting the information for spaces with convex polyhedron geometry; it is designed to apply geometry computation for convex and non-convex polyhedron spaces. To this end, an accurate BECE-based graph is generated. Thus, the proposed graph encompasses accurate semantic, geometry, and space-adjacent information. Also, the proposed space-based graph does not have IFC model complexity and is compatible with data-driven algorithms.

6.1.2. Design a graph-based machine learning algorithm based on Building Energy Consumption Estimation space-based graphs to consider detailed information about building objects and their adjacent information in the learning process.

Chapter 4 introduces a novel GNN-based classification model called NE-GraphSAGE, designed to identify critical spaces (specifically spaces with high energy consumption) using a 3D building model. This model considers both the space's characteristics and the surrounding information during the training of the classification model. The outcomes demonstrate a superior model accuracy compared to both non-graph-based and conventional GraphSAGE models. To tackle the inherent trade-off between model accuracy and interpretability, We integrate the NE-GraphSAGE model with the SHAP (SHapley Additive exPlanations) model. SHAP is a technique employed for model interpretation, offering explanations for individual predictions.

By combining SHAP with the NE-GraphSAGE model, both high accuracy and interpretability have been achieved. This approach enables a clearer comprehension of the factors influencing the model's predictions, thereby enhancing the transparency and reliability of the results. Rather than forcing a choice between accuracy and interpretability, the proposed methodology presents a solution that allows for pushing the boundaries of model complexity and accuracy while still providing intuitive explanations for each space.

Furthermore, the suggested interpretation model can reveal the impact of individual features within a space and aggregated features across 3D adjacent spaces. The results indicate that the integration of the high-accuracy classification model (NE-GraphSAGE) with the SHAP model facilitates accurate insights for designers and engineers in their design and decision-making processes. However, it is essential to acknowledge two main limitations of this methodology, which are discussed in the limitation and future work sections.

6.1.3. Develop an explainable and prescriptive model to interpret the result of the graph-based prediction model so we can have a quantified-based interpretation for each critical space.

Chapter 5 devised a prescriptive model to offer solutions to improve the energy consumption of critical areas within a building. It achieves this by identifying cost-effective building materials, optimizing geometry, and enhancing mechanical and electrical systems using BIM. The study employs Genetic Algorithms (GA) as an optimization model to generate viable candidate solutions, conducting heuristic searches to approach optimal variables. Furthermore, it integrates value engineering principles to model energy consumption functionality and assess the cost-efficiency of these candidate solutions. Within the Genetic Algorithm (GA) framework, this study incorporates building construction codes as constraints, encompassing factors like R-value, window size, air freshening, plug load, ventilation type, and lighting into the optimization process.

Additionally, value engineering principles are employed to accurately gauge the influence of building materials and geometry on project costs. This approach marks a significant improvement over existing methods, as it thoroughly tackles the shortcomings commonly associated with them. Through a successful implementation in a real-world case study, the primary contributions of this paper can be summarized as follows:

(1) Various comparisons and analyses can be undertaken using the presented data, demonstrating the proposed method's capabilities to help engineer efficient and data-driven decision-making.

(2) it accounts for more variables such as construction, mechanical, and electrical variables that can influence the cost efficiency of the building design, captures their complex dependency, and determines the significance of their impacts on the project costs.

(3) Geometry and material variability in optimization were considered, and their impact on the project functionality and cost was evaluated through SFI and CI, respectively.

(4) it evaluated and selected the optimum solution based on both functionality and cost, which enabled the designer to evaluate the satisfaction of the subjective constraints and quantify the cost impacts of the solution; and (6) it provides different construction design scenarios, enabling the engineer to identify the most efficient plan based on the primary goal of the project.

6.2. LIMITATIONS

Despite the substantial contributions made by the methods proposed and algorithms developed within this study, certain limitations have been discerned in the research assumptions:

- The primary objective of the research is to analyze energy consumption at a granular level, specifically focusing on individual spaces (spaces). Consequently, the accuracy and reliability of the results are contingent upon the precision of space definitions within the inputs of the case study.

- In generating space relationships (topology-edge), the assumption is that only a single entity (such as a wall, window, roof, or floor) exists between any adjacent entities.
- The research assumes a sole material entity exists between two adjacent entities, and the IFC (Industry Foundation Classes) data does not encompass multi-material types. This assumption serves as the foundation for computing the R-value of IFC objects.

▪ Data-Driven Model Components

- Energy consumption is influenced by diverse factors, including environmental conditions, building electrical information, occupant behavior, heating, ventilation, air conditioning (HVAC) systems, and appliance usage. Occupants significantly influence energy consumption as they have more autonomy in controlling building systems like HVAC and appliances. However, due to the complexity, variability, uncertainty, and lack of data associated with occupant behavior, this research neglects the occupant behavior factor in BECE analysis.
- Limited historical data was an important limitation in this research. In our case study, a limited amount of space-based historical data is available for our modeling calibration and validation. Although we used RETScreen software to generate more historical data based on the case study environment, ensuring the model's calibration and reliability is challenging.
- The accuracy of the trained model is highly dependent on the data quality. If the input data (graph) used for the proposed modeling is inaccurate or contains errors, the model's classification will be unreliable.

- Prescriptive Model Components:
  - In this study, we utilize a genetic algorithm to produce potential outcomes for critical spaces identified through the graph classification model (outlined in Chapter 4). The solutions are derived from critical spaces at levels 1 and 2, with the option to incorporate other space classes into the genetic algorithm model. More than two levels of the neighborhood can be investigated to achieve better accuracy. However, it significantly increases computation run time.
  - The cost estimation for the case study relies on prevailing market rates and established structural and material costs. The study assumes a constant labor rate. Additionally, this research does not consider the long-term cost implications of the elite solution or offer insights into its influence on energy efficiency and potential cost savings over an extended period.

6.3. NOVEL CONTRIBUTIONS

This research investigated the solutions for a graph-based BECE using BIM data. The innovation and novel contributions of this research include:

1. A new framework is developed to solve the compatibility challenge of 3D models (BIM) with BECE-based data-driven models by introducing a space-based topological graph for assessing building energy consumption through BIM. The research novelty proposes an object-oriented data model that generates a space-based graph for building that is compatible with graph-based classification and prediction models encompassing 3D topology and relevant attribute information.
2. A graph classification model is established to improve the accuracy of space classification for BECE analysis by introducing NE-GraphSAGE and making the model interpretable. The novelty of this research lies in proposing a graph-based classification model that involves space topology

information in the training process that ranks spaces within a building in terms of energy consumption, utilizing a graph-based data structure derived from BIM. This model incorporates essential building parameters such as 3D geometry, material, and 3D topology information to identify critical spaces and suggests optimized, energy-efficient BIMs. Also, the result of the classification model is interpreted by the SHAP model to make it more transparent for engineers to analyze the impacted parameters in the classification result.

3. A prescriptive model is designed and implemented using a genetic algorithm to help engineer optimized decision-making. The prescriptive model proposes optimal solutions based on building codes, employing a value-engineering method to assess the cost-effectiveness of these solutions. The proposed model in this contribution balances energy efficiency and the project's cost by introducing a value index. It helps engineers make data-driven decisions from the model's output by exploring various solutions and scenarios based on diverse datasets and considering the cost of elite solutions.

6.4. FUTURE RESEARCH DIRECTIONS

Employing BIM in the analysis of BECE reveals numerous unexplored dimensions and facets. This study highlights several uncharted territories, proposing them as potential subjects for future research.

6.4.1. Regenerating non-convex polyhedron spaces in IFC model

The IFC model is capable of accommodating both convex and non-convex polyhedron spaces. Nevertheless, the prevailing tendency in most IFC models is to treat the geometry of spaces as convex polyhedral, overlooking the intricacies associated with non-convex polyhedron geometry. This oversight results in an inaccurate estimation of building energy consumption. In this study, we introduced a geo-computation algorithm designed to extract precise geometry information from non-convex polyhedron spaces promptly. However, it is imperative to note that future research endeavors are required to regenerate non-convex polyhedrons and redefine the original IFC model for a more comprehensive and accurate representation.

6.4.2. Finding the number of objects and related materials between two IFC spaces

In Chapter 3, we utilized the ray concept to determine the shared area and the associated R-value of the entirety situated between two adjacent spaces. Nevertheless, this methodology presupposed the existence of a singular entity with a predetermined, constant R-value. It is imperative, however, to underscore the necessity for further research endeavors in subsequent works. Enhanced precision in calculating the interstitial information between spaces can be achieved by incorporating multiple walls, roofs, or floors, each possessing variable R-values. This entails thoroughly examining objects and material characteristics between the two spaces. There is limited research in this area. Some examples of such studies are conducted by G. N. Lilis [139], A. A. Diakité, and S. Zlatanova [181].

6.4.3. Training the energy estimation model with multiple buildings

In developing a BECE model, the prediction model can be trained with multiple buildings across diverse datasets; it is imperative to incorporate a comprehensive set of outdoor parameters. These parameters encompass dynamic weather conditions, such as ambient and operative temperatures, humidity levels, wind speed, and direction, all influencing heating, ventilation, and air conditioning (HVAC) systems. Solar radiation parameters, including solar irradiance and angle, play a crucial role in assessing the impact of sunlight on cooling loads and building surfaces. Additionally, considerations for precipitation, geographical location, time of day, and seasonality are vital to understanding variations in energy demand. Achieving a robust prediction model requires meticulous integration of these outdoor parameters with indoor variables and historical energy consumption data, ensuring a holistic approach that captures the robust model dynamics of each building in the diverse dataset.

6.4.4. Lack of historical data in the training classification model

A drawback of this study lies in the absence of enough historical space-specific data available for training the graph classification model. Including indoor temperature data at the space (space) level across various periods could enhance the accuracy of the classification model. While this investigation addressed the gap by generating historical data using the RETScreen energy simulation software for a year, it is essential to underscore the necessity of gathering indoor energy consumption data for heating and cooling from smart meters and home sensors to enrich further and refine the model training process.

6.4.5. Apply various optimization methods.

In addition to investigating Genetic Algorithms (GAs) for designing optimum buildings in terms of energy consumption, which is implemented in Chapter 5, there is a need for future attempts to explore alternative heuristic optimization methods to identify elite solutions. The architectural and engineering complexities in building design necessitate a comprehensive examination of various heuristic approaches beyond GAs. Exploring alternatives such as simulated annealing, particle swarm optimization, or ant colony optimization can offer valuable insights into their suitability and effectiveness for addressing the intricacies of architectural optimization problems. This diversified exploration aims to uncover methods that may exhibit superior performance or unique advantages in optimizing building designs, ultimately providing architects and engineers with a richer set of tools to achieve optimal and innovative solutions. The comparative evaluation of these methods is essential for advancing state-of-the-art energy-efficient building design optimization, ensuring that the chosen heuristic aligns seamlessly with this domain's specific requirements and challenges for residential and commercial buildings.

6.4.6. Empowering data-driven model by using digital twin platforms.

Empowering energy estimation through data-driven models is revolutionized by integrating digital twin platforms into building design processes. Digital twins, dynamic virtual replicas of physical structures, provide a real-time and comprehensive representation of a building's components and systems. By leveraging this technology in conjunction with data-driven models, designers can tap into a wealth of information to enhance the accuracy and efficiency of energy estimation. Digital twin platforms facilitate the integration of real-world data, such as weather conditions, occupancy patterns, and equipment performance, into the energy estimation model. This enables a more precise understanding of how these variables impact the building's energy consumption. By continuously updating the digital twin with actual operational data, the energy estimation model becomes increasingly refined and reflective of the building's true behavior.

Additionally, digital twin platforms support machine learning algorithms, allowing the energy estimation model to learn and adapt over time. As the digital twin accumulates more operational data, the model can predict energy consumption patterns, identify optimization opportunities, and recommend design modifications for improved energy efficiency. This integration streamlines the energy estimation process and empowers designers to make informed decisions throughout the building design lifecycle. It enables the exploration of various design scenarios, the identification of energy-efficient solutions, and the validation of those solutions against real-world performance data. Ultimately, the synergy between data-driven models and digital twin platforms in building design enhances energy estimation capabilities and contributes to creating more sustainable and efficient buildings. There are some examples of research in this area by A. Clausen *et al*. [182], V. A. Arowoiya *et al.* [183], T. D. Nguyen, and S. Adhikari [184].

BIBLIOGRAPHY

[1] H. Gao, C. Koch, and Y. Wu, “Building information modelling based building energy modelling: A review,” *Applied Energy*. 2019.

[2] A. Borrmann, S. Schraufstetter, and E. Rank, “Implementing Metric Operators of a Spatial Query Language for 3D Building Models: Octree and B-Rep Approaches,” *Journal of Computing in Civil Engineering*, 2009.

[3] R. Jin, B. Zhong, L. Ma, A. Hashemi, and L. Ding, “Integrating BIM with building performance analysis in project life-cycle,” *Automation in Construction*, vol. 106, p. 102861, Oct. 2019.

[4] A. Ghaffarianhoseini *et al.*, “Building Information Modelling (BIM) uptake: Clear benefits, understanding its implementation, risks, and challenges,” *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 1046–1053, Aug. 2017.

[5] A. Porwal and K. N. Hewage, “Building Information Modeling (BIM) partnering framework for public construction projects,” *Automation in Construction*, vol. 31, pp. 204–214, May 2013.

[6] G. Herbert and X. Chen, “A comparison of the usefulness of 2D and 3D representations of urban planning,” *Cartography and Geographic Information Science*, vol. 42, no. 1, pp. 22–32, Jan. 2015.

[7] C. S. Dossick and H. W. Lee, “Optimization of Energy through BIM in the Construction Industry Ritu,” 2023.

[8] I. BuildingSMART, “BuildingSMART International,” 2015. .

[9] H. Kiavarz, M. Jadidi, A. Rajabifard, and G. Sohn, “An Automated Space-Based Graph Generation Framework for Building Energy Consumption Estimation,” *Buildings*, vol. 13, no. 2, p. 350, Jan. 2023.

[10] Y. Wei *et al.*, “A review of data-driven approaches for prediction and classification of building energy consumption,” *Renewable and Sustainable Energy Reviews*. 2018.

[11] S. Freitas, C. Catita, P. Redweik, and M. C. Brito, “Modelling solar potential in the urban environment: State-of-the-art review,” *Renewable and Sustainable Energy Reviews*. 2015.

[12] L. Pérez-Lombard, J. Ortiz, and C. Pout, “A review on buildings energy consumption information,” *Energy and Buildings*, vol. 40, no. 3, pp. 394–398, 2008.

[13] R. Nouvel, M. Zirak, H. Dagtageeri, V. Coors, and U. Eicker, “Urban energy analysis based on 3d city model for national scale applications,” *5th German-Austrian IBPSA Conference*, 2014.

[14] Y. Sun, F. Haghighat, and B. C. M. Fung, “A review of the-state-of-the-art in data-driven approaches for building energy prediction,” *Energy and Buildings*. 2020.

[15] A. Ciccozzi, T. de Rubeis, D. Paoletti, and D. Ambrosini, “BIM to BEM for Building Energy Analysis: A Review of Interoperability Strategies,” *Energies*, vol. 16, no. 23, p. 7845, Nov. 2023.

[16] A. Cemesova, C. J. Hopfe, and R. S. McLeod, “PassivBIM: Enhancing interoperability between BIM and low energy design software,” *Automation in Construction*, 2015.

[17] N. Van Tam, N. Quoc Toan, V. Van Phong, and S. Durdyev, “Impact of BIM-related factors affecting construction project performance,” *International Journal of Building*

Pathology and Adaptation, vol. 41, no. 2, pp. 454–475, Apr. 2023.

[18] V. Bazjanac and L. Berkeley, "Space boundary requirements for modeling of building geometry for energy and other performance simulation," *Proceedings of the CIB W78 2010: 27th International Conference*, 2010.

[19] T. Liebich, "IFC4 – the new buildingSMART Standard," *BuildingSMART*, 2013.

[20] J. Runge and R. Zmeureanu, "Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review," *Energies*, vol. 12, no. 17, p. 3254, Aug. 2019.

[21] B. Pradhan, "A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS," *Computers and Geosciences*, 2013.

[22] E. Mocanu, P. H. Nguyen, M. Gibescu, and W. L. Kling, "Deep learning for estimating building energy consumption," *Sustainable Energy, Grids and Networks*, 2016.

[23] Z. Liu *et al.*, "Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction," *Energy Exploration and Exploitation*. 2019.

[24] H. Kiavarz, M. Jadidi, and P. Esmaili, "A graph-based explanatory model for room-based energy efficiency analysis based on BIM data," *Frontiers in Built Environment*, vol. 9, Sep. 2023.

[25] H. Kiavarz, M. Jadidi, A. Rajabifard, and G. Sohn, "ROOM-BASED ENERGY DEMAND CLASSIFICATION OF BIM DATA USING GRAPH SUPERVISED LEARNING," *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLVI-4/W4-, pp. 97–100, Oct. 2021.

[26] K. Amasyali and N. M. El-Gohary, “A review of data-driven building energy consumption prediction studies,” *Renewable and Sustainable Energy Reviews*. 2018.

[27] N. Desk, “No Title,” 2015. [Online]. Available: http://verticaldistinct.com/new-engineered-solution-from-emc-information-infrastructure-pivotal-and-vmware-accelerates-and-automates-deployment-of-data-lakes.

[28] R. Shan, “Optimization for Heating, Cooling and Lighting Load in Building Façade Design,” *Energy Procedia*, vol. 57, pp. 1716–1725, 2014.

[29] W. Grosrey, “Building Information Modeling,” p. 36, 2013.

[30] S. Tang, D. R. Shelden, C. M. Eastman, P. Pishdad-Bozorgi, and X. Gao, “A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends,” *Automation in Construction*, vol. 101, no. June 2018, pp. 127–139, 2019.

[31] L. Chen and H. Luo, “A BIM-based construction quality management model and its applications,” *Automation in Construction*, vol. 46, pp. 64–73, Oct. 2014.

[32] Y. Liu, S. van Nederveen, and M. Hertogh, “Understanding effects of BIM on collaborative design and construction empirical study in China,” *International Journal of Project Management*, 2017.

[33] Z. Pezeshki and S. A. S. Ivari, “Applications of BIM: A Brief Review and Future Outline,” *Archives of Computational Methods in Engineering*, vol. 25, no. 2, pp. 273–312, 2018.

[34] M. Marcellino, G. Castelblanco, and A. De Marco, “Building information modeling for

construction project management: A literature review," 2023, p. 070012.

[35] 37–41. http://doi.org/10.1037/a0022390 Tuma, J. M., & Pratt, J. M. (1982). Clinical child psychology practice and training: A survey. \ldots of Clinical Child & Adolescent Psychology, 137(August 2012) *et al.*, "Representation of Electronic Mail Filtering Profiles: A User Study," *Intelligent User Interfaces IUI00*, 2000.

[36] G. Gröger, T. Kolbe, C. Nagel, and K.-H. Häfele, "OGC City Geography Markup Language (CityGML) En-coding Standard," *OGC project document*, 2012.

[37] N. Mowafy, M. El Zayat, and M. Marzouk, "Parametric BIM-based life cycle assessment framework for optimal sustainable design," *Journal of Building Engineering*, vol. 75, p. 106898, Sep. 2023.

[38] J. H. Lee, M. J. Ostwald, S. Arasteh, and P. Oldfield, "BIM-Enabled Design Collaboration Processes in Remote Architectural Practice and Education in Australia," *Journal of Architectural Engineering*, vol. 29, no. 1, Mar. 2023.

[39] S. U. Rehman, I. Kim, and J. Choi, "Data-driven integration framework for four-dimensional building information modeling simulation in modular construction: a case study approach," *Journal of Computational Design and Engineering*, vol. 10, no. 6, pp. 2288–2311, Nov. 2023.

[40] Y. Chen, X. Wang, Z. Liu, J. Cui, M. Osmani, and P. Demian, "Exploring Building Information Modeling (BIM) and Internet of Things (IoT) Integration for Sustainable Building," *Buildings*, vol. 13, no. 2, p. 288, Jan. 2023.

[41] J. Crowther and S. O. Ajayi, "Impacts of 4D BIM on construction project performance,"

International Journal of Construction Management, vol. 21, no. 7, pp. 724–737, Jul. 2021.

[42] M. Garramone, N. Moretti, M. Scaioni, C. Ellul, F. Re Cecconi, and M. C. Dejaco, “BIM AND GIS INTEGRATION FOR INFRASTRUCTURE ASSET MANAGEMENT: A BIBLIOMETRIC ANALYSIS,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. VI-4/W1-20, pp. 77–84, Sep. 2020.

[43] H. Kim, Z. Shen, I. Kim, K. Kim, A. Stumpf, and J. Yu, “BIM IFC information mapping to building energy analysis (BEA) model with manually extended material information,” *Automation in Construction*, 2016.

[44] J. Yan, K. Kensek, K. Konis, and D. Noble, “CFD Visualization in a Virtual Reality Environment Using Building Information Modeling Tools,” *Buildings*, vol. 10, no. 12, p. 229, Dec. 2020.

[45] M. J. Sani and A. Abdul Rahman, “GIS AND BIM INTEGRATION AT DATA LEVEL: A REVIEW,” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XLII-4/W9, pp. 299–306, Oct. 2018.

[46] A.-H. Hor, A. Jadidi, and G. Sohn, “BIM-GIS INTEGRATED GEOSPATIAL INFORMATION MODEL USING SEMANTIC WEB AND RDF GRAPHS,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III–4, pp. 73–79, Jun. 2016.

[47] V. Pereira, J. Santos, F. Leite, and P. Escórcio, “Using BIM to improve building energy efficiency – A scientometric and systematic review,” *Energy and Buildings*, vol. 250, p. 111292, Nov. 2021.

[48] H. Deng, D. Fannon, and M. J. Eckelman, “Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata,” *Energy and Buildings*, 2018.

[49] W. Jeong, J. B. Kim, M. J. Clayton, J. S. Haberl, and W. Yan, “Translating building information modeling to building energy modeling using model view definition,” *Scientific World Journal*, 2014.

[50] J. Lee, M. McCuskey Shepley, and J. Choi, “Exploring the effects of a building retrofit to improve energy performance and sustainability: A case study of Korean public buildings,” *Journal of Building Engineering*, vol. 25, p. 100822, Sep. 2019.

[51] Y. Arayici, T. Fernando, V. Munoz, and M. Bassanino, “Interoperability specification development for integrated BIM use in performance-based design,” *Automation in Construction*, 2018.

[52] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, “Physical energy and data-driven models in building energy prediction: A review,” *Energy Reports*, vol. 8, pp. 2656–2671, Nov. 2022.

[53] M. L. Chalal, M. Benachir, M. White, and R. Shrahily, “Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review,” *Renewable and Sustainable Energy Reviews*, vol. 64, pp. 761–776, Oct. 2016.

[54] M. Bourdeau, X. Qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, “Modeling and forecasting building energy consumption: A review of data-driven techniques,” *Sustainable Cities and Society*. 2019.

[55] S. Mohammad E. Saryazdi, A. Etemad, A. Shafaat, and A. M. Bahman, “Data-driven performance analysis of a residential building applying artificial neural network (ANN) and multi-objective genetic algorithm (GA),” *Building and Environment*, vol. 225, p. 109633, Nov. 2022.

[56] R. E. Edwards, J. New, and L. E. Parker, “Predicting future hourly residential electrical consumption: A machine learning case study,” *Energy and Buildings*, 2012.

[57] X. Li, L. Ding, J. Lǔ, G. Xu, and J. Li, “A novel hybrid approach of KPCA and SVM for building cooling load prediction,” in *3rd International Conference on Knowledge Discovery and Data Mining, WKDD 2010*, 2010.

[58] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, “Applying support vector machine to predict hourly cooling load in the building,” *Applied Energy*, 2009.

[59] S. Karatasou, M. Santamouris, and V. Geros, “Modeling and predicting building’s energy use with artificial neural networks: Methods and results,” *Energy and Buildings*, vol. 38, no. 8, pp. 949–958, Aug. 2006.

[60] P. A. González and J. M. Zamarreño, “Prediction of hourly energy consumption in buildings based on a feedback artificial neural network,” *Energy and Buildings*, 2005.

[61] J. M. Hellerstein, “Quantitative Data Cleaning for Large Databases,” *United Nations Economic Commission for Europe*, 2008.

[62] D. Carrión, A. Lorenz, and T. H. Kolbe, “Estimation of the energetic rehabilitation state of buildings for the city of Berlin using a 3D city model represented in cityGML,” *International Archives of Photogrammetry, Remote Sensing and Spatial Information*

Sciences, 2010.

[63] L. C. Garcia and B. Kamsu-Foguem, “BIM-oriented data mining for thermal performance of prefabricated buildings,” *Ecological Informatics*, vol. 51, no. February, pp. 61–72, 2019.

[64] P. Werner and A. Mahdavi, “Building Morphology, Transparence and Energy Performance,” *Building Simulation 2003*, 2003.

[65] U. Eicker, D. Monien, É. Duminil, and R. Nouvel, “Energy performance assessment in urban planning competitions,” *Applied Energy*, 2015.

[66] C. Andrews, “BIM-GIS Interoperability.”

[67] A. A. Bhalekar and R. Sneha, “Assessment of indoor & outdoor air quality of school buildings located close to urban roadway in Manipal (Karnataka),” *International Journal of Civil Engineering and Technology*, 2018.

[68] V. Shekar and M. Krarti, “Control strategies for dynamic insulation materials applied to commercial buildings,” *Energy and Buildings*, 2017.

[69] B. Dave, A. Buda, A. Nurminen, and K. Främling, “A framework for integrating BIM and IoT through open standards,” *Automation in Construction*, vol. 95, no. August 2017, pp. 35–45, 2018.

[70] H. Tashakkori, A. Rajabifard, and M. Kalantari, “A new 3D indoor/outdoor spatial model for indoor emergency response facilitation,” *Building and Environment*, 2015.

[71] R. Olu-Ajayi, H. Alaka, H. Owolabi, L. Akanbi, and S. Ganiyu, “Data-Driven Tools for Building Energy Consumption Prediction: A Review,” *Energies*, vol. 16, no. 6, p. 2574,

Mar. 2023.

[72] S. Torabi Moghadam, J. Toniolo, G. Mutani, and P. Lombardi, “A GIS-statistical approach for assessing built environment energy use at the urban scale,” *Sustainable Cities and Society*, 2018.

[73] J. Mennis and D. Guo, “Spatial data mining and geographic knowledge discovery-An introduction,” *Computers, Environment and Urban Systems*, 2009.

[74] T. Ahmad, H. Chen, Y. Guo, and J. Wang, “A comprehensive overview on the data-driven and large scale based approaches for forecasting of building energy demand: A review,” *Energy and Buildings*, vol. 165, pp. 301–320, Apr. 2018.

[75] F. Biljecki, G. B. M. Heuvelink, H. Ledoux, and J. Stoter, “Propagation of positional error in 3D GIS: estimation of the solar irradiation of building roofs,” *International Journal of Geographical Information Science*, 2015.

[76] C. Fan, F. Xiao, and Y. Zhao, “A short-term building cooling load prediction method using deep learning algorithms,” *Applied Energy*, 2017.

[77] C. Fan, J. Wang, W. Gang, and S. Li, “Assessment of deep recurrent neural network-based strategies for short-term building energy predictions,” *Applied Energy*, 2019.

[78] O. Abdel-Hamid, A. R. Mohamed, H. Jiang, L. Deng, G. Penn, and D. Yu, “Convolutional neural networks for speech recognition,” *IEEE Transactions on Audio, Speech and Language Processing*, 2014.

[79] A. Foucquier, S. Robert, F. Suard, L. Stéphan, and A. Jay, “State of the art in building modelling and energy performances prediction: A review,” *Renewable and Sustainable*

Energy Reviews. 2013.

[80] I. Fernández, C. E. Borges, and Y. K. Penya, “Efficient building load forecasting,” in *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA*, 2011.

[81] L. Xuemei, L. Jin-hu, D. Lixing, X. Gang, and L. Jibin, “Building Cooling Load Forecasting Model Based on LS-SVM,” in *2009 Asia-Pacific Conference on Information Processing*, 2009, pp. 55–58.

[82] Li Xuemei, Deng Yuyan, Ding Lixing, and Jiang Liangzhong, “Building cooling load forecasting using fuzzy support vector machine and fuzzy C-mean clustering,” in *2010 International Conference on Computer and Communication Technologies in Agriculture Engineering*, 2010, pp. 438–441.

[83] D. M. Solomon, R. L. Winter, A. G. Boulanger, R. N. Anderson, and L. L. Wu, “Forecasting energy demand in large commercial buildings using support vector machine regression,” *Department of Computer Science, Columbia University, Tech. Rep. CUCS-040-11*, 2011.

[84] K. Song, N. Kwon, K. Anderson, M. Park, H.-S. Lee, and S. Lee, “Predicting hourly energy consumption in buildings using occupancy-related characteristics of end-user groups,” *Energy and Buildings*, vol. 156, pp. 121–133, Dec. 2017.

[85] J. Massana, C. Pous, L. Burgas, J. Melendez, and J. Colomer, “Short-term load forecasting in a non-residential building contrasting models and attributes,” *Energy and Buildings*, vol. 92, pp. 322–330, Apr. 2015.

[86] Y. K. Penya, C. E. Borges, and I. Fernandez, “Short-term load forecasting in non-residential Buildings,” in *IEEE Africon ’11*, 2011, pp. 1–6.

[87] H. X. Zhao and F. Magoulès, “Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption,” *Journal of Algorithms & Computational Technology*, vol. 4, no. 2, pp. 231–249, Jun. 2010.

[88] D. Liu and Q. Chen, “Prediction of building lighting energy consumption based on support vector regression,” in *2013 9th Asian Control Conference (ASCC)*, 2013, pp. 1–5.

[89] J. Ma and J. C. P. Cheng, “Estimation of the building energy use intensity in the urban scale by integrating GIS and big data technology,” *Applied Energy*, 2016.

[90] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, “Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy,” *Applied Energy*, 2014.

[91] R. Platon, V. R. Dehkordi, and J. Martel, “Hourly prediction of a building’s electricity consumption using case-based reasoning, artificial neural networks and principal component analysis,” *Energy and Buildings*, 2015.

[92] R. Mena, F. Rodríguez, M. Castilla, and M. R. Arahal, “A prediction model based on neural networks for the energy consumption of a bioclimatic building,” *Energy and Buildings*, 2014.

[93] S. S. K. Kwok and E. W. M. Lee, “A study of the importance of occupancy to building cooling load in prediction by intelligent approach,” *Energy Conversion and Management*,

vol. 52, no. 7, pp. 2555–2564, Jul. 2011.

[94] R. Yokoyama, T. Wakui, and R. Satake, “Prediction of energy demands using a neural network with model identification by global optimization,” *Energy Conversion and Management*, vol. 50, no. 2, pp. 319–327, Feb. 2009.

[95] J. Yang, H. Rivard, and R. Zmeureanu, “On-line building energy prediction using adaptive artificial neural networks,” *Energy and Buildings*, vol. 37, no. 12, pp. 1250–1259, Dec. 2005.

[96] Y. Chen, T. Hong, and M. A. Piette, “Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis,” *Applied Energy*, 2017.

[97] A. Krüger and T. H. Kolbe, “BUILDING ANALYSIS FOR URBAN ENERGY PLANNING USING KEY INDICATORS ON VIRTUAL 3D CITY MODELS – THE ENERGY ATLAS OF BERLIN,” *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2012.

[98] H. N. Rafsanjani and A. H. Nabizadeh, “Towards digital architecture, engineering, and construction (AEC) industry through virtual design and construction (VDC) and digital twin,” *Energy and Built Environment*, vol. 4, no. 2, pp. 169–178, Apr. 2023.

[99] D. Robinson *et al.*, “SUNtool - A new modeling paradigm for simulating and optimizing urban sustainability,” *Solar Energy*, 2007.

[100] Z. Li, Y. Han, and P. Xu, “Methods for benchmarking building energy consumption against its past or intended performance: An overview,” *Applied Energy*, vol. 124, pp.

325–334, Jul. 2014.

[101] B. Dong, C. Cao, and S. E. Lee, “Applying support vector machines to predict building energy consumption in the tropical region,” *Energy and Buildings*, vol. 37, no. 5, pp. 545–553, May 2005.

[102] A. H. Neto and F. A. S. Fiorelli, “Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption,” *Energy and Buildings*, vol. 40, no. 12, pp. 2169–2176, Jan. 2008.

[103] M. Yalcintas and S. Akkurt, “Artificial neural networks applications in building energy predictions and a case study for tropical climates,” *International Journal of Energy Research*, vol. 29, no. 10, pp. 891–901, Aug. 2005.

[104] X. Li, C. P. Bowers, and T. Schnier, “Classification of Energy Consumption in Buildings With Outlier Detection,” *IEEE Transactions on Industrial Electronics*, vol. 57, no. 11, pp. 3639–3644, Nov. 2010.

[105] X. Liang, S. Chen, X. Zhu, X. Jin, and Z. Du, “Domain knowledge decomposition of building energy consumption and a hybrid data-driven model for 24-h ahead predictions,” *Applied Energy*, vol. 344, p. 121244, Aug. 2023.

[106] L. Zhang *et al.*, “Analysis of energy consumption prediction for office buildings based on GA-BP and BP algorithm,” *Case Studies in Thermal Engineering*, vol. 50, p. 103445, Oct. 2023.

[107] H. H. Hosamo, M. S. Tingstveit, H. K. Nielsen, P. R. Svennevig, and K. Svidt, “Multiobjective optimization of building energy consumption and thermal comfort based

on integrated BIM framework with machine learning-NSGA II," *Energy and Buildings*, vol. 277, p. 112479, Dec. 2022.

[108] H. X. Li, Z. Ma, H. Liu, J. Wang, M. Al-Hussein, and A. Mills, "Exploring and verifying BIM-based energy simulation for building operations," *Engineering, Construction, and Architectural Management*, vol. 27, no. 8, pp. 1679–1702, Jun. 2020.

[109] S. Durdyev, G. Dehdasht, S. R. Mohandas, and D. J. Edwards, "Review of the Building Information Modelling (BIM) Implementation in the Context of Building Energy Assessment," *Energies*, vol. 14, no. 24, p. 8487, Dec. 2021.

[110] S. Dawood, R. Lord, and N. Dawood, "Development of a visual whole life-cycle energy assessment framework for the built environment," in *Proceedings of the 2009 Winter Simulation Conference (WSC)*, 2009, pp. 2653–2663.

[111] J. A. Fonseca and A. Schlueter, "Integrated model for characterization of spatiotemporal building energy consumption patterns in neighborhoods and city districts," *Applied Energy*, 2015.

[112] R. Nouvel, C. Schulte, U. Eicker, D. Pietruschka, and V. Coors, "CITYGML-based 3D city model for energy diagnostics and urban energy policy support," in *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association*, 2013.

[113] D. L. Marino, K. Amarasinghe, and M. Manic, "Building energy load forecasting using Deep Neural Networks," in *IECON Proceedings (Industrial Electronics Conference)*, 2016.

[114] K. P. Amber, R. Ahmad, M. W. Aslam, A. Kousar, M. Usman, and M. S. Khan, “Intelligent techniques for forecasting electricity consumption of buildings,” *Energy*, 2018.

[115] G. Zhang, C. Tian, C. Li, J. J. Zhang, and W. Zuo, “Accurate forecasting of building energy consumption via a novel ensembled deep learning method considering the cyclic feature,” *Energy*, 2020.

[116] A. Waqar, I. Othman, N. Saad, M. Azab, and A. M. Khan, “BIM in green building: Enhancing sustainability in the small construction project,” *Cleaner Environmental Systems*, vol. 11, p. 100149, Dec. 2023.

[117] K.-U. Ahn, Y.-J. Kim, C.-S. Park, I. Kim, and K. Lee, “BIM interface for full vs. semi-automated building energy simulation,” *Energy and Buildings*, vol. 68, pp. 671–678, Jan. 2014.

[118] V. Bazjanac, “Implementation of semi-automated energy performance simulation: building geometry,” *CIB W*, vol. 78, no. 2009, pp. 595–602, 2009.

[119] J. Lin, Z. Hu, and J. Zhang, “BIM Oriented Intelligent Data Mining and Representation,” *Proceedings of the 30th CIB W78 International Conference*, 2013.

[120] G. K, “Big Data in Building Information Modeling Research: Survey and Exploratory Text Mining,” *MOJ Civil Engineering*, 2017.

[121] M. Needham and A. E. Hodler, *Graph Algorithms - Practical Examples in Apache Spark and Neo4j*. 2019.

[122] F. Xia *et al.*, “Graph Learning: A Survey,” *IEEE Transactions on Artificial Intelligence*,

vol. 2, no. 2, pp. 109–127, Apr. 2021.

[123] S. Daum, A. Borrmann, C. Langenhan, and F. Petzold, “Automated generation of building fingerprints using a spatial-semantic query language for building information models,” in *eWork and eBusiness in Architecture, Engineering, and Construction - Proceedings of the 10th European Conference on Product and Process Modelling, ECPPM 2014*, 2015.

[124] Z. Pezeshki, A. Soleimani, and A. Darabi, “Application of BEM and using BIM database for BEM: A review,” *Journal of Building Engineering*, vol. 23, pp. 1–17, May 2019.

[125] F. C. Collins, A. Braun, M. Ringsquandl, D. M. Hall, and A. Borrmann, “Assessing IFC classes with means of geometric deep learning on different graph encodings,” *Proceedings of the 2021 European Conference on Computing in Construction*, vol. 2, pp. 332–341, 2021.

[126] H. Y. Chong, C. Y. Lee, and X. Wang, “A mixed review of the adoption of Building Information Modelling (BIM) for sustainability,” *Journal of Cleaner Production*. 2017.

[127] A. Abou Diakité and S. Zlatanova, “Valid Space Description in BIM for 3D Indoor Navigation,” *International Journal of 3-D Information Modeling*, 2017.

[128] A. Andriamamonjy, D. Saelens, and R. Klein, “An automated IFC-based workflow for building energy performance simulation with Modelica,” *Automation in Construction*, 2018.

[129] N. Fumo, “A review on the basics of building energy estimation,” *Renewable and Sustainable Energy Reviews*. 2014.

[130] S. Yumusak, E. Dogdu, H. Kodaz, A. Kamilaris, and P. Y. Vandenbussche, “SpEnD:

Linked data SPARQL endpoints discovery using search engines," *IEICE Transactions on Information and Systems*, 2017.

[131] C. Jin, M. Xu, L. Lin, and X. Zhou, "Exploring BIM Data by Graph-based Unsupervised Learning," in *ICPRAM 2018 - Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods*, 2018.

[132] V. Bazjanac and T. Maile, "IFC HVAC interface to EnergyPlus - A case of expanded interoperability for energy simulation," in *Building Sustainability and Performance Through Simulation*, 2004.

[133] R. J. Hitchcock and J. Wong, "Transforming IFC architectural view BIMs for energy simulation: 2011," in *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, 2011.

[134] P. Remmen *et al.*, "An open framework for integrated bim-based building performance simulation using Modelica," in *14th International Conference of IBPSA - Building Simulation 2015, BS 2015, Conference Proceedings*, 2015.

[135] Y. N. Bahar, C. Pere, J. Landrieu, and C. Nicolle, "A thermal simulation tool for building and its interoperability through the Building Information Modeling (BIM) platform," *Buildings*, 2013.

[136] B. Dong, K. P. Lam, Y. C. Huang, and G. M. Dobbs, "A comparative study of the IFC and gbXML informational infrastructures for data exchange in computational design support environments," in *IBPSA 2007 - International Building Performance Simulation Association 2007*, 2007.

[137] V. Bazjanac, T. Maile, and C. Nytsch-Geusen, “Generation of building geometry for energy performance simulation using Modelica,” in *Proceedings of CESBP / BauSim 2016*, 2016.

[138] M. Gunduz *et al.*, “A review of building information modeling (BIM) and the Internet of things (IoT) devices integration: Present status and future trends,” *Automation in Construction*, vol. 95, no. June 2018, pp. 127–139, 2019.

[139] G. N. Lilis, G. I. Giannakis, and D. V. Rovas, “Automatic generation of second-level space boundary topology from IFC geometry inputs,” *Automation in Construction*, vol. 76, pp. 108–124, 2017.

[140] S. Kim, P. A. Zadeh, S. Staub-French, T. Froese, and B. T. Cavka, “Assessment of the Impact of Window Size, Position and Orientation on Building Energy Load Using BIM,” in *Procedia Engineering*, 2016.

[141] J. Lee and M. -P. Kwan, “A combinatorial data model for representing topological relations among 3D geographical features in micro-spatial environments,” *International Journal of Geographical Information Science*, vol. 19, no. 10, pp. 1039–1056, Nov. 2005.

[142] Z. Yu, B. C. M. Fung, and F. Haghighat, “Extracting knowledge from building-related data - A data mining framework,” *Building Simulation*, vol. 6, no. 2, pp. 207–222, 2013.

[143] G. Boeing, “OSMnx: A Python package to work with graph-theoretic OpenStreetMap street networks,” *The Journal of Open Source Software*, vol. 2, no. 12, p. 215, Apr. 2017.

[144] Scellato, “NetworkX : Network Analysis with Python Today ’ 's Outline,” no. February, pp. 1–51, 2013.

[145] L. Saini, C. S. Meena, B. P. Raj, N. Agarwal, and A. Kumar, "Net Zero Energy Consumption building in India: An overview and initiative toward sustainable future," *International Journal of Green Energy*, vol. 19, no. 5, pp. 544–561, Apr. 2022.

[146] A. S. Ahmad *et al.*, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renewable and Sustainable Energy Reviews*. 2014.

[147] U. Atila, I. R. Karas, and A. A. Rahman, "A 3D-GIS implementation for realizing 3D network analysis and routing simulation for evacuation purpose," in *Lecture Notes in Geoinformation and Cartography*, 2013.

[148] Z. Wang and R. S. Srinivasan, "A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models," *Renewable and Sustainable Energy Reviews*, vol. 75, pp. 796–808, Aug. 2017.

[149] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Applying support vector machine to predict hourly cooling load in the building," *Applied Energy*, vol. 86, no. 10, pp. 2249–2256, Oct. 2009.

[150] Z. Wang, R. Sacks, and T. Yeung, "Exploring graph neural networks for semantic enrichment: Room type classification," *Automation in Construction*, no. November, p. 104039, 2021.

[151] P. Pauwels and W. Terkaj, "EXPRESS to OWL for the construction industry: Towards a recommendable and usable ifcOWL ontology," *Automation in Construction*, vol. 63, pp. 100–133, Mar. 2016.

[152] A. Khalili and D. K. H. Chua, "IFC-Based Graph Data Model for Topological Queries on

Building Elements," *Journal of Computing in Civil Engineering*, vol. 29, no. 3, p. 04014046, May 2015.

[153] D. Simeone and S. Cursi, "A Platform for Enriching BIM Representation Through Semantic Web Technologies," in *Lean and Computing in Construction Congress - Volume 1: Proceedings of the Joint Conference on Computing in Construction*, 2017, pp. 423–430.

[154] M. Saad, Y. Zhang, J. Tian, and J. Jia, "A graph database for life cycle inventory using Neo4j," *Journal of Cleaner Production*, vol. 393, p. 136344, Mar. 2023.

[155] A. Ismail, B. Strug, and G. Ślusarczyk, "Building Knowledge Extraction from BIM/IFC Data for Analysis in Graph Databases," 2018, pp. 652–664.

[156] W. Shi and R. Rajkumar, "Point-GNN: Graph Neural Network for 3D Object Detection in a Point Cloud," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 1708–1716.

[157] W. L. Hamilton, R. Ying, and J. Leskovec, "Inductive representation learning on large graphs," in *Advances in Neural Information Processing Systems*, 2017.

[158] M. Xu, "Understanding Graph Embedding Methods and Their Applications," *SIAM Review*, vol. 63, no. 4, pp. 825–853, Jan. 2021.

[159] M. Z. Naser, "An engineer's guide to eXplainable Artificial Intelligence and Interpretable Machine Learning: Navigating causality, forced goodness, and the false perception of inference," *Automation in Construction*, vol. 129, p. 103821, Sep. 2021.

[160] J. Wang, J. Wiens, and S. Lundberg, "Shapley Flow: A Graph-based Approach to

Interpreting Model Predictions," vol. 130, 2020.

[161] S. Celik, R. Family, and M. P. Menguc, "Analysis of perlite and pumice-based building insulation materials," *Journal of Building Engineering*, 2016.

[162] R. Goel, X. Athalye, H. Wang, and M. Zhang, "Enhancements to ASHRAE standard 90.1 prototype building models," 2014.

[163] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," in *34th International Conference on Machine Learning, ICML 2017*, 2017.

[164] A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, and Y. LeCun, "The loss surfaces of multilayer networks," in *Journal of Machine Learning Research*, 2015.

[165] L. Wu *et al.*, "Learning to teach with dynamic loss functions," in *Advances in Neural Information Processing Systems*, 2018.

[166] L. L. Linyuan and T. Zhou, "Link prediction in complex networks: A survey," *Physica A: Statistical Mechanics and its Applications*. 2011.

[167] S. M. Oh, Y. J. Kim, C. S. Park, and I. H. Kim, "Process-driven BIM-based optimal design using the integration of EnergyPlus, genetic algorithm, and Pareto optimality," *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, pp. 894–901, 2011.

[168] S. Gou, V. M. Nik, J.-L. Scartezzini, Q. Zhao, and Z. Li, "Passive design optimization of newly-built residential buildings in Shanghai for improving indoor thermal comfort while reducing building energy demand," *Energy and Buildings*, vol. 169, pp. 484–506, Jun.

2018.

[169] B. Abbasnejad, M. P. Nepal, A. Ahankoob, A. Nasirian, and R. Drogemuller, “Building Information Modelling (BIM) adoption and implementation enablers in AEC firms: a systematic literature review,” *Architectural Engineering and Design Management*, vol. 17, no. 5–6, pp. 411–433, Nov. 2021.

[170] J. Abualdenien, B. Harder, and J. Clever, “A prescriptive parametric model supporting performance-based design exploration at the early stages,” *Proc. of the 32nd Forum Bauinformatik*, no. July, 2021.

[171] Y. Song *et al.*, “Trends and Opportunities of BIM-GIS Integration in the Architecture, Engineering and Construction Industry: A Review from a Spatio-Temporal Statistical Perspective,” *ISPRS International Journal of Geo-Information*, vol. 6, no. 12, p. 397, 2017.

[172] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, “A Comprehensive Survey on Graph Neural Networks,” *IEEE Transactions on Neural Networks and Learning Systems*, 2020.

[173] E. D. Goodman, “Introduction to genetic algorithms,” in *Proceedings of the Companion Publication of the 2014 Annual Conference on Genetic and Evolutionary Computation*, 2014, pp. 205–226.

[174] W. Yu, B. Li, H. Jia, M. Zhang, and D. Wang, “Application of the multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design,” *Energy and Buildings*, vol. 88, pp. 135–143, Feb. 2015.

[175] W. T. Chen, H. C. Merrett, S.-S. Liu, N. Fauzia, and F. N. Liem, “A Decade of Value Engineering in Construction Projects,” *Advances in Civil Engineering*, vol. 2022, pp. 1–13, Mar. 2022.

[176] M. Hamdy, A. Hasan, and K. Siren, “Applying a multi-objective optimization approach for Design of low-emission cost-effective dwellings,” *Building and Environment*, vol. 46, no. 1, pp. 109–123, Jan. 2011.

[177] R. Dondi, G. Mauri, and I. Zoppis, “Graph Algorithms,” in *Encyclopedia of Bioinformatics and Computational Biology: ABC of Bioinformatics*, 2018.

[178] K. Xu, C. Li, Y. Tian, T. Sonobe, K. I. Kawarabayashi, and S. Jegelka, “Representation learning on graphs with jumping knowledge networks,” in *35th International Conference on Machine Learning, ICML 2018*, 2018.

[179] T. N. Carnes, “Indoor sound criteria according to the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)—An introduction,” *The Journal of the Acoustical Society of America*, vol. 118, no. 3_Supplement, pp. 1843–1843, Sep. 2005.

[180] D. Listokin and D. B. Hattis, “Building Codes and Housing,” *Cityscape*, vol. 8, no. 1, pp. 21–67, 2005.

[181] A. A. Diakité and S. Zlatanova, “EXTRACTION OF THE 3D FREE SPACE FROM BUILDING MODELS FOR INDOOR NAVIGATION,” *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-2/W1, pp. 241–248, Oct. 2016.

[182] A. Clausen *et al.*, "A digital twin framework for improving energy efficiency and occupant comfort in public and commercial buildings," *Energy Informatics*, vol. 4, no. S2, p. 40, Sep. 2021.

[183] V. A. Arowoiya, R. C. Moehler, and Y. Fang, "Digital twin technology for thermal comfort and energy efficiency in buildings: A state-of-the-art and future directions," *Energy and Built Environment*, May 2023.

[184] T. D. Nguyen and S. Adhikari, "The Role of BIM in Integrating Digital Twin in Building Construction: A Literature Review," *Sustainability*, vol. 15, no. 13, p. 10462, Jul. 2023.

APPENDICES

APPENDIX A: PYTHON CODE FOR CHAPTER 3

https://github.com/HamidKiavarz/IFC_ObjectOriented_Framework.git

APPENDIX B: PYTHON CODE FOR CHAPTER 4

https://github.com/HamidKiavarz/graphsage-_classification.git

APPENDIX C: PYTHON CODE FOR CHAPTER 5

https://github.com/HamidKiavarz/GA-Optimization.git

APPENDIX D: ENERGY CONSUMTION ESTIMATION FOR ALL SPACES (SPACES)

ID	Space Name	Description	Floor	Degree	TWA	EWA	Direction	IWA	EWINA	Floor_Area	Lighting Load (W/m2)	Plug Load (W/m2)	Plug Load (W)	Roof_Area	Type	People Density #/100 m2	# of People	People Vent (cfm/person)	Area Vent Rate (cfm/sqft)	Total Fresh Air (cfm)	Internal Heat Gain (W)	Heating (kWh)	Cooling (kWh)	Energy (kWh)	EUI (kWh/m2)
1	149	Space	1	3	162.788	0		162.788	0	25.97	10	5	129.85	25.97	Office	5	1.3	5	0.06	23.3	169	184.49	112.84	1728.54	78.01
2	150	Space	1	5 + 1	153.951	7.522	S	146.429	10.851	55.013	13	5	275.065	55.013	Conference Space	50	27.5	5	0.06	173.1	3576	2573.18	1402.99	12649.86	305.60
3	151	Space	1	4	44.885	0	S	37.439	0	10.768	10	5	53.84	10.768	Office	5	0.5	5	0.06	9.7	70	92.20	46.60	713.00	79.10
4	152	Space	1	7 + 1	47.189	2.75	S	27.269	5.125	14.03	10	5	70.15	14.03	Office	5	0.7	5	0.06	12.6	91	600.00	316.00	931.00	131.65
5	153	Space	1	5 + 1	85.696	16.326	S	64.084	20.671	112.208	10	5	561.04	112.208	Office	5	5.6	5	0.06	100.6	729	2075.00	1521.00	7459.00	98.52
6	154	Space	1	6	201.208	112.626	0.25 S, 0.25 N, 0.5 E	88.582	157.752	194.652	10	5	973.26	194.652	Office	5	9.7	5	0.06	174.6	1265	10608.00	4644.00	12941.00	303.60
7	155	Space	1	4	218.552	23.462	E	195.09	9.929	61.176	13	5	305.88	61.176	Conference Space	50	30.6	5	0.06	192.5	3976	2864.49	1329.00	14067.47	298.50
8	156	Space	1	2	102.042	35.111	N, E	66.931	23.929	110.864	8	5	554.32	110.864	Lounge	100	110.9	7.5	0.18	1046.6	14412	16897.00	4767.00	42254.00	576.54
9	157	Space	1	3	107.068	0	N, E	107.068	0	58.53	8	5	292.65	58.53	Lounge	100	58.5	5	0.06	330.5	7609	277.00	1569.00	22307.00	535.60
10	158	Kitchen	1	3 + 1	60.022	16.987	N	43.035	1.7	18.255	10	5	91.275	18.255	Kitchen	20	3.7	5	0.06	30.1	475	1348.00	317.00	2144.11	208.66
11	159	Space	1	2	66.87	14.71	N, W	45.47	19.172	39.677	13	5	198.385	39.677	Meeting Space	50	19.8	5	0.06	124.9	2579	2996.00	1396.00	9138.00	341.00
12	160	Space	1	3	120.346	11.276	W	78.267	12.986	61.274	13	5	306.37	61.274	Meeting Space	50	30.6	5	0.06	192.8	3983	2864.00	1634.00	14092.00	303.39
13	161	Space	1	5	125.103	8.377	W	116.726	43.091	185.959	10	5	929.795	185.959	Office	5	9.3	5	0.06	166.8	1209	3980.00	2986.00	12365.00	103.95
14	162	Main Lobby	1	6	72.89012	2.086	W	70.80412	8.668	38.23456	7	5	191.1728	38.23456	Main Lobby	10	3.8	5	0.06	43.8	497	830.00	190.00	2770.00	99.12
15	163	Lift Lobby	1	4	45.65904	0	S	45.65904	0	28.512	7	5	142.56	28.512	Lift Lobby	10	2.9	5	0.06	32.7	371	276.74	142.01	2067.46	87.20
16	164	Space	1	4	72.489	0	S	62.72	0	35.744	13	5	178.72	35.744	Meeting Space	50	17.9	5	0.06	112.5	2323	1660.00	562.00	8216.00	292.02
17	165	Space	1	6	88.611	0	N, W	70.171	0	49.26	13	5	246.3	49.26	Meeting Space	50	24.6	5	0.06	155.0	3202	2306.20	774.60	11329.00	292.53
18	166	Space	1	8	55.67	0	S, W	55.67	0	20.164	10	5	100.82	20.164	Office	5	1.0	5	0.06	18.1	131	92.20	87.60	1342.00	75.47
19	167	Space	1	4	46.622	0	W	34.30	0	11.92	13	5	59.6	11.92	Meeting Space	50	6.0	5	0.06	37.5	775	553.00	188.00	2745.00	292.45
20	168	Space	1	4	63.547	0	W	38.47	0	11.81	10	5	59.05	11.81	Office	5	0.6	5	0.06	10.6	77	92.20	51.40	787.00	78.80
21	169	Space	1	4	50.693	0	S, W	38.47	0	11.208	10	5	56.04	11.208	Office	5	0.6	5	0.06	10.1	73	92.20	48.70	745.00	79.04

22	170	Space	1	4	53.487	0	S	45.49	0	11.81	10	5	59.05	11.81	Office	5	0.6	5	0.06	10.6	77	92.20	51.30	785.00	78.62
23	171	Space	1	4	61.64	0	N	54.10	0	11.04	10	5	55.2	11.04	Office	5	0.6	5	0.06	9.9	72	92.20	48.00	735.00	79.28
24	172	Space	1	4	99.368	0	E	86.66	0	42.154	10	5	210.77	42.154	Office	5	2.1	5	0.06	37.8	274	184.00	183.00	2803.00	75.20
25	173	Space	1	3	48.322	0	E	48.32	0	9.316	10	5	46.58	9.316	Office	5	0.5	5	0.06	8.4	61	92.20	40.50	753.00	95.07
26	174	Exit Stairway 2	1	5	200.49	0	N	200.49	5.829	47.138	7	5	235.69	47.138	Corridor+ Exit Stairway	0	0.0	5	0.06	30.5	0	767.37	304.52	1813.53	61.21
27	175	Space	1	4	118.833	0	S	96.40		67.093	13	5	335.465	67.093	Lobby	30	20.1	5	0.06	144.0	2617	1845.00	732.00	10880.18	200.58
28	176	Space	1	3	87.416	0	S	87.42	0	55.771	10	5	278.855	55.771	Washspace	0	0.0	5	0.06	0.0	0	0.00	0.00	2764.00	49.56
29	177	Space	1	4	47.502	13.5	W	27.55	15.094	21.98	10	5	109.9	21.98	Office	5	1.1	5	0.06	19.7	143	2017.00	876.00	1463.00	198.18
30	178	Space	1	5	125.451	4.514	N	110.06	0	30.747	7	5	153.735	30.747	Corridor	0	0.0	5	0.06	19.9	0	106.00	80.40	1185.00	44.60
31	179	Space	1	8	44.623	0	N	44.62	0	59.987	7	5	299.935	59.987	Corridor	0	0.0	5	0.06	38.8	0	146.00	2315.00	2461.00	82.05
32	180	Space	1	4	52.527	0	W	44.77	0	12.832	10	5	64.16	12.832	Office	5	0.6	5	0.06	11.5	83	92.20	53.00	850.00	77.56
33	181	Space	1	4	61.862	0	W	61.86	0	3.961	7	5	19.805	3.691	Storage	0	0.0	5	0.06	2.4	0	0.00	9.47	151.27	40.58
34	182	Vestibule	1	3 + 1 (outdoor)	33.093	25.994	W	7.10	0	8.65	7	5	43.25	8.65	Vestibule	0	0.0	5	0.06	5.6	0	1492.72	53.96	335.17	217.56
35	183	Space	1	2 + 1(outdoor)	21.564	6.516	S, W	8.60	8.548	6.663	10	5	33.315	6.663	Office	5	0.3	5	0.06	6.0	43	769.02	351.32	443.48	234.70
36	184	Exit Stairway 1	1	4	63.832	0	S, W	63.83	0	14.168	7	5	70.8	14.168	Exit Stairway 1	0	0.0	5	0.06	9.2	0	0.00	0.00	363.00	25.62
39	200	Space	2	3	96.404	0	S, W	96.40	0	208.6802974	13	5	1043.4		Meeting Space	50	104.3	5	0.06	656.7	13564	184.49	112.84	1728.54	358.62
40	201	Space	2	4	253.0816	165.666	N, W	87.42	46.9	250.9228625	10	5	1254.6		Office	5	12.5	5	0.06	225.0	1631	2573.18	1402.99	12649.86	118.35
41	202	Space	2	4	129.7112	102.159	S, W	27.55	36.43	147.3094796	10	5	736.5		Office	5	7.4	5	0.06	132.1	958	92.20	46.60	713.00	125.78
42	203	Space	2	6	181.502	71.44	S	110.06	26.4	176.0343866	10	5	880.2		Office	5	8.8	5	0.06	157.9	1144	600.00	316.00	931.00	98.99
43	204	Space	2	2	44.623		W	44.62		21.30576208	10	5	106.5		Washspace	0	0.0	5	0.06	13.8	0	2075.00	1521.00	7459.00	52.66
44	205	Wash Space	2	5	44.771		N	44.77		20.67286245	10	5	103.4		Washspace	0	0.0	5	0.06	13.4	0	10608.00	4644.00	12941.00	52.73
45	206	Space	2	4	61.862		N	61.86		19.64498141	10	5	98.2		Office	5	1.0	5	0.06	17.6	128	2864.49	1329.00	14067.47	75.47
46	207	Space	2	2	7.099		W	7.10		20.85223048	10	5	104.3		Office	5	1.0	5	0.06	18.7	136	16897.00	4767.00	42254.00	76.98

47	208	Space	2	3	8.598		W	8.60		14.41263941	10	5	72.1	Office	5	0.7	5	0.06	12.9	94	277.00	1569.00	22307.00	77.26
48	209	Space	2	2	68.3		W	68.30		9.251858736	7	5	46.3	Storage	0	0.0	5	0.06	6.0	0	1348.00	317.00	2144.11	41.31
49	211	Space	2	2	101.624	74.072	N	27.55	26.46	131.7416357	10	5	658.7	Office	5	6.6	5	0.06	118.1	856	2996.00	1396.00	9138.00	382.53
50	212	Space	2	3	221.8092	111.747	%15N, %75E	110.06	38.262	156.267658	10	5	781.3	Office	5	7.8	5	0.06	140.1	1016	2864.00	1634.00	14092.00	131.41
51	213	Space	2	5	44.623		N	44.62		41.26765799	13	5	206.3	Meeting Space	50	20.6	5	0.06	129.9	2682	3980.00	2986.00	12365.00	292.38
52	214	Space	2	4	208.031	163.26	N, E	44.77	50.32	208.679368	10	5	1043.4	Office	5	10.4	5	0.06	187.1	1356	830.00	190.00	2770.00	121.74
53	215	Space	2	4	61.862		S	61.86		26.97118959	13	5	134.9	Meeting Space	50	13.5	5	0.06	84.9	1753	276.74	142.01	2067.46	294.57
54	216	Space	2	3	7.099		W	7.10		22.59851301	13	5	113.0	Meeting Space	50	11.3	5	0.06	71.1	1469	1660.00	562.00	8216.00	299.88
55	217	Space	2	6	8.598		N	8.60		30.93122677	10	5	154.7	Office	5	1.5	5	0.06	27.7	201	2306.20	774.60	11329.00	76.90
56	218	Space	2	2	63.832		N	63.83		10.22304833	13	5	51.1	Meeting Space	50	5.1	5	0.06	32.2	664	92.20	87.60	1342.00	289.55
57	220	Space	2	8	96.404		W	96.40		20.64498141	13	5	103.2	Meeting Space	50	10.3	5	0.06	65.0	1342	553.00	188.00	2745.00	291.22
58	221	Space	2	4	87.416		W	87.42		53.01208178	8	5	265.1	Lounge	100	53.0	7.5	0.06	431.9	6892	92.20	51.40	787.00	568.55
59	222	Space	2	4	28.52		W	27.55		21.20074349	13	5	106.0	Meeting Space	50	10.6	5	0.06	66.7	1378	92.20	48.70	745.00	313.13
60	223	Space	2	2	28.96		N	27.55		10.75092937	10	5	53.8	Office	5	0.5	5	0.06	9.6	70	92.20	51.30	785.00	80.40
61	224	Space	2	4	110.062		S	110.06		10.57713755	10	5	52.9	Office	5	0.5	5	0.06	9.5	69	92.20	48.00	735.00	81.21
62	225	Space	2	3	44.623		W	44.62		10.69330855	10	5	53.5	Office	5	0.5	5	0.06	9.6	70	184.00	183.00	2803.00	81.13
63	226	Space	2	3	44.771		N	44.77		10.54182156	13	5	52.7	Meeting Space	50	5.3	5	0.06	33.2	685	92.20	40.50	753.00	289.17
64	227	Space	2	5	76.902	15.04	N	61.86	8.94	62.29832714	7	5	311.5	Corridor	0	0.0	5	0.06	40.3	0	767.37	304.52	1813.53	64.53
65	228	Space	2	4	56.23		W	7.10		16.35315985	7	5	81.8	Corridor	0	0.0	5	0.06	10.6	0	1845.00	732.00	10880.18	82.83
66	229	Space	2	3	8.598		W	8.60		11.46096654	10	5	57.3	Office	5	0.6	5	0.06	10.3	74	0.00	0.00	2764.00	78.88
67	230	Space	2	4	63.832		W	63.83		11.71524164	10	5	58.6	Office	5	0.6	5	0.06	10.5	76	2017.00	876.00	1463.00	78.62
68	231	Space	2	6	96.404		N	96.40		9.303903346	7	5	46.5	Corridor	0	0.0	5	0.06	6.0	0	106.00	80.40	1185.00	41.15
69	232	Space	2	5	87.416		S	87.42		13.48327138	13	5	67.4	Meeting Space	50	6.7	5	0.06	42.4	876	146.00	2315.00	2461.00	373.49
70	233	Space	2	2	27.552		W	27.55		34.15799257	7	5	170.8	Corridor	0	0.0	5	0.06	22.1	0	92.20	53.00	850.00	45.96
71	234	Exit Stairway	2	4	56.5		N	25.60		13.12267658	7	5	65.6	Stairwell	0	0.0	5	0.06	8.5	0	0.00	9.47	151.27	41.13
72	235	Space	2	3	110.062		N	110.06		4.343866171	10	5	21.7	Office	5	0.2	5	0.06	3.9	28	1492.72	53.96	335.17	91.60

73	236	Space	2	8	44.623		W	44.62		4.558550186	10	5	22.8	Office	5	0.2	5	0.06	4.1	30	769.02	351.32	443.48	91.72
74	237	Space	2	3	44.771		W	44.77		4.442379182	10	5	22.2	Office	5	0.2	5	0.06	4.0	29	0.00	0.00	363.00	91.26
75	238	Space	2	4	61.862		W	61.86		32.30576208	7	5	161.5	Corridor	0	0.0	5	0.06	20.9	0	184.49	112.84	1728.54	84.54
76	239	Space	2	5	56.23		N	8.90		11.59572491	13	5	58.0	Meeting Space	50	5.8	5	0.06	36.5	754	2573.18	1402.99	12649.86	293.49
77	240	Space	2	4	8.69		S	8.69		9.348513011	7	5	46.7	Corridor	0	0.0	5	0.06	6.0	0	92.20	46.60	713.00	40.97
78	241	Space	2	4	63.832		W	63.83		10.74349442	10	5	53.7	Office	5	0.5	5	0.06	9.6	70	600.00	316.00	931.00	79.28
79	242	Space	2	3	100.02		N	100.02		5.752788104	10	5	28.8	Office	5	0.3	5	0.06	5.2	37	2075.00	1521.00	7459.00	86.95
80	243	Space	2	7	89.5		N	89.50		10.16728625	10	5	50.8	Office	5	0.5	5	0.06	9.1	66	10608.00	4644.00	12941.00	80.00
81	245	Space	2	6	25.6		W	25.60		2.298327138	7	5	11.5	Storage	0	0.0	5	0.06	1.5	0	2864.49	1329.00	14067.47	41.16
82	246	Space	2	4	56.5		W	44.77		2.249070632	7	5	11.2	Storage	0	0.0	5	0.06	1.5	0	16897.00	4767.00	42254.00	40.55
83	247	Space	2	2	61.862		W	61.86		2.178438662	7	5	10.9	Storage	0	0.0	5	0.06	1.4	0	277.00	1569.00	22307.00	41.41
84	248	Space	2	5	185.20		N	8.90		7.607806691	7	5	38.0	Storage	0	0.0	5	0.06	4.9	0	1348.00	317.00	2144.11	41.09
85	249	Space	2	3	65.89		S, W	8.69		20.36245353	7	5	101.8	Corridor	0	0.0	5	0.06	13.2	0	2996.00	1396.00	9138.00	41.23
86	250	Space	2	4	190.58		N, W	63.83		14.78624535	7	5	73.9	Corridor	0	0.0	5	0.06	9.6	0	2864.00	1634.00	14092.00	41.19
87	340	Space	3	4	195.65	106.25	N	100.02	20.47	157.581	13	5	787.9	Office	5	7.9	5	0.06	141.3	1024	3980.00	2986.00	12365.00	49.56
88	302	Space	3	5	228.62	93.62	S	89.50	30.17	210.79	13	5	1054.0	Office	50	105.4	5	0.06	663.3	13701	830.00	190.00	2770.00	154.48
89	311	Space	3	3	201.36		S	25.60		20.36245353	10	5	101.8	Lab	5	1.0	5	0.06	18.3	132	276.74	142.01	2067.46	118.68
90	346	Space	3	5 + 1	198.69		W	44.77		14.78624535	13	5	73.9	Office	5	0.7	5	0.06	13.3	96	1660.00	562.00	8216.00	292.73
91	341	Space	3	4	285.69		N	61.86		157.581	13	5	787.9	Meeting Space	5	7.9	5	0.06	141.3	1024	2306.20	774.60	11329.00	49.58
92	303	Space	3	7 + 1	136.56		N	8.90		110.91	7	5	554.6	Office	5	5.5	5	0.06	99.5	721	92.20	87.60	1342.00	135.48
93	312	Space	3	5 + 1	256.16	128.61	W	8.69	38.553	152.337	13	5	761.7	Office	50	76.2	5	0.06	479.4	9902	553.00	188.00	2745.00	293.92
94	347	Space	3	6	282.41	175.54	E	63.83	34.8	211.94	10	5	1059.7	Office	100	211.9	5	0.06	1196.8	27552	92.20	51.40	787.00	49.51
95	334	Space	3	4	298.29	169.05	E	100.02	49	259.895	10	5	1299.5	Office	100	259.9	5	0.06	1467.6	33786	92.20	48.70	745.00	289.49
96	301	Space	3	2	134.11	4.72	S, E	89.50	0	47.933	10	5	239.7	Corridor	20	9.6	5	0.06	78.9	1246	92.20	51.30	785.00	292.92
97	313	Space	3	3	125.60		N, E	25.60	0	50.23	13	5	251.2	Entrance	50	25.1	5	0.06	158.1	3265	92.20	48.00	735.00	132.70
98	306	Space	3	3 + 1	281.40		S, W	25.60	0	69.56	7	5	347.8		50	34.8	5	0.06	218.9	4521	184.00	183.00	2803.00	78.63
99	343	Space	3	2	265.90		N, E	44.77	0	120.36	7	5	601.8	Kitchen	5	6.0	5	0.06	107.9	782	92.20	40.50	753.00	291.58
100	307	Space	3	3	209.05	80.51	S, W	61.86	26.51	132.568	10	5	662.8	Office	10	13.3	5	0.06	152.0	1723	767.37	304.52	1813.53	49.47
101	314	Space	3	5	256.16		W	8.90		157.581	10	5	787.9	Meeting Space	10	15.8	5	0.06	180.7	2049	1845.00	732.00	10880.18	82.29

102	308	Space	3	6	282.41		S	8.69		210.79	7	5	1054.0	Office	50	105.4	5	0.06	663.3	13701	0.00	0.00	2764.00	81.66
103	335	Space	3	4	298.29		S	63.83		20.36245353	13	5	101.8	Office	50	10.2	5	0.06	64.1	1324	2017.00	876.00	1463.00	81.60
104	309	Space	3	4	134.11		W	100.02		14.78624535	7	5	73.9	Cooridor	5	0.7	5	0.06	13.3	96	106.00	80.40	1185.00	340.98
105	315	Space	3	6	145.20		N	99.30		157.581	13	5	787.9	Cooridor	50	78.8	5	0.06	495.9	10243	146.00	2315.00	2461.00	49.67
106	310	Space	3	8	142.30		N	25.60		110.91	13	5	554.6	Office	5	5.5	5	0.06	99.5	721	92.20	53.00	850.00	82.17
107	321	Space	3	4	208.90		W	25.60		152.337	10	5	761.7	Office	5	7.6	5	0.06	136.6	990	0.00	9.47	151.27	49.48
108	316	Space	3	4	209.05		E	44.77		211.94	13	5	1059.7	Office	5	10.6	5	0.06	190.1	1378	1492.72	53.96	335.17	99.85
109	332	Space	3	4	298.29		E	61.86		259.895	13	5	1299.5	Office	5	13.0	5	0.06	233.1	1689	769.02	351.32	443.48	57.21
110	317	Space	3	4	134.11		S, E	8.90		47.933	7	5	239.7	Office	5	2.4	5	0.06	43.0	312	0.00	0.00	363.00	676.58
111	322	Space	3	4	142.60		N, E	11.20		50.23	13	5	251.2	Cooridor	5	2.5	5	0.06	45.0	326	184.49	112.84	1728.54	168.84
112	318	Space	3	4	142.30		S, W	63.83		69.56	10	5	347.8	Office	0	0.0	5	0.06	45.0	0	2573.18	1402.99	12649.86	92.95
113	342	Space	3	3	208.90		N, E	98.60		120.36	10	5	601.8	Meeting Space	30	36.1	5	0.06	258.4	4694	92.20	46.60	713.00	93.30
114	319	Space	3	5	208.50		S, W	89.50		132.568	10	5	662.8	Office	0	0.0	5	0.06	85.7	0	600.00	316.00	931.00	190.30
115	323	Space	3	4	134.11		W	26.30		157.581	13	5	787.9	Storage	5	7.9	5	0.06	141.3	1024	2075.00	1521.00	7459.00	80.60
116	320	Space	3	3	125.60		W	44.77		210.79	7	5	1054.0	Storage	0	0.0	5	0.06	136.3	0	10608.00	4644.00	12941.00	81.66
117	333	Wash Space	3	4	281.40		E	61.86		20.36245353	7	5	101.8	Wash Space	0	0.0	5	0.06	13.2	0	2864.49	1329.00	14067.47	57.11
118	324	Wash Space	3	5	265.90		S, W	8.90		14.78624535	10	5	73.9	Wash Space	5	0.7	5	0.06	13.3	96	16897.00	4767.00	42254.00	130.50
119	336	Space	3	8	209.05		N, E	8.69		157.581	10	5	787.9		0	0.0	5	0.06	101.9	0	277.00	1569.00	22307.00	293.64
120	331	Space	3	4	298.29		S, W	63.83		110.91	7	5	554.6	Office	0	0.0	5	0.06	71.7	0	1348.00	317.00	2144.11	49.58
121	348	Space	3	4	134.11		S, W	100.02		152.337	10	5	761.7	Office	5	7.6	5	0.06	136.6	990	2996.00	1396.00	9138.00	88.83
122	326	Space	3	2 + 1 (outdoor)	142.60	89.02	S, W	89.50	26.51	211.94	7	5	1059.7	Office	0	0.0	5	0.06	137.1	0	2864.00	1634.00	14092.00	93.33
123	337	Space	3	4 + 1(outdoor)	142.30	71.5	N, W	25.60	26.51	259.895	13	5	1299.5	Storage	50	129.9	5	0.06	817.8	16893	3980.00	2986.00	12365.00	54.36
124	327	Space	3	5	208.90		E	25.60		47.933	13	5	239.7	Cooridor	5	2.4	5	0.06	43.0	312	830.00	190.00	2770.00	534.05
125	344	Exit Stair	3	4	208.50		S, E	44.77		50.23	10	5	251.2	Stairwell	5	2.5	5	0.06	45.0	326	276.74	142.01	2067.46	293.94
126	328	Space	3	3	134.11		N, E	56.30		69.56	12	5	347.8	Office	5	3.5	5	0.06	62.4	452	1660.00	562.00	8216.00	289.84
127	338	Space	3	5	125.60		S, W	61.86		120.36	13	5	601.8	Office	0	0.0	5	0.06	77.8	0	2306.20	774.60	11329.00	554.42
128	329	Space	3	5	281.40		N, E	8.90		132.568	6	5	662.8	Storage	5	6.6	5	0.06	118.9	862	92.20	87.60	1342.00	49.44
129	300	Space	3	3	265.90		S, W	12.20		157.581	13	5	787.9	Cooridor	50	78.8	5	0.06	495.9	10243	553.00	188.00	2745.00	49.48
130	330	Space	3	3	209.05		W	63.83		210.79	10	5	1054.0	Office	5	10.5	5	0.06	189.0	1370	92.20	51.40	787.00	292.05

131	339	Space	**3**	4	298.29		E	98.50		20.36245353	10	5	101.8			5	1.0	5	0.06	18.3	132	92.20	48.70	745.00	130.70
132	345	Space	**3**	7	134.11		N, E	87.60		14.78624535	10	5	73.9		Office	5	0.7	5	0.06	13.3	96	92.20	51.30	785.00	80.22
133	304	Space	**3**	4	142.60		S, W	26.30		157.581	13	5	787.9		Office	5	7.9	5	0.06	141.3	1024	92.20	48.00	735.00	81.32

APPENDIX E: FEATURES

Category	**Features**
Building Properties	# of Floors
	Year
	Occupancy Percentage
	Location
Geometry	Outer Wall Area
	Inner Wall Ara
	Outer Window Area
	Floor Area
	Roof Area
	Space Orientation
Isolation (Material)	Wall R-Value
	Window U-Value
	Roof R-value
Mechanical	Total Fresh Air
Electrical	Lighting Load

	Plug Load
Topology	Common Area
	Common R-Value
	# of Neighbour (Space Degree) Space Degree