TOWARDS THE APPLICATION OF A MACHINE LEARNING FRAMEWORK FOR BUILDING LIFE CYCLE ENERGY ASSESSMENT

A Dissertation

by

VARUSHA VENKATRAJ

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Manish Kumar Dixit
Co-Chair of Committee,	Wei Yan
Committee Members,	Stephen Caffey
	Petros Sideris
	Ashrant Aryal
Head of Department,	Phil Lewis

May 2022

Major Subject: Construction Management

ABSTRACT*

Nearly half of the global annual energy supply is consumed by buildings in their construction, operation, and maintenance, indicating an enormous potential to minimize the carbon footprint. During its life cycle, a building consumes energy in the form of embodied and operational energy. Embodied energy (EE) is expended in processes during construction, this includes extraction of raw material, transportation, manufacturing, etc. Operating energy (OE) is spent on operating and maintaining the building to ensure occupant comfort. Studies show that improving the operational efficiency of a building may have serious implications for EE. Building life cycle energy assessments (LCEA) is, therefore, essential to understanding the dichotomy between EE and OE.

Traditionally, data-driven approaches such as simulation-based optimization techniques are used for design space exploration. Literature shows that these data-driven approaches are error-prone, time-consuming, and computationally expensive, and fail to provide real-time feedback to the user. Besides, EE and OE assessment tools are disjointed

---

and suffer from the issues of interoperability. These limitations restrict design space exploration, which eventually hinders the design decision-making process.

In recent years, increased availability and accessibility of large-scale data have made machine learning (ML) techniques a popular choice for building performance assessment. In this context, numerous articles have developed prediction models to assess or optimize OE. While this work is significant, there remains a lack of studies that have utilized ML techniques for building LCEA mainly due to the lack of a large-scale LCEA database. This study proposed to generate a simulation-based building energy dataset for different building typologies using a parametric framework. The synthetically generated database was then used in the development of the ML model. The artificial neural network (ANN) model developed in this research would provide quick and reliable results related to the buildings' EE intensity and OE intensity. Furthermore, the application of the ANN prediction model was demonstrated using a case study. The experimental results of the case study show that the developed prediction model achieved high prediction performance using minimal inputs that are available during the early design phase. The results of this research indicate that ML techniques can indeed be used to instantaneously estimate building LCE performance. The practical implementation of this research would help designers with no experience in using simulation tools select design options with minimal LCE consumption.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my committee chair, Dr. Dixit for motivating me to pursue a Ph.D. while I was enrolled in the master's program. Dr. Dixit has extensive knowledge in the field of climate change, embodied energy, and building life cycle energy. His continuous support and guidance during the five years of my Ph.D. inspired me to conduct research and maintain a good publication record. During my time at A&M, he provided me with financial assistance and resources needed to complete my Ph.D. research for which I am forever grateful. I would also like to express my sincere thanks to my committee co-chair, Dr. Yan. He has dedicated valuable time to give me constructive feedback, guidance, and encouragement throughout these years. I appreciate Dr. Yan for patiently going through my programming code and helping me resolve certain issues. His recommendations helped me develop a more holistic research plan. My heartfelt thanks to Dr. Caffey for being extremely understanding and enthusiastic. His advice and discussions helped me broaden my perspective. Last but not the least, I would like to thank Dr. Sideris, and Dr. Aryal, for their guidance and support throughout the course of this research. I would also like to thank Dr. Lewis, Liz Smith, and Ginger White for their encouragement and motivation during my time here at Texas A&M University.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience. Finally, thanks to my brother, mother, and father for their encouragement and to my husband for his patience and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis (or) dissertation committee consisting of Professors Manish Kumar Dixit (Chair) and Ashrant Aryal of the Department of Construction Science, Professors Wei Yan (Co-Chair), and Stephan Caffey of the Department of Architecture, and Professor Petros Sideris of the Zachry Department of Civil and Environmental Engineering.

The data analyzed for Chapter IV was provided by Professor Manish Kumar Dixit. The analyses depicted in Chapter IV were conducted in part by Professor Manish Kumar Dixit of the Department of Construction Science and were published in 2021. All other work conducted for the dissertation was completed by the student with guidance from Professor Manish Kumar Dixit.

Funding Sources

This material is based upon work partly supported by the National Science Foundation under Grant No. 2048093. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION[*]

1.1 Background and motivation

The construction sector consumes nearly 50% of the global energy supply each year and accounts for 40% of the greenhouse gas (GHG) emissions within the United States (US) (EIA, 2019). Buildings consume several fossil fuel-based energy sources during their life cycle, and this demand keeps on increasing every year (Dixit, 2017). This rising energy demand has led to several adverse environmental implications (Huang et al., 2021). Few of these implications include energy scarcity, depletion of fossil fuels, and greenhouse gas (GHG) emissions leading to climate change and global warming (Abbasi and Noorzai, 2021; Fathi et al., 2020; Fumo 2014). The global energy crisis coupled with the uncertain climate future highlights the importance and urgency of reducing building

---

[*] Part of this chapter is reprinted with permission from:

"Challenges in implementing data-driven approaches for building life cycle energy assessment: A review" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 160, 112327, Copyright (2022) by Elsevier

"Life cycle embodied energy analysis of higher education buildings: A comparison between different LCI methodologies" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 144, 110957, Copyright (2021) by Elsevier

"Evaluating the impact of operating energy reduction measures on embodied energy" by Varusha Venkatraj, Manish Kumar Dixit, Wei Yan, and Sarel Lavy, Energy and Buildings, 226, 110340, Copyright (2020) by Elsevier

"Evaluating the temporal representativeness of embodied energy data: A case study of higher education buildings" by Pranav Pradeep Kumar, Varusha Venkatraj, and Manish Kumar Dixit, Energy and Buildings, 254, 111596, Copyright (2022) by Elsevier

energy and carbon footprints (Abediniangerabi et al., 2021). Building practitioners use the process of life cycle energy assessments (LCEA) to evaluate and optimize a building's energy performance and environmental impacts (Zeng and Chini, 2017; Ramesh et al., 2010).

The total life cycle energy (LCE) of a building consists of embodied and operating energy components (Seyrfar et al., 2021). Embodied energy (EE) is used indirectly through construction materials, equipment, and products and directly by several processes associated with manufacturing, transportation, construction, maintenance, repair, and final demolition (Chastas et al., 2016; Cabeza et al., 2014). Operating energy (OE) is consumed during the operational phase in the form of lighting loads, heating, and cooling loads, plug loads, and occupancy loads (Asl et al., 2015). The relative proportions of EE and OE in total building LCE is an ongoing debate within the research community (Venkatraj et al., 2020). However, several studies have come to a consensus that improving the operational efficiency of a building is often associated with an increase in EE (Venkatraj et al., 2020; Dixit, 2017). For example, adding more insulation to the exterior wall to decrease OE, may tremendously impact EE. These trade-offs become more pronounced in energy-efficient buildings, carbon-neutral buildings, and net-zero energy buildings, thereby, indicating the importance of evaluating the interdependencies between EE and OE components for overall building LCE reduction.

The dichotomy between the two LCE components has created a profound interest in the field of LCEA across the world. Mandatory laws and guidelines have been passed in several states across the US to reduce carbon emissions and embodied impacts

(Thilakarathna et al., 2020). For example, the 'Buy Clean California Act' which was implemented in 2020 requires that building materials used in all public projects within California should undergo a complete LCEA (CLF, 2019). Hence, several building professionals, researchers, and policymakers have focused their attention on emerging trends and strategies to improve the overall LCE performance (Kiss and Szalay, 2020). In this context, numerous tools and scientific methods have been developed using cutting-edge technologies for reliable building energy performance assessment (Wang et al., 2021; Fumo, 2014). These initiatives assist researchers as well as designers with the decision-making process, subsequently, resulting in the construction of high-performance buildings (Asl et al., 2017).

1.2 Problem statement

For the last few decades, governments all over the world have been working to improve energy performance and reduce carbon emissions produced by the built environment. The use of energy-intensive material to reduce OE creates a paradox between the energy spent to construct vs operate the building (Ramesh et al., 2010; Hernandez and Kenny, 2011). Therefore, evaluating both EE and OE is important for overall LCE reduction (Stephan and Stephan, 2016). Several tools and technologies exist to evaluate the OE and EE of a building (Basbagill et al., 2013). Among commonly used OE tools include e-Quest and Energy Plus, whereas Athena Impact Estimator, SimaPro, Tally, and One Click LCA are some of the LCA tools used for EE calculation (Aygenc, 2019; Attia et al., 2012). Despite the availability of these tools, design professionals still

face numerous challenges in incorporating LCA-based design decisions mainly due to a lack of expertise or resources (Asl et al., 2017).

Literature shows that existing building energy simulation tools are not very well integrated into the design platform and do not provide immediate feedback to the designer (Venkatraj et al., 2020; Asl et al., 2017; Feng et al., 2019; Kim et al., 2017; Attia et al., 2012). A simple energy simulation run generates large volumes of information, making it difficult for designers to compare and correlate different combinations of energy-saving factors (Amasyali and El-Gohary, 2018). They often must switch back and forth between OE and EE tools to understand trade-offs and make informed decisions (Shadram and Mukkavaara, 2018). Tools that simulate OE require information such as the material thermal properties, HVAC specifications, occupancy schedules, etc., whereas EE tools require material quantities, construction specifications, etc., This information is usually unavailable during the early design phase (Ngo, 2019). Moreover, the limited availability of EE data and unstandardized EE calculation methods hinder the evaluation of EE-OE trade-offs.

These issues with existing simulation methodologies and the design-decision-making process impede a systemic reduction in a building's life cycle energy use. To address these issues, several studies have suggested the use of data-driven methods to improve building energy performance assessment (Sharif and Hammad, 2019). In recent years, increased accessibility to large-scale data has made data-driven approaches more popular, mainly due to the ease of use. These approaches have emerged as an alternative to physics-based engineering approaches (Seyrfar et al., 2021; Ngo, 2019). Several studies

showcasing the application of building energy consumption prediction models have been published in recent years (Ahmad et al., 2014; Amasyali and El-Gohary, 2018; Zhao and Magoules, 2012; Sun et al., 2020; Bordeau et al., 2019; Wei et al., 2018; Yildiz et al., 2017; Mohandes et al., 2019; Deb et al., 2017; Fumo, 2014; Zhang et al., 2021; Wang et al., 2021; Qiao et al., 2021). While this work is significant, there remains a lack of studies that focus on data-driven approaches from a building LCE perspective (Mohandes et al., 2019). Therefore, the application of machine learning techniques, specifically for evaluating EE and OE is still an open research problem. As a result, designers explore a limited set of design options and fail to assess building energy from a life cycle perspective.

1.3 Thesis organization

This thesis is organized into nine chapters as seen in Figure I-1:

Chapter II presents a rigorous review of literature that provides a basic understanding of building LCE, EE, OE, calculation methods, trade-offs between embodied and operating energy components, building LCE performance assessment, available simulation tools, and machine learning-based approaches for building load prediction. This chapter also elaborates upon the current state of research: gaps and challenges some of which will be addressed by this research.

Chapter III outlines the overall research scope, assumptions, and limitations. It provides a holistic overview of the study design and methods that were implemented in this study.

Chapter IV highlights the variation in EE values caused by using different sources of data and LCI techniques (process-based, IOH-aggregated, and IOH-disaggregated). We quantify EE-OE trade-offs for different LCI calculation methods using two case study buildings. Ultimately, we compare and discuss how such trade-offs differ across the two methods. The results signify the importance of understanding the advantages and drawbacks of each LCI technique before evaluating EE-OE tradeoffs. In other words, a particular assembly may show different embodied impacts based on which EE inventory is applied to the calculation. In conclusion, the IOH-disaggregated approach is considered the most complete since it covers a larger system boundary. Therefore, justifying our intention to utilize the IOH-disaggregated approach in our parametric framework to generate synthetic data.

Chapter V outlines the process of developing a parametric framework to generate a simulation-based building LCE database. The data generated using this framework will be used to develop the machine learning model.

Chapter VI describes the development of the machine learning model for building LCE performance assessment. This chapter has two main parts. The first part focuses on developing the supervised ANN model for multi-output regression (EE and OE) using the data generated in the previous chapter. This includes (i) data cleaning, data pre-processing, and data transformation; (ii) determining the train-test split ratio, optimal network architecture, hyper-parameters, and performance metrics used to evaluate the model. We also utilize early stopping as a regularization technique to prevent overfitting the model.

The second part demonstrates the application and validation of the developed ANN model using a case study building.

Chapter VII summarizes the results and discusses the findings of this study. The findings of this study highlight the factors that influence the performance of the machine learning model. More importantly, the results of the case study indicate that machine learning models can indeed accurately predict building EE intensity and OE intensity using minimal information.

Chapter VIII provides a discussion regarding the various challenges we faced while conducting this study. Furthermore, we also elaborate upon the shortcomings and limitations of this model.

Chapter IX summarizes the research contributions and findings presented in this dissertation and provides recommendations for future work.

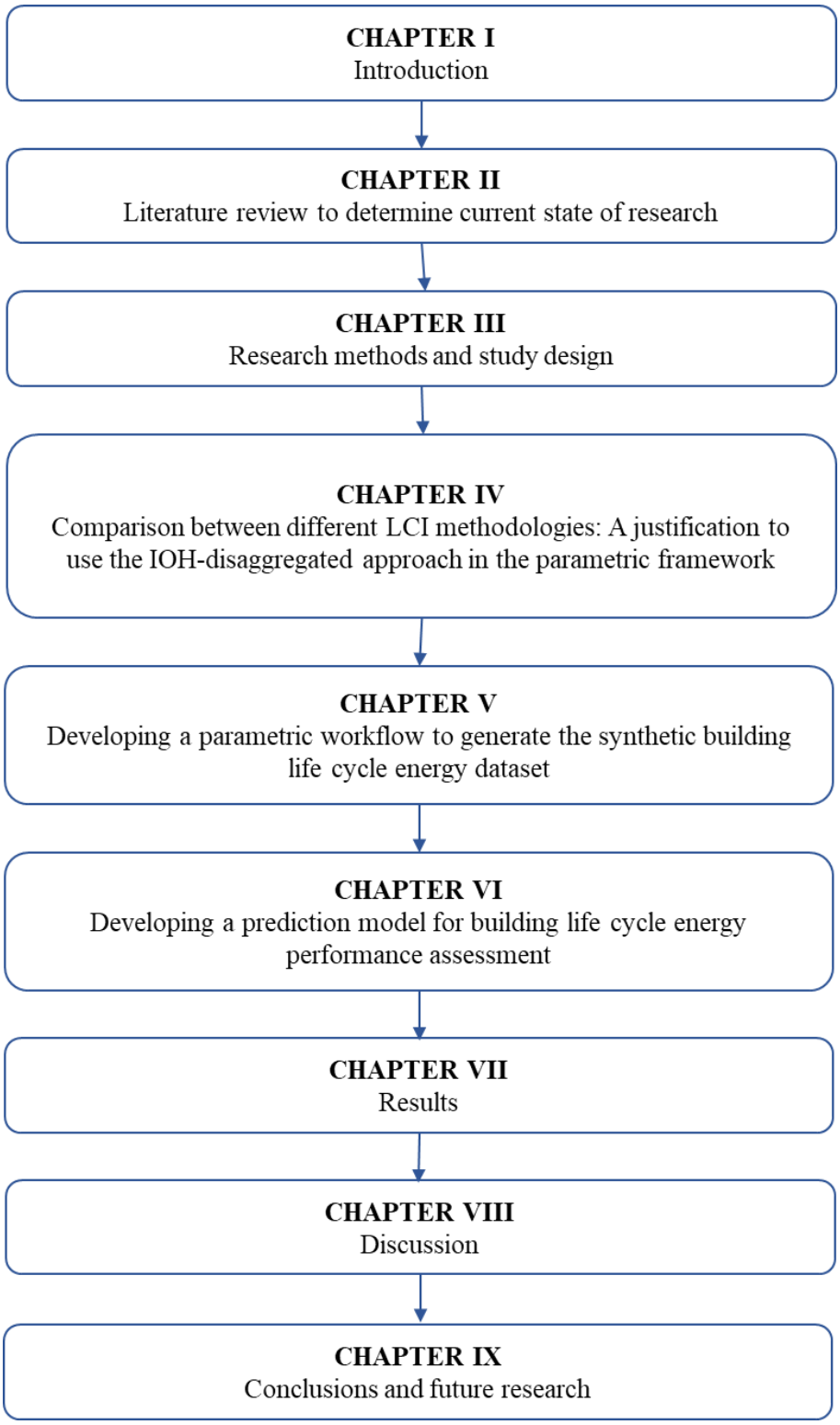


Figure I-1 Thesis organization

CHAPTER II

LITERATURE REVIEW*

2.1 Literature review method

We performed a comprehensive keyword-based search using the Google Scholar search engine to source relevant articles, peer-reviewed journal papers, book chapters, conference papers, dissertations, and technical reports. Few keywords that were used in the search are: building life cycle energy (LCE), embodied energy (EE), operating energy (OE), primary energy consumption, building energy prediction, building energy forecasting, parametric energy modeling, machine learning (ML), data-driven energy modeling, artificial neural network (ANN), support vector machine (SVM), surrogate modeling, genetic algorithm (GA), multi-objective optimization (MOO). Furthermore, these keywords were used in combination (ex: ANN + building energy) to improve the

---

* Part of this chapter is reprinted with permission from:

"Challenges in implementing data-driven approaches for building life cycle energy assessment: A review" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 160, 112327, Copyright (2022) by Elsevier

"Life cycle embodied energy analysis of higher education buildings: A comparison between different LCI methodologies" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 144, 110957, Copyright (2021) by Elsevier

"Evaluating the impact of operating energy reduction measures on embodied energy" by Varusha Venkatraj, Manish Kumar Dixit, Wei Yan, and Sarel Lavy, Energy and Buildings, 226, 110340, Copyright (2020) by Elsevier

"Evaluating the temporal representativeness of embodied energy data: A case study of higher education buildings" by Pranav Pradeep Kumar, Varusha Venkatraj, and Manish Kumar Dixit, Energy and Buildings, 254, 111596, Copyright (2022) by Elsevier

search and coverage of articles. Eventually, this search retrieved more than 850 studies from 1995 to 2021. We then performed a manual screening of the abstracts to eliminate studies that were irrelevant to the goals and objectives of this study. The following criteria were used to include relevant studies: (i) the study must address building energy consumption and its components; (ii) the research method should either focus on engineering or data-driven approaches for building energy performance assessment. Following this, we also included studies that were cited in the studies we screened first. These new articles were also screened based on the two inclusion criteria mentioned above. We then reviewed these articles based on the input and output targets, type and size of the dataset, ML model, temporal granularities, and prediction performance metrics used in their study. Furthermore, we analyze and discuss the factors that restrict the practical implementation of ML models for building LCEA research.

2.2 Building life cycle energy

The amount of energy spent during a building's lifetime consists of EE and OE as seen in Figure II-1 (Cabeza et al., 2014; Karimpour et al., 2014). Conventionally, studies use a life cycle energy analysis (LCEA) approach to measure environmental impacts and estimate net energy savings over the building's lifetime (Ramesh et al., 2010; Fay et al., 2000). Previous studies have consistently shown that LCEA results vary based on a multitude of factors such as the system boundary definitions, life cycle inventory (LCI) method, and type and form of energy included in EE calculations (Doh and Panuwatwanich, 2014; Dixit et al., 2013). A system boundary defines processes and

energy flows of a product's life cycle that are included in LCEA (Dixit et al., 2013). Frequently used system boundary definitions used in LCEA include ‘cradle to gate’, ‘cradle to site’, and ‘cradle to grave’ (Dixit et al., 2013). The cradle-to-gate assessment for a product includes all processes from raw material extraction and main manufacturing through the final product leaving the factory gate. The cradle-to-site system boundary covers additional activities such as transporting the final product from the factory gate to the construction site, on-site fabrication, administration, disposal of waste, etc. The cradle to grave system boundary includes operation, maintenance, renovation, retrofit, demolition, and other end-of-life activities, along with the cradle to site processes (Zuo et al., 2017; Cabeza et al., 2014). A system boundary definition also varies based on direct and indirect embodied energy components that are covered in LCEA (Dixit et al., 2010; Shrivastava and Chini, 2012). Direct energy refers to the energy consumed by the main processes, such as on-site and off-site construction activities, equipment, and material transportation (Stephan and Stephan, 2016). Indirect energy refers to non-energy inputs such as building materials, assemblies, packaging, equipment, etc., that are installed in a building (Praseeda et al., 2016).

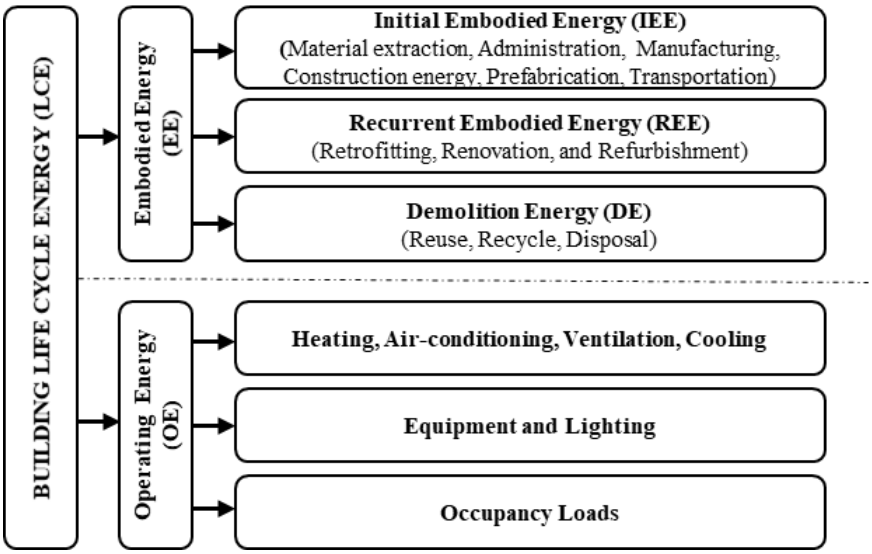


Figure II-1 Components of building life cycle energy (Reprinted with permission from Venkatraj and Dixit, 2022)

2.2.1 Embodied energy

Embodied energy is the sum of all energy embedded in products and processes used in the construction, maintenance, replacement, and disposal of a building (Copiello et al., 2016; Shrivatsava and Chini, 2012; Dixit, 2017). LCEE comprises of three primary components: initial embodied energy (IEE), recurrent embodied energy (REE), and demolition energy (DE) (Thomas et al., 2016). These EE components are spread over the three major building life cycle stages namely initial construction, operation and maintenance, and end of life/demolition (Stephan and Stephan, 2016; Thomas et al., 2016; Vukotic et al., 2010). IEE refers to the energy utilized directly and indirectly in a building's design and construction processes (Copiello et al., 2016). REE pertains to the

direct and indirect energy spent on maintenance, repair, replacement, and renovation activities, whereas DE is associated with the end-of-life activities such as deconstructing a building, recycling/reusing its building systems, or disposing them (Stephan and Stephan, 2016; Hernandaz and Kenny, 2011).

2.2.1.1 Calculation methods and issues

Embodied energy calculation methods include process-based, input-output (IO) based, hybrid, and statistical methods (Chang et al., 2014; Rauf and Crawford, 2015; Optis and Wild, 2010; Guan et al., 2016).

Process-based approach: In a process-based approach, actual energy use data is collected from construction sites and manufacturers and summed to compute the total EE. In an IO-based method, monetary flows between energy and other industry sectors are converted into physical energy flows using energy tariffs. In comparison to other LCI methods, the process-based method is relatively more reliable since it uses actual energy use data from the manufacturer, whereas IO-based approaches are considered complete but unreliable since the EE is computed for an aggregated industry sector (Davies et al., 2014; Guan 2016; Stephan et al., 2013). The process-based method includes direct and indirect energy flows and material inputs of each upstream process (Stephan et al., 2013; Crawford 2004; Omar et al., 2014). Tracking the major and minor energy flows for the entire supply chain is difficult and time-consuming (Stephan et al., 2013). Due to the issue of data unavailability, the system boundary in the process-based approach is incomplete (Crawford 2004; Acquaye et al., 2017; Lenzen 2000). This causes 'truncation error'

leading to underestimation of EE values (Crawford et al., 2019; Majeau-Bettez et al., 2011; Lenzen, 2000).

Input-output based approach: The IO-based method overcomes the issue of incomplete data by using economic IO data of monetary transactions between the industry sectors (Crawford 2004; Stephan et al., 2013; Treloar 1998). These transactions are converted into energy flows by using the product price and energy intensity of its manufacturing sector (Horowitz and Planting, 2009). Although the IO-based method has a relatively complete system boundary, it still suffers from several methodological issues such as the assumptions of proportionality and homogeneity that make its results unreliable (Acquaye et al., 2017). The IO-based approach uses energy tariffs and product prices for calculations which may over/underestimate EE values; furthermore, the energy intensities are assigned for the entire sector rather than the product level, resulting in 'aggregation error'(Stephan et al., 2013; Dixit et al., 2015; Crawford 2004; Davies et al., 2014; Guan et al., 2016).

Hybrid approach: A hybrid method was developed to resolve the limitations of the process-based and IO-based methods (Menzies and Tsolaki, 2016; Krogmann 2008). This method utilizes process data until the stage where complete information is available and IO data beyond that. Based on the framework used for calculating EE, the hybrid method is either a process-based hybrid or an IO-based hybrid (Dixit et al., 2015). As the name suggests, the process-based hybrid method uses the process-based framework with IO data, while the IO-based hybrid uses the IO-based framework integrated with process data (Acquaye et al., 2017; Crawford 2004). Since the hybrid method uses reliable process data

and the wider system boundary of the IO-based method, the hybrid method is regarded as more accurate and complete (Acquaye et al., 2017; Crawford 2004.; Treloar 1998).

2.2.1.2 Tools and software used to evaluate embodied energy

LCA tools are used to assess the energy and environmental impacts of building products, services, and processes across various life cycle phases (i.e., from raw material extraction to disposal). In recent times, tools such as Embodied Carbon in Construction Calculator (EC3), Tally™ and One Click LCA have become a crucial resource for building professionals designing zero energy buildings. These tools are integrated with BIM tools. Tally™ utilizes process-based EE data from GaBi 8.5 database (Tally, 2020). The GaBi database has been used frequently by multiple studies to conduct the LCA of built facilities around the globe (Martínez-Rocamora et al., 2016). Tally™ reports cradle-to-cradle results of a building's energy use under A1-A5 (IEE), B1-B5 (REE), and C1-C4 (DE) life cycle stages as defined by EN15804:2012 and EN15978:2011 (Tally, 2020). The One Click LCA tool also utilizes an automated process to map material information in the BIM to its LCA database and generate a detailed environmental impact report. This tool reports environmental impacts using a cradle-to-grave (life cycle stages from A-D) system boundary that is in compliance with the EN15804 or ISO14025 standards (One Click LCA, 2021).

2.2.2 Operating energy

Operating energy (OE) refers to the energy spent on operating and maintaining the building. This includes heating, ventilation, air-conditioning (HVAC), lighting, and plug loads (Karimpour et al., 2014; Thormark, 2002).

2.2.2.1 Tools and software used to evaluate operating energy

Energy simulation tools are used to evaluate energy use, system level performance, and life cycle cost impacts of buildings. These tools use physics-based modeling to predict and analyze building energy consumption. OE simulation tools analyze 3-dimensional computer models to estimate sub-hourly, hourly, monthly, or annual building energy consumption for buildings of any size or complexity. These tools also require detailed information about the building geometry, orientation, building envelope characteristics, shading devices, glazing systems, building systems, occupancy, and climate parameters to run the simulations (Amasyali and El-Gohary, 2018). Weather data is obtained from local weather stations or the ESSAT-EM tool. The Building Energy Software Tools Directory has recognized over four hundred OE simulation tools (IBPSA, 2019). Some of the popular tools include EnergyPlus, E-Quest, Ecotect, TRNSYS, DOE-2, etc., (Fay et al., 2010). The EnergyPlus simulation tool is developed, maintained, and distributed by NREL. OpenStudio is a software development kit that supports energy modeling in EnergyPlus.

More recently energy analysis plug-ins such as Autodesk Insight 360, Integrated Environmental Solutions Virtual Environment (IES VE), and DesignBuilder have gained

more traction because of their integration with BIM tools. These plug-ins are used to conduct whole building energy analysis which includes heating, cooling, and daylighting. These plug-ins are extremely useful in eliminating issues related to data transfer between BIM and BPA tools. The Autodesk Insight 360 and DesignBuilder simulation tools utilize the EnergyPlus engine to perform simulations (Elnabawi, 2020) whereas IES VE utilizes the APACHE engine for its analysis (IES VE, 2021). These plug-ins help designers to compare and assess several design options in a quick and seamless manner. In addition, IES VE utilizes cloud based operational dashboards which is extremely useful for portfolio management and collaboration across multiple stakeholders (IES VE, 2021). Few simulation plug-ins such as Ladybug and Honeybee are integrated with parametric modeling tools. These plug-ins connect the building geometry developed in Rhino and the parametric functionality of Grasshopper with the EnergyPlus, Radiance, Daysim, and OpenStudio simulation engines for building energy and daylighting analysis (Roudsari and Park, 2013).

2.2.3 Trade-offs between embodied and operating energy components

Conventionally, LCEA studies focus on reducing the operating energy since it is considered to constitute a larger proportion of building LCE (Ramesh et al., 2010; Karimpour et al., 2014; Zeng and Chini (2017); Chastas et al., 2016; Goia, 2016). However, recent research showed that improving building design or system efficiency to reduce OE resulted in increasing the proportion of EE in the total LCE (USGBC 2008; Monteiro et al., 2016; Ghattas et al., 2013). This trend of higher EE proportion in building

LCE can be profoundly observed in carbon-neutral and energy-efficient buildings (Copiello 2016). The studies conducted by Sartori and Hestnes (2007), Chastas et al. (2016) and Ramesh et al. (2010) estimated that the EE proportion in building LCE increases from approximately 5-20% in a conventional building to approximately 74-100% in a net zero energy building. Despite the shift in energy consumption from OE to EE, it is important to note that the overall LCE of energy-efficient buildings are generally lower in comparison to conventional buildings. For example, Stephan et al. (2017) assessed the LCE impacts for 87 variations of an apartment complex that was created by changing the design features of the building envelope. Their study revealed that the 40% wall to window ratio, double glazed window system, and heavyweight walls had the least LCE. Even though these variations used energy-intensive material, the increased IEE was offset by the OE savings.

Several studies focus on increasing the level of insulation in the exterior walls to reduce operating energy (Blengini and Di Carlo, 2010; Stephan et al., 2013; Rodrigues and Freire, 2017; Yohanis and Nortan, 2002; Lu et al., 2015). In most of these studies, implementing higher insulation levels results in net LCE savings despite the increase in IEE. For instance, Utama and Gheewala (2009) evaluated the LCE implications of changing the exterior wall type of a high-rise residential building located in Indonesia. They found that the wall type with a higher level of insulation has more LCE benefits despite using additional material. Implementing operating energy measures such as increasing the level of insulation, changing the glazing system, installing energy-efficient equipment, etc., utilize energy-intensive material. Figure II-2 further illustrates the

variation in EE and OE components across building types that have varying level of energy-efficiency measures incorporated in their design. Beyond a certain extent, it is impractical to achieve lower OE by utilizing energy intensive materials or systems. The excessive use of such measures may increase the IEE and, in some cases, become counterproductive to the goal of reducing total building LCE (Ramesh et al., 2010). These trade-offs highlight the need to comprehensively evaluate both EE and OE.

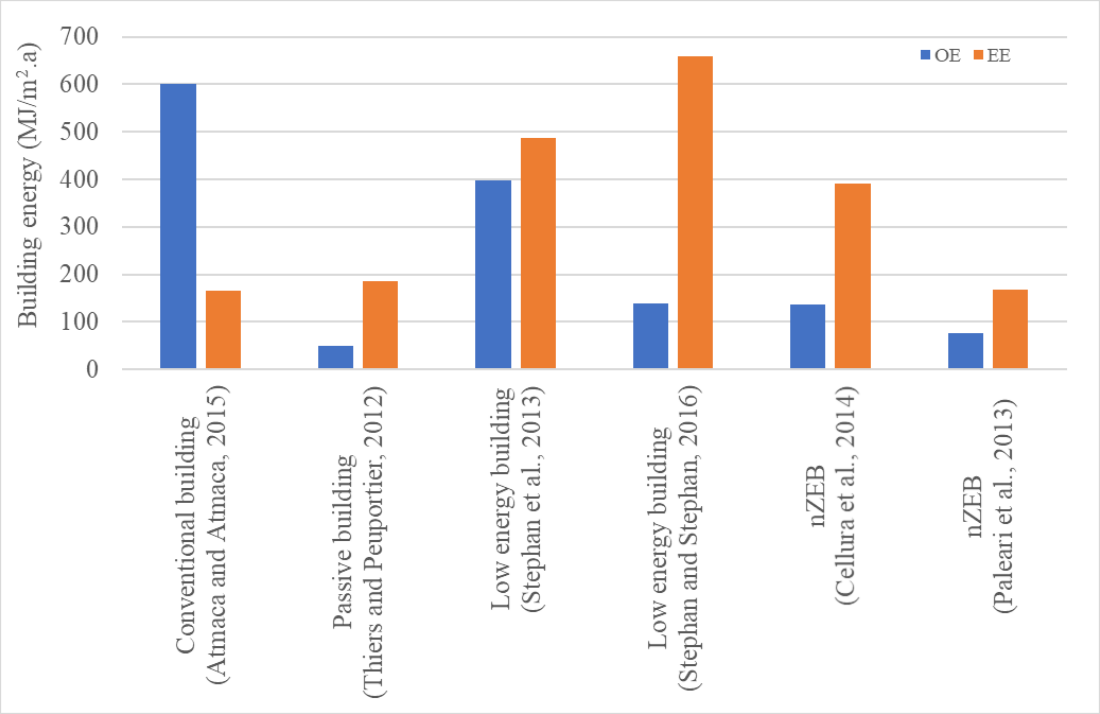


Figure II-2 EE vs OE trade-offs across building types

2.3 Building life cycle energy performance assessment

The methods used for building energy performance assessment can be segregated into two main categories, they are (i) Engineering (classical) approaches, and (ii) Data-driven approaches (Kiss and Szalay, 2020; Ngo 2019; Zhao and Magoules, 2012).

2.3.1 Engineering approaches

The engineering approach a.k.a. white-box approach utilizes complex mathematics, thermodynamic calculations, and simulation-based modeling techniques to evaluate the energy performance of a building (Fumo, 2014). Building performance simulation (BPS) tools in particular use 3D computer models and require detailed information about the building material quantities, construction, mechanical system, and climate parameters for their analysis (Amasyali and El-Gohary, 2018). Over the last decade, building information modeling (BIM) tools have become a popular choice for assessing the LCE performance of a building (Kavitha and Molykutty, 2021; Cavalliere et al., 2019; Chong et al., 2017). These BIM tools are integrated with BPS software such as EnergyPlus, Autodesk Green Building Studio, Tally, One Click LCA, etc., to conduct performance assessments (Abbasi and Noorzai, 2021; Wang et al., 2021).

Traditionally, BPS tools focus on evaluating a limited design space consisting of a few design options before finding the optimal design. To address these concerns, BPS tools are now integrated with parametric modeling tools (Rhino + Grasshopper) and optimization algorithms to further expand the design space and evaluate thousands of design options (Seyedzadeh et al., 2019; Touloupaki and Theodosiou, 2017). Optimization

algorithms are used to determine the minimum or maximum value of the objective function under a given set of constraints (Abbasi and Noorzai, 2021). Buildings are considered complex design problems with several conflicting design objectives; therefore, multi-objective optimization (MOO) approaches are used to quantify trade-offs between these different objectives (Shadram and Mukkavaara, 2018). Stochastic population-based algorithms such as genetic algorithm (GA), particle swarm optimization (PSO), and other hybrid algorithms are commonly used in the building industry (Nguyen et al., 2014). This approach generates several design alternatives based on the objective functions. Ultimately, the designer selects the most optimum building design from the Pareto-set of solutions (Wang et al., 2005).

2.3.1.1 Building life cycle energy optimization: key studies

Shadram and Mukkarva (2018) presented a MOO framework to explore the trade-offs between EE and OE of a low energy building in Sweden. Their study used a multi-objective genetic algorithm (MOGA) to show that small reductions in OE (140GJ) can result in large increases in EE (340GJ). Similarly, Wang et al. (2005) implemented a MOGA to show that a 5% increase in EE resulted in a similar operational energy decrease. Azari et al. (2016) used the GA to optimize the LCE performance of the building envelope for a low-rise office building located in Washington. The study evaluated combinations of six design features such as window type, wall to window ratio, insulation level, etc., of the building facade. Taghizade et al. (2019) applied a MOO approach to evaluate the environmental impacts of more than a hundred varieties of glazing systems. Hong et al.

(2019) developed a MOO model using NSGA-II to evaluate the thermal comfort for occupants, thermal energy loads, and life cycle economic and environmental impacts. Table II-1 summarizes MOO studies based on their system boundary and objective functions.

Table II-1 Summary of MOO studies for LCEA from literature (Reprinted with permission from Venkatraj and Dixit, 2022)

Study	Building type	Tools used	Optimization algorithm	System boundary	Objective functions
Sharif and Hammad, 2019		GA + energy simulation tool	GA	Envelope; HVAC; lighting	LCC, LCA
Fang and Cho, 2019	Commercial	Rhino + GH + LB + HB + Octopus	GA	Geometry; envelope; Weather	EUI
Kiss and Szalay, 2020	Apartment	Rhino + GH +LB + HB + Octopus + Ecoinvent database	GA	Geometry; Envelope; HVAC;	LCA indicators, OE
Shadram and Mukkavaara, 2018	Residential	BIM + Dynamo + GH + Archsim + Octopus	GA	Envelope; Service life; Occupancy; HVAC; Material quantities	EE, OE
Najjar et al., 2019	Residential	BIM + Tally + Green Building Studio	Mathematical optimization	Envelope; Occupancy; HVAC; Weather	OE, fuel, and electricity cost, constructability
Hollberg and Ruth, 2016	Residential	Rhino + GH + GOAT	CRS2	Geometry; Envelope; Occupancy; Weather data	LCA indicators
Lobaccaro et al., 2018	Residential	Rhino + GH + LB + Octopus + Galapagos	GA	Geometry; Envelope; Weather	EE, OE
Gilles et al., 2017	Residential	TRNSYS + Matlab	NSGA-II	Envelope; Weather; Heating systems	Primary energy consumption, CO_2 emissions; LCC, Durability, Comfort
Abbasi and Noorzai, 2021	Apartment	Rhino + GH + LB + HB + Octopus	HypE GA	Envelope; energy supply system	EE, OE
Touloupaki and Theodosiou, 2017		Rhino + GH + LB + HB + Galapagos	GA	Geometry; Envelope; HVAC;	LCE

2.3.2 Data-Driven approaches

Several studies have suggested the use of surrogate or data-driven models to overcome the limitations of the traditional physics-based modeling techniques (Ngo, 2019; Seyedzadeh et al., 2019; Singaravel et al., 2018). Black box-approaches a.k.a. statistical regression modeling techniques predict building energy performance based on historic/simulated energy use patterns. These models are purely data-driven and do not necessarily require detailed building information for prediction (Sayedzadeh et al., 2018). That said, it is always better practice to feed data-driven models with detailed building information to potentially obtain more prediction accuracy. In recent times, machine learning (ML) models (a subcategory of statistical modeling) have attracted substantial attention across several industries (Asl et al., 2017; Asl et al., 2016). ML approaches require initial efforts to develop, train, and test computational algorithms. Once successfully tested, these algorithms offer more efficient and faster assessment compared to other traditional simulation-based approaches. ML models require very minimal domain knowledge and are computationally inexpensive after the training process is complete (Liu et al., 2020). This data-driven approach utilizes considerable amounts of historical or simulated data to predict output targets for unseen data samples (Amasyali and El-Gohary, 2018). The flexibility and efficiency of ML models have made this approach popular in comparison to the other approaches.

2.3.2.1 Machine learning: key terms and concepts

With the increase in volumes and varieties of data, affordable data storage, and improvements in computational processing power, machine learning has become a key technique for quickly analyzing data (mathworks.com). Machine learning (ML) is a term that is used to describe computational algorithms that learn from existing data points without being programmed explicitly (Seyedzadeh et al., 2018; Ongsulee, 2017; Samuel, 1959). These algorithms have evolved based on pattern recognition and computational learning theories (Ongsulee, 2017).

Two of the most popular ML methods include supervised and unsupervised learning (Seyedzadeh et al., 2018; Ongsulee, 2017). Supervised learning algorithms are trained using labeled datasets (input features and output labels are known). These algorithms are utilized in applications where correlated input and output values are available for forecasting (Ongsulee, 2017). On the other hand, unsupervised algorithms are trained using unlabeled datasets. In this approach, the algorithm needs to recognize hidden patterns or intrinsic structures in the input features to predict outputs (Seyedzadeh et al., 2018). These algorithms are used for cluster analysis such as gene sequence analysis, market research, etc. (mathworks.com). To further clarify, supervised learning algorithms require an input dataset (represented as X) and output dataset (represented as Y), while unsupervised learning extracts meaning and patterns from just the input dataset (X) (Fan et al., 2017). Both supervised and unsupervised learning methods are used to develop classification and regression-based predictive models (mathworks.com). Classification

tasks predict discrete class label outputs, whereas regression tasks focus on predicting the continuous quantity outputs.

Literature shows supervised learning methods are conventionally used in the field of building energy performance prediction. Developing a supervised ML model consists of four major steps, they are data collection, data pre-processing/transformation, model training, and model validation (Amasyali and El-Gohary, 2018; Chalal et al., 2016). The process of data collection begins with gathering historical (real), benchmark, or simulated data containing correlated input and output values. Historic or real data refers to data collected from weather stations, smart sensors, utility bills, etc. Benchmark data consists of publicly available datasets such as the ASHRAE's Great Building Energy Predictor Shootout (Bourdeau et al., 2019), whereas simulation data is gathered by running energy simulations of building models on software tools such as EnergyPlus, TRNSYS, Ecotect, etc. (Amasyali and El-Gohary, 2018). During the process of data collection, it is important to ensure that the size of the dataset is representative of the entire sample space (Amasyali and El-Gohary, 2018). The collected data is often in its raw form and cannot be used directly in the ML model. To prepare the dataset, missing, repeated, incomplete, irrelevant, and noisy parts of the data are removed. After that, certain data pre-processing techniques such as normalization or smoothing are applied to improve data quality. This dataset is then split into training, testing, and validation datasets based on the users' discretion (Table II-2) (Wahid and Kim, 2016). The ML algorithm is then trained using the training dataset to identify patterns and relationships between the input features and output targets (Seyedzadeh et al., 2018). The main task of the ML algorithm during the training phase is

to learn the mapping function (Y = f(X)) from the input features (X) to the output labels (Y). After approximating the mapping function, the ML algorithm can predict output labels for new input features.

Table II-2 shows that the input features used in ML models built for energy prediction can be classified into six broad categories, weather conditions, geometry, building envelope, occupancy schedules, time, and HVAC systems. Weather refers to parameters such as dry bulb temperature, humidity, wind speeds, solar radiation, heating degree days (HDDs), cooling degree days (CDDs), etc. Geometry parameters describe the building shape and form, they include orientation, length, width, and height information. Building envelope parameters include information regarding the wall to window ratio (WWR), thermal conductivity (U-value) of the walls, windows, and roofs, shading depth, glazing type, and material composition of the envelope walls. Time parameters are used to denote the type of day (weekend or weekday). Few studies that developed prediction models for LCEA included additional features such as environmental indicators, source of primary energy, mass, and cost of construction materials (Feng et al., 2019; D'Amico et al., 2019). Ultimately, the prediction accuracy of the model is determined using performance metrics such as the coefficient of variation (CV), coefficient of multiple determination (R^2), percentage error (% error), mean absolute percentage error (MAPE), and root mean square error (RMSE). Table II-2 summarizes each study in terms of building type, type of data used, ML algorithm, input features, dataset size, train-test split, output target, and performance metrics.

Generalization is defined as the ability of the ML model to handle unseen data. Commonly used generalization techniques applied to resolve overfitting include changing the model complexity (number of weights or values of the weights), early stopping (using cross-validation), dropout, noise, data augmentation, hyperparameter tuning, etc.

Few popular supervised ML algorithms include k-nearest neighbor (kNN), linear regression (LR), decision trees (DT), support vector machines (SVM), and artificial neural networks (ANN) (Runge and Zmeureanu, 2019; Amasyali and El-Gohary, 2018). SVM and ANN models are extensively used in the building sector to predict short-term, medium-term, and long-term energy loads, enhance building control and improve building energy system design (Wei et al., 2018). These models have proven to be far more superior than other comparable ML algorithms in terms of handling high-dimensional or non-linear data. Moreover, they also have higher prediction accuracy and generalizability in comparison to other regression-based methods (Asl et al., 2017). The review conducted by Amasyali and El-Gohary (2018) shows that 47% and 25% of the research articles trained their models using ANN and SVM, respectively. Therefore, our review predominantly focuses on research articles that deploy SVM or ANN models for building energy prediction.

Table II-2 Summary of studies focused on ML-based building energy prediction (Reprinted with permission from Venkatraj and Dixit, 2022)

Study	Building type	Data used	ML model	Input features	Size of the dataset	Output parameter	Performance metric
Li et al., 2019	Commercial	Simulated	ANN	Geometry, Envelope, HVAC	Total: 172,980 Training: 150,000 Test: 22,980	CL; HL; LL	R^2
Asl et al., 2017	Commercial	Simulated	ANN BDT	Geometry, Envelope, HVAC	Total: 180,000 Training: 67% Test: 33%	OE	R^2
Sharif and Hammad, 2019	Educational	Simulated	DNN	Envelope, HVAC, Occupancy, Lighting	Total: 463 Training: 325 (70%) Test:138 (30%)	OE; LCC; LCA	MSE
Fan et al., 2017	Educational	Historic	DNN	Weather, HVAC, Time	Total: 15,792 Training: 70% Test: 15% Validation: 15%	CL	MAE RMSE CV-RMSE
Feng et al., 2019	Commercial	Simulated, EP; ICE database	ANN FCM ELM	Geometry, Envelope, Weather, environmental indicators	Total: 1152 Training: 1024 Test: 128	Environmental uncertainty	MAE RMSE CV-RMSE
Khalil et al., 2019			DNN	Geometry, Envelope	Total: 768 Training: 519 Test: 249	HL; CL	
Kumar et al., 2018		Benchmark	ELM OSELM	Geometry, Envelope	Total: 768	HL; CL	MAE
Deng et al., 2018	Commercial	CBECS	LR SVM RF ANN	Geometry, Envelope, HVAC, Weather, Lighting, Occupancy	Total: 1024 Training: 50% Validation: 25% Testing: 25%	EUI	MAE RMSE

Continued

Study	Building type	Data used	ML model	Input features	Size of the dataset	Output parameter	Performance metric
Edwards et al., 2012	Residential	Historic	LR FFNN SVM LS-SVM FCM with FFNN	HVAC, Occupancy	Training: 85% Test: 15%	Electricity	CV MBE MAPE
Naji et al., 2016	Residential	Simulated EP	ELM ANN	Envelope	Total: 180	HL; CL	RMSE r R^2
Chou and Bui, 2014		Simulated Ecotect	SVR ANN LR	Geometry, Envelope	Total: 768	HL; CL	RMSE MAE MAPE
Tsanas and Xifara 2012	Residential	Simulated Ecotect	IRLS RF	Geometry, Envelope	Total: 768	HL; CL	MAE MRE MSE
Marino et al., 2016	Residential	Benchmark	LSTM	Electricity consumption	Total: 2075259	EUI	RMSE
Fan et al., 2019	Educational	Historic	MLR SVR ANN XGBDT	Weather data, HVAC, Time	Total: 17040 Training: 70% Testing: 30%	CL	RMSE MAE CV-RMSE
Somu et al., 2021	Educational		*k*-CNN-LSTM	Weather, HVAC	Total: 420 Training: 60% Test: 20% Validation: 20%	OE	MAE MAPE MSE RMSE
Amarasinghe et al., 2017	Residential	Benchmark	CNN	Electricity consumption	Total: 34608 Training: 75% Testing: 25%	OE	RMSE
Ahmad et al., 2017	Commercial	Historic	FFBP-ANN RF	Weather, Occupancy, Time	Total: 10972	OE	RMSE MAPD CV MAPE

Continued

Study	Building type	Data used	ML model	Input features	Size of the dataset	Output parameter	Performance metric
							MAD R^2
Lee et al., 2019	Residential	Historic; Simulated EP	ANN	Occupancy, Demography	Total: 5192	OE	MSE
Wang et al., 2018	Educational	Historic	RF RT SVR	Weather, Occupancy, Time	Total: 8760 Training: 80% Test 20%	Hourly-electric usage	RMSE MAPE R^2
Wang et al., 2020	Educational	Historic	RF XGBoost GBDT SVR kNN	Weather, Time	Training:70% Testing: 30%	HL	RMSE MAE MAPE CV-RMSE
Biswas et al., 2016	Residential	Historic	ANN	Weather, HVAC, electricity		OE; Heat pump consumption	SSE MSE R^2
Aydinalp et al., 2004	Residential	Benchmark Historic	ANN	Weather, HVAC	Total: 1228 (HL) Total: 563 (DHW) Training: 75% Test 25%	HL; DHW	R^2
Michalakakou et al., 2002	Residential	Historic	ANN	Weather data	Training: 80% Test: 20%	HL; CL	RE
Kaligirou et al., 1997	Educational	Historic	ANN	Envelope, HVAC	Total: 250	HL	R^2
Ben-Nakhi and Mahmood, 2004	Commercial	Simulated ESP-r	GRNN	Weather	Training: 80% Test: 20%	CL	R^2
Sebestyen and Tyc, 2020		Simulated LB	ANN	Envelope, Weather	Total: 10000 Total: 4500	Sunlight hours; Radiation	MSE

Continued

Study	Building type	Data used	ML model	Input features	Size of the dataset	Output parameter	Performance metric
Cheng-wen and Jian, 2010	Residential	Simulated DeST	ANN BPNN	Envelope, Weather	Total 132 Training: 75% Testing: 25%	HL; CL	MAE CV-RMSE
Robinson et al., 2017	Commercial	Benchmark; CBECS	LR SVM XGBoost	Envelope, Weather, HDD, CDD, occupancy	Total: 6720 Total: 13223	OE	MAE
Singaravel et al., 2018	Commercial	Simulated EP	CBML ANN LSTM	Envelope, Geometry, Orientation, Weather	Total: 9600	OE	R^2 CV
Ilbeigi et al., 2020	Commercial	Simulated LB, HB	ANN	Envelope, Weather data, Occupancy	Total: 1602 Training: 70% Testing: 15% Validation: 15%	EUI	MSE
Li et al., 2019	Commercial	Historic	ELM BPNN SVR GRBFNN MLR		Total: 17469 Total: 8734	OE	RMSE
Yan and Yao, 2010	Residential	Simulated EP	BPNN	Envelope, Weather		OE	
Ngo 2019	Commercial	Simulated TRACE 700	LR ANN SVR CART	Geometry, Envelope, Weather, Occupancy	Total: 243	CL	MAE R MAPE RMSE
Wong et al., 2010	Commercial	Simulated EP	MLP-ANN	Time, Envelope, Weather	Total: 11315 Training: 70% Testing: 30%	OE; HL; CL, LL; Electricity	CV-RMSE MBE RMSE
Turhan et al., 2014	Residential	Simulated KEP-IYTE-ESS	BPNN	Geometry, Envelope, Weather	Total: 148 Training: 103 Testing: 45	HL	R^2 MSE MAPE

Continued

Study	Building type	Data used	ML model	Input features	Size of the dataset	Output parameter	Performance metric
Sun and Han, 2013	Commercial	Simulated Ecotect	BPNN	Geometry, Envelope		HL	
Dong et al., 2018	Commercial	Simulated EP	BPNN	Geometry, Envelope	Total: 500 Training:70% Testing: 15% Validation: 15%	OE; Construction cost	MSE
Li et al., 2010	Commercial	Simulated DeST	LS-SVM	Weather		CL	RME RMSE MARE
Paudel et al., 2017	Residential	Simulated TRNSYS	SVM	Weather		HL	RMSE R^2
Zhang et al., 2017	Commercial	Simulated EP	SVM MLR	Weather, Time, Envelope	Total: 48 Training: 36 Test: 12	CL	MAE MAPE RMSE CV-RMSE
Dong et al., 2005	Commercial	Historic	SVM	Weather	Total: 5 years Training: 4 years Testing: 1 year	OE	MSE S-MSE CV-RMSE
Setiawan et al., 2009	Commercial	Historic	SVM BPNN	Time	Total: 26784 Training: 17856 Testing: 8928	Electricity	MAE RAE MAPE KPI
Zhao and Magoules, 2009	Commercial	Simulated EP	SVR	Weather, Envelope, Occupancy, HVAC	Total: 3624 Training: 3576 Testing: 48	OE	MSE SCC
Zhang et al., 2016	Commercial	Historic	SVR	HVAC, Time, Chilled water	Total: 480 Training: 384 (80%) Testing: 96 (20%) Total: 261	OE	MAPE

Continued

Study	Building type	Data used	ML model	Input features	Size of the dataset	Output parameter	Performance metric
					Training: 209 Testing: 52		
Li et al., 2009	Commercial	Simulated DeST	SVM	Weather	Total: 3696 Training: 2952 Testing: 744	CL	RMSE MRE
Seo et al., 2019	Commercial	Simulated EP	ANN	Weather	Training:Test = 9:1	CL	CV-RMSE
Chari and Christodoulou, 2017	Residential	Simulated	ANN	Weather, Envelope, Lighting, HVAC	Total: 100000; 90000; 70000; 40000 Training: 70% Test: 15% Validation: 15%	Primary energy consumption, CO_2 emissions	CE
Elbeltagi and Wefki, 2021	Residential	Simulated EP	BPNN	Envelope, Geometry, Occupancy, Lighting, HVAC	Total: 12000 Training: 10200 Validation: 1800	EUI	MAPE R^2
Cuilla et al., 2019	Commercial	Simulated TRNSYS	ANN	Weather, Envelope, Time	Total: 2184 Training: 85% Testing: 15%	OE	MAE MSE MAPE
D'Amico et al., 2019	Commercial	Simulated TRNSYS	ANN	Geometry, Location, Weather, Envelope; Energy source, Mass of construction materials	Total: 28,080 Training: 85% Testing: 15%	HL; LCA indicators	RMSE MAPE R^2
Martellotta et al., 2017	Apartments	Simulated EP	ANN	Weather, Lighting, Equipment	Total: 3288 Training: 85% Testing: 15%	HL	MSE R^2
Fan et al., 2014	Commercial	Historic	SVR MLR ARIMA	Weather, Time	Total: 34616 Training: 70% Testing: 30%	OE; Peak power demand	RMSE MAE MAPE

Continued

Study	Building type	Data used	ML model	Input features	Size of the dataset	Output parameter	Performance metric
			RF MLP BT MARS kNN				
Luo et al., 2020	Commercial	Simulated TRNSYS	ANN SVM LSTM	Weather, Occupancy	Total: 8760	HL; CL; LL; BIPV electrical power production	MAPE
Amber et al., 2018	Commercial	Historic	MR GP ANN DNN SVM	Weather, Time	Total: 1825 Training: 1460 Testing: 365	Electricity	MAPE
Ye et al., 2018	Commercial	Historic	BPNN	Weather Time	Total: 60 Training: 40 Testing: 20	Electricity	RMSE
Lei et al., 2021	Commercial	Historic	DNN BPNN FNN	Weather, Envelope, Occupancy	Total: 8176 Training: 7840 Testing: 336 Total: 341 Training: 311 Testing: 30	OE	MAPE RMSPE
Zhong et al., 2019	Residential	Historic	SVR	Weather	Total: 52 days Training: 70% Testing: 30%	CL	R MAE RMSE RAE RRSE
Seyrfar et al., 2021	Residential	Benchmark	BPNN RF XGBoost	Envelope, Demographic, Socio-economic, Energy use	Total: 1325	EUI	MAPE MSE R^2

Continued

2.3.2.2 Support Vector Machine (SVM): Theory and key studies

SVMs are considered the most robust, accurate, and computationally efficient data mining models in the research community (Chalal et al., 2016; Ahmad et al., 2014; Cortes and Vapnik, 1995). SVM models were developed based on the principles of statistic learning theory (SLT) and structural risk minimization (SRM) (Ahmad et al., 2014; Dong et al., 2005). These principles are known to have higher generalization performance in comparison to the traditional Empirical Risk Minimization (ERM) principle, used in neural networks. SVM models are, therefore, less prone to issues caused by generalization and overfitting (Chalal et al., 2016; Cortes and Vapnik, 1995). SVMs deployed for regression tasks are referred to as support vector regression (SVR) models (Yildiz et al., 2017).

SVMs utilize a kernel-based learning algorithm, which can solve both linear and non-linear problems (Wu et al., 2008). To solve non-linear problems, SVMs transform the non-linearity between the input features (x_i) and output target (y_i) using linear mapping in two steps. In the first step, the non-linear problem is projected onto a kernel-induced high dimensional space a.k.a. feature space. After that, the function f(x) that best fits the problem in the high dimensional space is determined. In the second step, the kernel function is applied to convert the complex non-linear map into a linear problem (Amasyali and El-Gohary, 2018). Some frequently used kernel functions include radial bias function (RBF) a.k.a. Gaussian kernel, linear function, and polynomial function (Kavaklioglu, 2011; Jain et al., 2014; Zhong et al., 2019). Selecting an appropriate kernel function is extremely important since it impacts the generalizability, learning ability, and prediction

accuracy of the SVM algorithm (Chalal et al., 2016; Wang and Srinivasan, 2015). According to the literature, the least-square SVM (LS-SVM) is a commonly used sub-type of SVM models.

Shao et al. (2020) established an SVM-based model that utilizes the RBF to predict the overall energy consumption of hotel buildings. Weather and operating parameters were measured on-site and used as inputs for the SVM model. The performance of the developed SVM model was evaluated using MSE and R^2 values which were 2.22% and 0.94, respectively. Similarly, Ma et al. (2018) developed an SVM model to forecast building energy consumption in China. Their model was trained using historic weather data and statistical economic indicators. The model deployed in their study had an MSE of 0.001% and an R^2 value higher than 0.99, therefore, indicating good accuracy. Dong et al. (2005) presented an SVM model to predict the monthly energy consumption of commercial buildings in Singapore. The model was trained and tested using weather data and utility bills. Their study used performance metrics such as CV and % error to evaluate the model. Other studies that have used SVMs to assess building energy consumption include Massana et al. (2016), Lai et al. (2008), Jain et al. (2014), and Solomon et al. (2011). Several studies also utilized least square support vector machines (LS-SVMs) – a subtype/reformulation of traditional SVMs. LS-SVMs are known to perform better than SVMs in terms of reducing the complexity and the training speed (Kaytez, 2020; Han et al., 2019). Li et al. (2010) developed an LS-SVM model to forecast the cooling loads of a commercial building in China. Similarly, Yi et al. (2010) used economic indicators as input features to build the LS-SVM regression model which predicts primary energy

consumption. Chen et al. (2021) developed a building energy consumption prediction model using building envelope parameters as input parameters. The LS-SVM model has an RMSE of 0.0273 and an R^2 of 0.9549.

2.3.3.3 Artificial Neural Network (ANN): Theory and key studies

ANNs are computational models designed to mimic the neural network of the human brain (Chalal et al., 2016). The theoretical framework of a generic ANN model consists of three sequential layers: the input layer, the hidden layer, and the output layer as illustrated in Figure II-3(a) (Ekici and Aksoy, 2009; Amasyali and El-Gohary, 2018). Each layer consists of interconnected nodes/neurons where computation occurs (Kumar et al., 2013). Labeled data is fed into the nodes of the input layer, the weighted sum of values from the input nodes are processed in the nodes of the hidden layer, and the result is propagated to the nodes of the output layer using an activation function (Mohandes et al., 2019; As et al., 2018). At first, the nodes in the hidden layer are initialized with random weights, these weights are then iteratively manipulated and updated during the training phase by comparing the model's output with the actual output (Runge and Zmeureanu, 2019). These weights either amplify or dampen the effect of a particular input feature based on the task the algorithm is trying to learn (Khalil et al., 2019). This iterative process of training continues until it is terminated by a certain criterion. The user may define the stopping criterion based on the maximum number of iterations (epochs), loss, or error (Sharif and Hammad, 2019). The model architecture described above is also referred to as a feed-forward neural network (FFNN) since information flows only in one direction

(Jovanovic et al., 2015). Back-propagation neural networks (BPNN) on the other hand pass information in a cyclic manner (Wang et al., 2014). Here, the error of the output is propagated as negative feedback to each node in the hidden layer. This feedback mechanism helps in adjusting the weights and biases of the hidden layer nodes which subsequently results in minimizing the total loss.

Deep learning is a sub-field within ML, which utilizes a 'deep' model architecture for prediction or classification (Fan et al., 2017). This means that the input features are transformed several times in a linear or non-linear fashion before obtaining the output (Fan et al., 2017). Conventionally, ML algorithms are considered 'shallow' since the input features undergo only one or two rounds of transformation as seen in Figure II-3(a) (Li et al., 2017). Since deep neural networks (DNN) contain more than one hidden layer they obtain great feature learning abilities and have better performance as illustrated in Figure II-3(b) (Qiao et al., 2021; Ongsulee, 2017). These algorithms are used to analyze data for more complex tasks such as image recognition, object classification (computer vision), speech recognition, etc. To summarize, (i) conventional ML algorithms mostly rely on structured labeled data, while deep learning networks rely on the layers of the ANN; (ii) conventional ML algorithms need human intervention when a result is incorrect, whereas deep learning algorithms learn from their own mistakes. The study by Singaravel et al., (2018) shows that the benefits of deep learning algorithms can only be observed in very large datasets, otherwise, their results are comparable to conventional ML algorithms. Ultimately, the quality of data plays the most important role and determines the accuracy of the output. Some other types of popular ANNs include radial basis function neural

networks (RBFNN) (Lee and Ko, 2009), extreme learning machine (ELM) (Roy et al., 2018), multi-layer perceptron (MLP) (Amasyali and El-Gohary, 2018), recurrent neural networks (RNN) (Mocanu et al., 2016; Bourdeau et al., 2019).

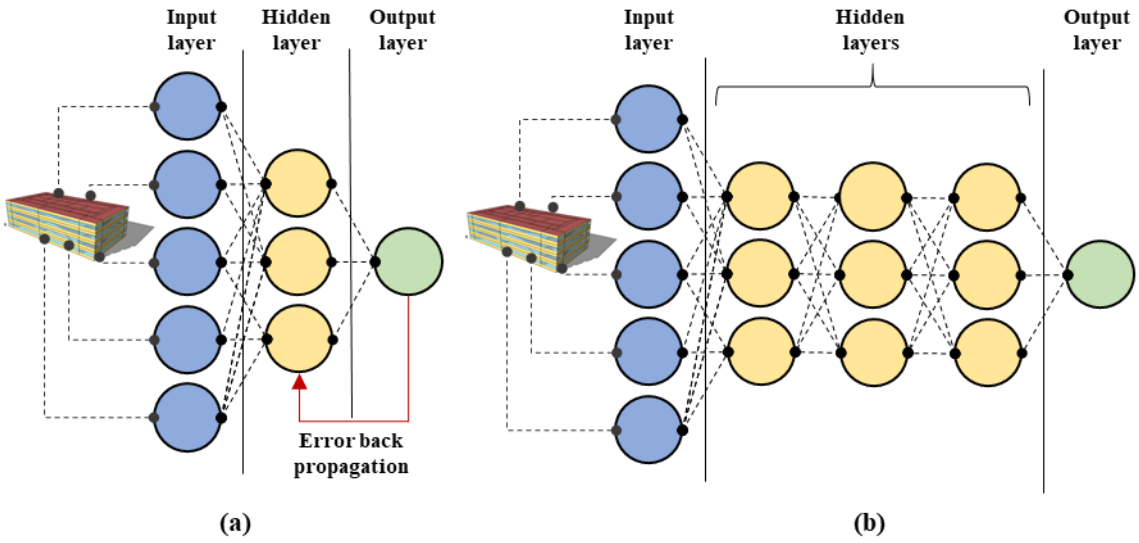


Figure II-3 Conceptual structure of (a) shallow and (b) deep neural network (Reprinted with permission from Venkatraj and Dixit, 2022)

Khalil et al., (2019) evaluated the effect of input features such as relative compactness, roof area, overall height, surface area, glazing area, wall area, and orientation on the output variable – the heating and cooling loads of the building. The training dataset used in their study included 768 residential buildings, and the accuracy of

prediction was found to be 99.60%. Similarly, Mena et al. (2014) trained an ANN model to forecast the overall energy demand of a non-residential building located in Spain. Ngo (2019) developed an ML model to predict building cooling loads. The input features included in their study consist of parameters such as aspect ratio, window to wall ratio, number of floors, outdoor air rate, floor area, and window glass U-value. The cooling loads of 243 commercial buildings located in Taiwan were used to train the ANN model. In this case, the MAPE between the observed and predicted value was 6.17%. Similarly, Yezioro et al. (2008) and Neto et al. (2008) developed an FFNN model to predict cooling loads using historic data. Kwok et al. (2011) and Paudel et al. (2014) developed an MLP model to estimate cooling and heating loads, respectively. The studies by Bagnasco et al. (2015) and Yokoyama et al. (2009) trained BPNN models to forecast the cooling demand of non-residential buildings. Wong et al. (2010) developed an ANN model to predict heating, cooling, artificial lighting, and total electric loads. The input features were related to the external weather conditions and building envelope design. Their study utilized the EnergyPlus platform to run simulations and generate the database for the ANN model. The developed ANN model was able to predict building loads with a percentage error ranging from 3 to 5.6%. Similarly, Yan and Yao (2010) also utilized a simulation-based dataset to develop a BPNN model to predict the overall energy consumption. Few other studies that have used ANNs to predict the heating and cooling loads in a building include Alam et al. (2016); Kumar et al. (2013); Sun and Han (2013); Azari et al. (2016).

2.3.3.4 Comparison-based studies and Ensemble models

Several studies compare the performance of different algorithms that were used to develop the ML prediction model. For example, Jain et al. (2016) compared MLR and SVM; Deng et al. (2018) compared LR, Lasso regression, SVM, RF, and ANN; Kumar et al. (2018) compared ELM and OSELM; Naji et al. (2016) compared ELM and ANN; Li et al. (2019) compared ELM, BPNN, SVR, GRBFNN, and MLR; Setiawan et al. (2009) compared SVM and BPNN; Lei et al. (2021) compared DNN, BPNN, and FNN; Amber et al. (2018) compared ANN, DNN, and SVM; Luo et al., 202 compared ANN, SVM, and LSTM; Fan et al. (2014) SVR, MLR, ARIMA, RF, MLP, BT, kNN; Zhang et al. (2017) compared SVM and MLR; and Massana et al. (2015) compared MLP, MLR, and SVM.

Since single data-driven models (as seen in Figure II-4(a)) may have certain limitations, a composite data mining approach called ensemble learning (Figure II-4(b)) has been introduced lately (Ovadia et al., 2019; Wang et al., 2017; Runge & Zmeureanu, 2019; Wang and Srinivasan, 2015). In essence, this approach combines multiple single prediction algorithms, to improve the performance, generalizability, and stability of the overall model (Qiao et al., 2021; Bourdeau et al., 2019). For instance, Chou and Bui (2014) constructed prediction models using 768 datapoints. Their study considered eight input features to predict the heating and cooling loads. They found that the ensemble model (SVM+ANN) had the least percentage error (less than 4%). Similarly, Alobaidi et al. (2018) developed an ensemble learning framework to forecast daily building energy consumption. They found that the MLP-FFNN-based ensemble model had better performance (MAPE=14.4%) and generalization ability in comparison to the single ANN

model (MAPE=18.3%). Jovanović et al. (2015) combined FFNN, RBFNN, and Adaptive neuro-fuzzy inference system (ANFIS) methods to create an ensemble model. The results show that the ensemble model predicted heating energy consumption with an R^2 and MAPE value of 0.9843 and 5.3%, respectively.

Singaravel et al. (2018) used a component-based machine learning model to improve generalization. Here, several ANN models were developed for specific tasks such as heating load estimation, cooling load estimation, etc. Then, high-level abstraction was implemented to integrate various ANN models to estimate total energy consumption. Studies also recommend using a Bayesian learning framework that introduces probability distribution over the neural network and a Gaussian likelihood function. This allows the use of posterior distributions to obtain predictions for out-of-sample predictions (Brusaferri et al., 2019). Transfer learning is another approach that can be used to leverage data learned in a previous task to be transferred to a similar/related task (Singaravel et al., 2018).

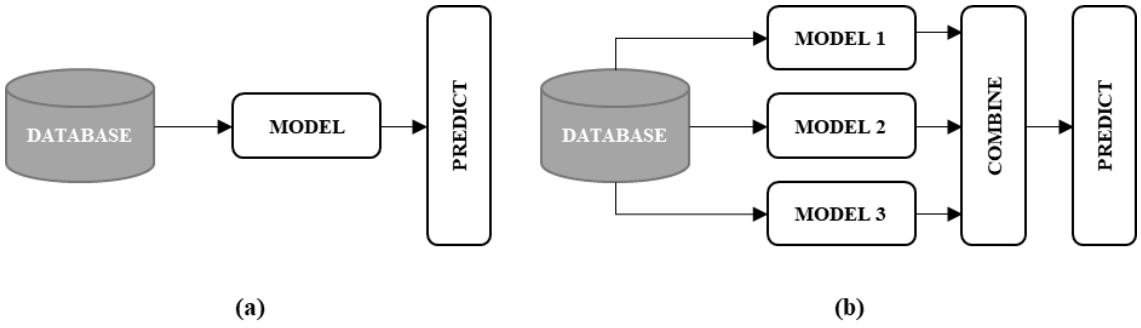


Figure II-4 Conceptual structure of (a) traditional ML model (b) ensemble ML model (Reprinted with permission from Venkatraj and Dixit, 2022)

2.4 Challenges in implementing data-driven approaches for LCE prediction

Our comprehensive review of past and recent literature shows that data-driven approaches for predicting building energy are generally focused on the operational building energy consumption. Based on our analysis of existing literature, we noticed that barely 7% of the articles have included EE or other environmental indicators in their study. In this section, we identified and examined gaps, challenges, and issues that exist within the realm of data-driven approaches (Figure II-5). Moreover, we also discuss the factors that hinder the implementation of ML-driven solutions for building LCEA.

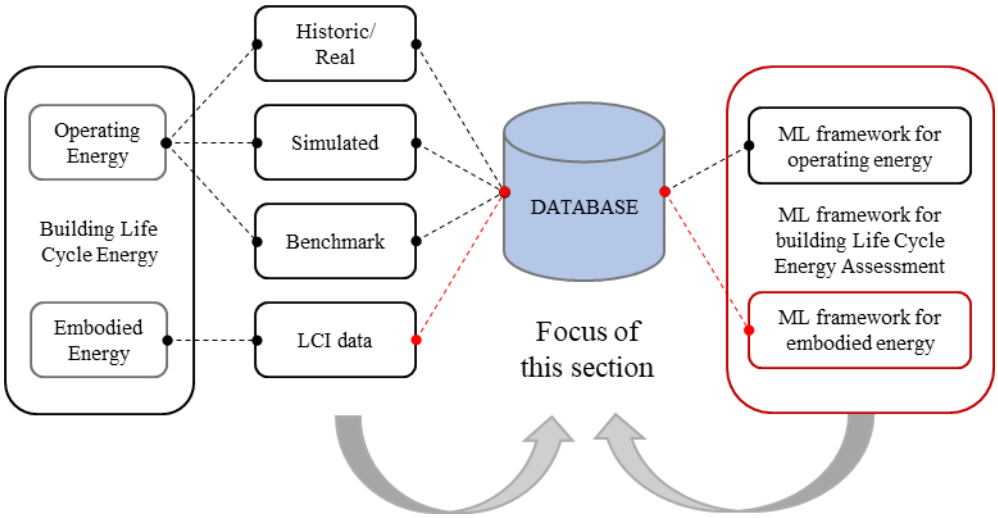


Figure II-5 Illustration highlighting the research gaps and challenges (Reprinted with permission from Venkatraj and Dixit, 2022)

2.4.1 Methodological issues

Literature has highlighted that EE computations suffer from a wide range of methodological issues such as varying building service life and system boundary definitions (energy inputs, products, and processes included in the study), high levels of uncertainty, and the lack of a globally standardized EE calculation protocol. These issues make EE computations much more complicated, time-intensive, and incomparable (Cabeza et al., 2021). As mentioned earlier, the three commonly used methods to evaluate EE include process-based, input-output (IO) based, and input-output based hybrid (IOH) approach. Each of these methods varies in terms of the type/source of data used for calculations, accuracy, and system boundary coverage (Dixit et al., 2010). As a result, it

is extremely difficult to compare the results of different studies (Optis and Wild 2010). Furthermore, all these EE calculation approaches suffer from certain limitations. For instance, process-based calculations utilize material and energy information for each process in the supply chain; accurately tracking this information for every downstream and upstream process is impractical (Stephan et al., 2016). Thus, process-based approaches are known to suffer from incompleteness and truncation errors due to the issue of missing data (Khasreen et al., 2009; Crawford 2009). Similarly, the IO-based approach utilizes economic IO data of the construction sector. This approach is considered unreliable since it utilizes aggregated data and energy price data that may over or under-estimate EE values.

The white-box approaches heavily rely on experiential knowledge and require detailed information regarding various building parameters to make accurate predictions (Touloupaki and Theodosiou, 2017). This information may not be available during the early design phase, making it difficult to integrate BPS into the design process (Yousif and Bolojan, 2021). Moreover, the interdependencies of EE and OE components are not quantified making the design process focused on either EE or OE. We notice that discrepancies in methodology and system boundary coverage also exist between different LCA tools (Dixit, 2017). In addition, these tools have their custom database embedded into the software to perform calculations. Most often, the user does not have access to the database used for LCE calculations (Martínez-Rocamora et al., 2016). As a result, comparing LCE data obtained from different tools is not possible due to the uncertainty associated with the database as well as the LCI technique. Athena IE utilizes the US EPA

TRACI methodology, GaBi database, and covers a ‘cradle to grave’ system boundary definition (Abbasi and Noorzai, 2021) whereas Tally utilizes a custom database and considers a cradle to grave system boundary excluding construction processes (Nizam et al., 2018). Some studies develop custom plug-ins and embed their database to evaluate EE.

Parametric simulation-based LCEA studies do not provide any information regarding the level of development (LOD) used to generate data. The five LODs recognized by the American Institute of Architects (AIA) include LOD 100, LOD 200, LOD 300, LOD 400, and LOD 500 (Soust-Verdaguer et al., 2017). LODs are generally chosen based on the modeler’s skill, availability of time, and computational resources, therefore, causing variations in simulation-based datasets. Table II-1 shows that commonly used parametric modeling tools are Grasshopper for Rhino and Dynamo for Autodesk Revit. Studies that compute EE using the parametric approach do not follow a standard methodology. Feng et al. (2019) utilized the Inventory of Carbon and Energy (ICE) database and considered a cradle to gate system boundary, whereas Kiss and Szalay. (2020) utilized the Ecoinvent database and covered a cradle to gate system boundary excluding transportation. Lobaccaro et al. (2018) utilized the Norwegian EPDs and SimaPro tool and covered a cradle to grave system boundary. All the above-mentioned studies also used mathematical formulas in Microsoft Excel to compute EE. Current efforts to standardize LCEA methodologies are fragmented, subsequently, making it difficult to establish an ML-based LCEA framework.

2.4.2 Issues of data collection, quality, and availability

The robustness of any data-driven solution is purely dependent upon the quality and quantity of collected data (Singaravel et al., 2018). Garbage in, garbage out (GIGO) is a term often used to represent poorly labeled or inaccurate data, data that is based on underlying human prejudices, and incomplete data. The issue of GIGO has, therefore, been a major pitfall in the field of ML (Geiger et al., 2020). As a result, developers spend most of their time preparing/cleaning the data to ensure that the datasets used for training are accurate and complete. Researchers also agree that the process of data collection is more complicated than training the ML model itself (Bourdeua et al., 2019). On a microscale, the proliferation of smart sensors, energy management systems, metering devices, wireless transmission, cloud computing, and other data acquisition technologies, has helped researchers gather large volumes of building energy data (Qiao et al., 2021; Seyedzadeh et al., 2018). However, acquiring energy data for building stocks on a macroscale is extremely difficult (Ye et al., 2021; Wei et al., 2018). Here, it is important to ensure that the collected data is representative of different building types, operating conditions, climatic zones, and thermal characteristics (Bourdeua et al., 2019). Nonetheless, it has been widely noted that only specific data patterns associated with important building energy-related parameters are monitored (Wang et al., 2021). As a result, data about certain building types or input features are sparsely available or unavailable. For instance, Table II-2 shows that a large percentage (62%) of the studies focused on commercial building types. This is mainly because commercial buildings are well equipped with smart sensing meters and other technologies, whereas residential energy data is gathered from

monthly utility bills (less granularity) (Edwards et al., 2012). Sometimes, certain features are either missing or under-represented during model development due to the difficulties associated with data collection and data quality. The prediction model developed in this manner is considered biased, skewed, or limited (Ye et al., 2021). This issue tremendously increases the risk of overfitting the model and reducing its generalization ability (Chahal et al., 2016).

As a common practice, researchers combine data from multiple sources and databases. Gathering data in this manner is riddled with issues of privacy, reliability, and quality (Wang et al., 2021). For this reason, researchers select input features, time-steps, and scale of the model based on the availability of data. The difference in data composition and volume makes it impossible to compare the performance of different ML models. The lack of a uniform data collection strategy, data quality assessment metrics, and data structure format further limits the application of the ML model for building energy prediction (Wang et al., 2021). According to the literature, the process of acquiring building energy data is extremely cumbersome and time-consuming (Bourdeua et al., 2019).

Specifically, our review of the literature indicates that procuring reliable EE data is an ongoing challenge (D'Amico et al., 2019). This is because researchers use different methodologies, background data, and system boundary settings to acquire data (Birgisdottir et al., 2017; Dixit et al., 2010; Dixit et al., 2012). Tracking information about environmental impacts, energy, and material flows that occur during the non-operational life cycle phases (ex: raw material extraction, building materials manufacturing,

transportation, construction, waste disposal, etc.) is time-consuming and nearly impossible (Thilakarathna et al., 2020; Hollberg et al., 2020; Stephan et al., 2016). Especially, due to the diversity of products, equipment, materials, services, and technologies used during the life cycle of a building (D'Amico et al., 2019). Most researchers find it hard to access primary data and often rely on incomplete secondary sources to obtain data (Dixit et al., 2010; Menzies et al., 2007).

To improve LCA data accessibility several embodied carbon databases, life cycle inventory (LCI) databases, and environmental product declarations (EPDs) have been published in recent years (D'Amico et al., 2019; Simonen et al., 2017). These LCA databases are published in accordance with guidelines established by the International Standardization Organization (ISO). Over the years several ISO LCA standards have been in use, they include, ISO 2006, ISO 14040, ISO 14025, ISO 21930, ISO 13790, etc., (Waldman et al., 2020; Reap et al., 2008; Citherlet and Defaux, 2007; Hammond and Jones, 2006). Each of these ISO LCA compliant databases varies widely in their specificity and provides conflicting/misleading information (Waldman et al., 2020; Birgisdottir et al., 2017; Dixit et al., 2010; Pullen et al., 1996). These uncertainties arise from differences in the temporal, technological, and geographical representativeness of EE data (Ramussen et al., 2018; Azari et al., 2016). For instance, the manufacturing technologies used to produce, transport, and deliver steel I-beams in Australia are most likely different from practices in Europe (Azari et al., 2016). Few other data quality issues that induce variation in LCEA results include (i) lack of data transparency and traceability (Waldman et al., 2020), (ii) insufficient, inconsistent, and unregulated reporting formats

(Ramussen et al., 2018), and (iii) absence of standardized protocols, guidelines, and data quality assessment metrics (Venkatraj et al., 2021). Many studies, therefore, agree that EE calculations are subject to high levels of uncertainty compared to OE estimations because of the issues mentioned above (Venkatraj et al., 2021; Azari et al., 2016). The biggest barrier that hinders the use of ML algorithms for building LCEA is the lack of large-scale LCA datasets (D'Amico et al., 2019). Some studies overcome the issue of limited data by generating synthetic data using a parametric modeling approach (Feng et al., 2019; Budig et al., 2020).

2.4.3 Issues of interoperability between different tools

Our review of literature shows that modeling tools and energy simulation software suffer from issues of interoperability (Venkatraj et al., 2020). The BPS tools used to evaluate EE and OE are inadequate, disjointed, and user-hostile (Abbasi and Noorzai, 2021; Santos et al., 2017; Elbeltagi et al., 2017; Attia et al., 2012). Design professionals are, therefore, required to use multiple tools to evaluate the energy performance of a building. LCA tools are not seamlessly integrated with parametric modeling tools (interoperability) causing loss of information (Elbeltagi and Wefki, 2021; Santos et al., 2017). As a result, parametric simulation-based LCEA models are known to have lower fidelity in comparison to real-world building data. Inconsistencies in EE calculations also arise due to the limitations of the parametric tools themselves. For instance, Dynamo cannot access volumetric information regarding building systems such as ducts, pipes, etc., (Hollberg et al., 2020). In addition, running thousands of simulations is error-prone,

computationally expensive, and time-consuming (Harkouss et al., 2018). As a result, design professionals are unable to assess the life cycle energy performance of a building in a quick and seamless manner. The absence of real-time feedback to the designer during the early design phase is often the cause for constructing underperforming buildings in terms of energy use (Asl et al., 2017; Lützkendorf et al., 2015; Cabeza et al., 2014; Yu et al., 2015).

2.5 Research gaps

In chapter II, we reviewed the current state of research in the field of building life cycle energy performance assessment. This review of literature highlighted three main challenges that need to be addressed (as seen in Figure II-6). First, the lack of standardized calculation protocols and system boundary definitions have caused inconsistencies in EE assessments. Second, the issues associated with the lack of LCEA data have hindered the development of ML models for the purpose of building LCE prediction. Third, interoperability between different LCA tools limits design space exploration, especially during the early design phase.

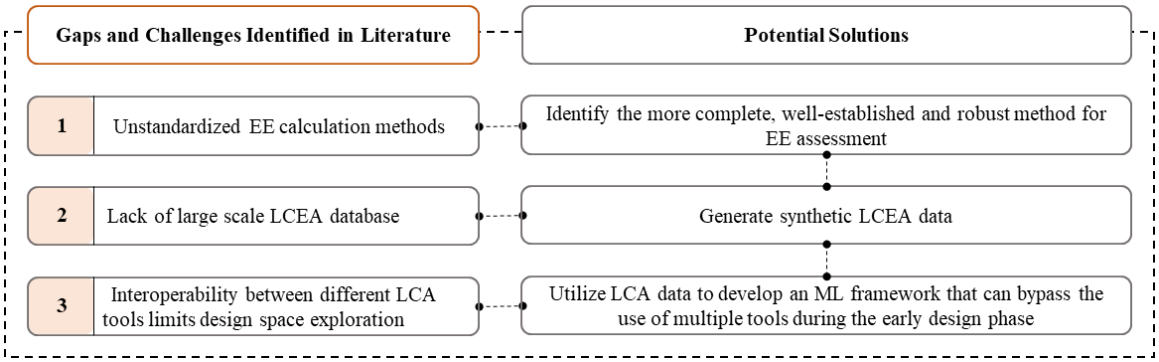


Figure II-6 Overview of research gaps and challenges identified in literature

CHAPTER III

OVERVIEW OF RESEARCH DESIGN AND METHODOLOGY*

3.1 Introduction

Our review of relevant literature helped us identify knowledge gaps and potential solutions in the field of data-driven building life cycle energy performance assessment. The goal of this study is to develop a streamlined method to quickly estimate EE intensity and OE intensity using a few numerical inputs that are known in the early design phase. This goal would be achieved by pursuing the following research objectives:

(1) Conduct a preliminary study to identify the more complete, well-established, and robust method for EE assessment.

(2) Develop a framework to generate a simulation-based building energy dataset for different building typologies.

(3) Develop, train, and test the building ANN prediction model using the simulation-based dataset to comprehensively predict EE intensity and OE intensity.

---

* Part of this chapter is reprinted with permission from:

"Challenges in implementing data-driven approaches for building life cycle energy assessment: A review" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 160, 112327, Copyright (2022) by Elsevier

"Life cycle embodied energy analysis of higher education buildings: A comparison between different LCI methodologies" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 144, 110957, Copyright (2021) by Elsevier

"Evaluating the impact of operating energy reduction measures on embodied energy" by Varusha Venkatraj, Manish Kumar Dixit, Wei Yan, and Sarel Lavy, Energy and Buildings, 226, 110340, Copyright (2020) by Elsevier

(4) Demonstrate the application of the prediction model using two case study buildings.

3.2 Research scope and limitations

In this section, we define the scope for the preliminary study, simulation-based building energy dataset, and the building LCE prediction model.

3.2.1 Preliminary study

This study was solely focused on computing the variation in EE values across the different LCI methods and did not quantify the uncertainty associated with each of these methods. Furthermore, the study only covers two types of buildings, a newly constructed and a newly renovated building, to calculate and analyze their respective EE factors. There are some limitations to this study. First, Athena IE uses a custom database, which is not publicly accessible. As a result, the LCEE values of the process-based approach completely rely on the accuracy of Athena-IE's database. Second, the IOH approaches to cover all activities for which monetary transactions have been recorded. If certain activities were carried out without such monetary transactions, they may not be included in the calculations. Third, temporal representativeness is another issue since IO tables are not updated in a timely fashion. Finally, the calculation of DE for the IOH-based aggregated and disaggregated approach was based on a generic demolition cost, which may not capture all costs associated with demolition, disposal, and hauling away materials for reuse or recycling.

3.2.2 Synthetic building energy dataset

We defined the scope of this dataset in terms of climate zone, building type and form, design parameters, and building loads.

Climate zone: The simulation-based building energy dataset was generated only for climate zone 2A and 5A.

Building type and form: According to Energy Information Administration (EIA), the residential and commercial sectors consumed 21 quadrillion British Thermal Units (Btu) of energy in 2019. Therefore, we were motivated to generate synthetic data only for commercial rectangular buildings. Considering other building types, forms, and taxonomies is beyond the scope of this study i.e., the dataset created in this study does not represent the diversity of the entire building stock.

Design parameters: The design parameters considered in our study can be broadly categorized into two groups: (i) geometry parameters, and (ii) construction parameters. Geometry parameters include orientation, length, and width of the building, number of floors, WWR, and glazing system. Construction parameters consist of ten different types of wall assemblies and three different types of glazing systems.

Building loads: We only consider building envelope loads such as heating, cooling, lighting, and equipment loads in our study since they contribute towards 75% of the total OE use. Moreover, factors such as occupancy loads, type of HVAC system, and schedule are unavailable to the designers during the early design phase (Cheng-wen et al., 2010). Since LCEE consists of a negligible proportion of DE (Stephan and Stephan, 2016),

we have only considered the IEE and REE of building materials in this study (as seen in Figure III-1).

Building LCE	**EE**	**Initial Embodied Energy:** Energy consumed in raw material extraction, on and off-site construction, fabrication, transportation,
		Recurrent Embodied Energy: Energy spent in maintenance, repair, replacement, and renovation activities.
	OE	**Operating Energy:** Building envelope loads such as heating loads, cooling loads, artificial lighting loads, and daylighting.

Figure III-1 System boundary of building loads

3.2.3 Building LCE prediction model

Finally, we defined the scope of the ML model in terms of the data used, type of model (static vs dynamic), purpose of prediction, and performance.

Data: The study utilizes synthetic data that was generated using building energy simulations. The data generated in this manner may be limited (i.e., representative of only certain scenarios) in comparison to real-world building energy data which is far more

complex since it accounts for daily changes in occupancy, HVAC system schedules, etc. That said, the simulation-based approach is advantageous since it allows researchers to build unlimited building models to obtain unlimited synthesized data for ML. Moreover, the study does not include weather variables in this study and utilizes the location parameter as a categorical variable.

Type of model: The building LCE prediction model developed in our study follows a supervised learning approach. Moreover, it utilizes static learning i.e., it does not account for changes in climate, energy production, economy, etc.,

Purpose of prediction: The ML model developed in this study would be able to predict OE and EE using limited information that is available during the early design phase. Predicting the impact of other environmental indicators such as global warming potential, embodied carbon, carbon emissions, acidification potential, eutrophication potential, etc., is beyond the scope of this dissertation.

Performance: First, the performance of this model is validated using a synthetic dataset that represents a subset of commercial rectangular buildings. While the development of this framework is significant, there are still other parameters (ex: location, weather, building types and forms, internal loads, schedules, systems, construction assemblies, etc.,) that need to be integrated with the framework developed in this study to completely generalize and validate the performance of this building LCE prediction model. Therefore, certain inconsistencies in the model performance may occur if the characteristics of the building are different from the dataset used to create the model.

Second, comparing the effectiveness of different ML models in terms of computational time and predictive accuracy is outside the scope of this research.

3.3 Research design and methods

We performed an initial experimental study to identify the more complete, well-established, and robust method for EE assessment. Next, an integrated framework to generate a synthetic building energy dataset, an ML model, and a validation study were developed to achieve our research goals. The framework of our study design has four interdependent parts that are closely tied to the four research objectives mentioned earlier. Each part has several sub-tasks which are elaborated in the relevant chapters of this dissertation. A brief outline of the overall study design (see Figure III-2) is summarized below:

Part 1. Identifying which method of LCI assessment (process-based, IOH-aggregated, or IOH-disaggregated) is well-established and more complete in terms of the system boundary using an experimental study. Furthermore, we will also be determining the impact of each LCI method on EE-OE trade-offs. We believe that the results of this study will justify our intention to utilize the IOH-disaggregated approach in our parametric framework. This process is elaborated further in Chapter IV of this dissertation.

Part 2. Developing an integrated parametric framework to generate a synthetic building energy dataset. The input (building geometry and construction) parameters and their range values will be defined based on our findings from literature. Next, we will be developing a simulation-based framework that (i) generates geometric iterations based on

the input parameters, and (ii) performs energy simulations. This framework will cohesively perform EE calculations using the IOH-disaggregated approach and OE calculations using the EnergyPlus. The inputs and their corresponding outputs will be recorded in a Microsoft Excel database. Eventually, this framework is utilized to generate a large-scale LCE database, which will be utilized by the machine learning model. Further description of the development of the parametric framework can be found in Chapter V of this dissertation.

Part 3. Developing a building LCE prediction model using a supervised learning approach. This includes cleaning, pre-processing, and transforming the data generated in the previous step. Next, we will focus on setting up the multi-output regression model using the procedural framework of a supervised ANN model. Setting up the model will include several sub-tasks such as determining the: (i) train-test split ratio, (ii) network architecture, (iii) hyper-parameters, (iv) regularization techniques, and (v) performance metrics. These sub-tasks are further explained in Chapter VI of this dissertation.

Part 4. Demonstrating the application and prediction accuracy of the developed building LCE prediction model using a case study (Chapter VI).

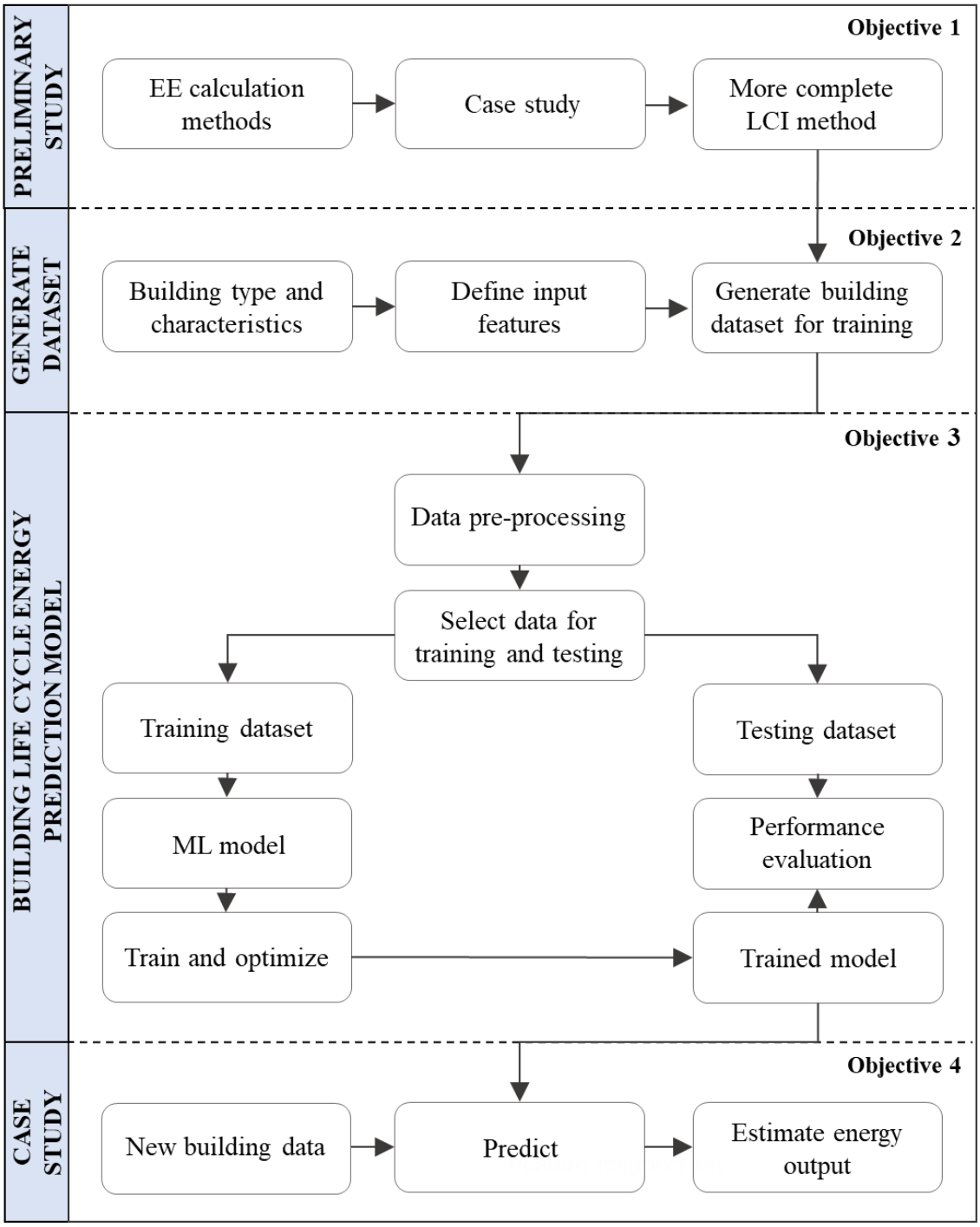


Figure III-2 Conceptual study design highlighting implementation steps

3.4 Main assumptions

We make the following assumptions in this study:

- The useful life span of the building is 60 years.
- The default assumptions for thermostat setpoint, thermal zoning, HVAC system type, lighting power density, equipment power density, and occupancy schedules are compliant with the ASHRAE 90.1-2019 standards.
- The parameters associated with climate, energy supply mix, and economy remain constant over the 60-year period.
- The building is a stand-alone structure i.e., the influence of the surrounding buildings is not considered for OE calculations.
- The EE of materials is the same across different geographic locations.
- The change in structural systems based on building height is ignored for simplicity.

CHAPTER IV

PRELIMINARY STUDY: IDENTIFYING THE MORE COMPLETE LCI METHOD FOR EMBODIED ENERGY ASSESSMENT*

4.1 Introduction

Previous studies (Venkatraj et al., 2020) utilized Autodesk Revit - a building information modeling (BIM) tool to model the building, the Tally plug-in was used to calculate EE using a process-based approach and GaBi database, whereas Autodesk Green Building Studio was used to perform the OE simulations. The results of the previous study have shown that: (1) there are trade-offs between a building's EE and OE for each design measure as well as at the building level; and (2) EE values differ based on the source of data and LCI technique used for its calculations, which may influence the EE-OE trade-offs. The extent of variation in EE values caused by using different LCI techniques (process-based, IOH-aggregated, and IOH-disaggregated) has not been discussed before.

The main goal of this study is to quantify EE-OE trade-offs of educational buildings using two different EE calculation methods and compare and study how such trade-offs differ across the two methods. This goal is met through two research objectives (as seen in Figure IV-1):

---

* This chapter is reprinted with permission from:

"Life cycle embodied energy analysis of higher education buildings: A comparison between different LCI methodologies" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 144, 110957, Copyright (2021) by Elsevier

(i) Calculate and compare the differences in EE values using a process-based approach and input-output-based hybrid (IOH) approach for one newly constructed and one renovated higher education building located in the United States.

(ii) Compute and compare EE factors for the design case and four study cases of the two buildings to discuss the differences in EE-OE trade-offs.

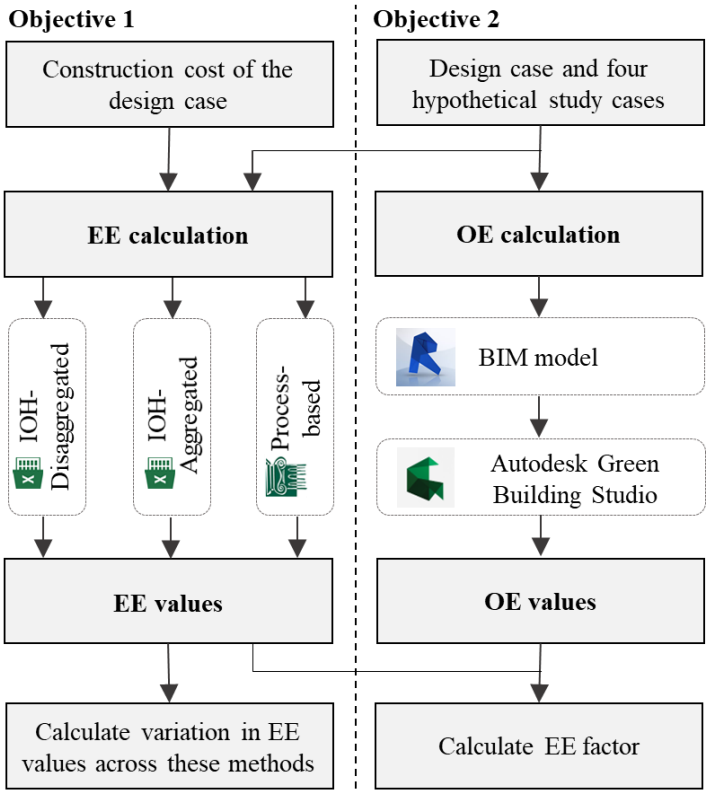


Figure IV-1 Research framework with the two research objectives (Reprinted with permission from Venkatraj and Dixit, 2021)

4.2 Research method

4.2.1 Description of the case study building and variations

The selected case study buildings are one new construction and one renovated educational building located at Texas A&M University. A brief description of the case study buildings can be found below:

4.2.1.1 Experiment 1: New construction

Building 1 is a newly constructed five-story building finished with brick veneer, stone, and aluminum panels on its exterior surfaces (Table IV-1). The building has a reinforced concrete structure with steel-framed construction. This building is designed in accordance with the Leadership in Energy and Environmental Design (LEED) silver standards. Building 1 spaces include theatres, classrooms, laboratories, recording studios, and rehearsal spaces.

4.2.1.2 Experiment 2: Renovation

Building 2, a three-story reinforced concrete structure with brick veneer and cast stone exterior was originally built in 1918. This building underwent a $6.5 million renovation in 2015 to upgrade the building according to modern building codes and standards (Table IV-1). This nearly 100-year-old building was renovated by completely stripping off the building interiors. New wall and roof systems, interior partitions, mechanical, engineering, and plumbing systems, windows, and elevators were added to the existing concrete frame and floors. The existing brick walls were kept in the original condition and interior walls

were added to their inner side. This building is equipped with laboratories, classrooms, auditoriums, conference rooms, and office spaces.

Table IV-1 Characteristics of the case study buildings (Reprinted with permission from Venkatraj and Dixit, 2021)

Building Characteristics	Building 1 (new construction)	Building 2 (renovation)
Cost ($ million)	34.76	6.21
Gross floor area (m^2)	11519	3423.47
Stories	5	3
Year of construction	2012	1918 (renovated in 2015)
Exterior wall construction	4" Brick Veneer 5/8" Sheathing 3" Mineral wool insulation Steel framed walls (2x4, 16" O.C.) 5/8" Gypsum board	4" Brick Veneer 1' Solid brick 5/8" Sheathing 1.5" Rigid Insulation 2.5" Mineral wool insulation Steel framed walls (2x4, 16" O.C.) 5/8" Gypsum board

Table IV-1 shows the characteristics and features of the existing case study buildings. In addition to the existing building designs, four variations of the exterior wall were analyzed to study the EE-OE trade-offs. The existing case study buildings served as the baseline cases to compare the four variations with hypothetical wall assemblies. The purpose is to investigate EE-OE interdependencies hypothetically, exploring differences if the walls were constructed differently. These wall types include: (i) Assembly 1: Stucco and mineral wool insulation, (ii) Assembly 2: Precast concrete panel and mineral wool insulation, (iii) Assembly 3: Precast concrete panel and expanded polystyrene insulation,

and (iv) Assembly 4: Limestone and polyurethane insulation (as seen in Figure IV-2). This resulted in five study cases, one with the original design and four with different exterior wall types for each building.

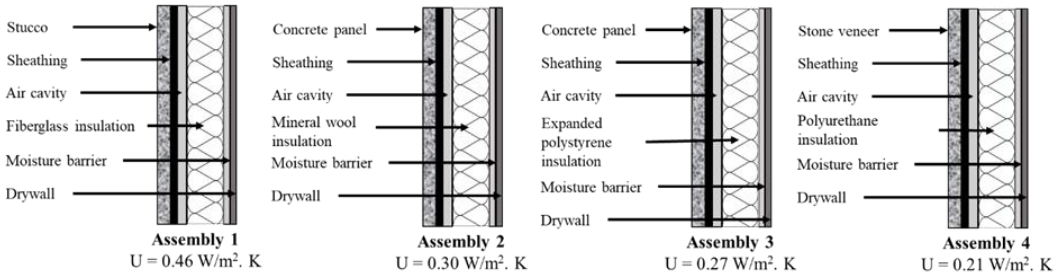


Figure IV-2 Cross-section of exterior wall assemblies used in the study cases (Reprinted with permission from Venkatraj and Dixit, 2021)

4.2.2 Tools and database

Because the appropriateness of an embodied energy calculation method is still debated (Dixit et al., 2015), we used both process-based and input-output-based hybrid (IOH) approaches to analyze the EE-OE trade-offs. Three approaches were applied to compute EE: (1) process-based; (2) aggregated IOH approach; and (3) disaggregated IOH approach.

4.2.2.1 Process-based approach

Prior to conducting EE simulations, we calculated material quantities by multiplying the total material costs by the cost per square foot obtained from the contractor's schedule of values (SOV) and National Constructor Estimator (Pray, 2016), respectively. These material quantities were uploaded to Athena Impact Estimator (IE) – a globally recognized software tool to estimate EE for buildings located in North America. Athena IE complies with the ISO 14040/14044 LCA method and utilizes process-based LCI data for its calculations (Estimator, 2017). Athena IE generates results for a cradle to grave system boundary in accordance with EN 15804/15978 (Estimator, 2017). These EE values are subsequently annualized for a 60-year building lifespan.

4.2.2.2 IOH approach

An IOH model was developed using the United States' 2007 Benchmark IO accounts, which include Make and Use tables. The Make table offers an industry-by-commodity matrix to show how each commodity is produced by different industry sectors. The Use table, on the other hand, is a commodity-by-industry matrix that lists inputs required by each industry sector. We collected data of energy use by each industry sector and integrated it into the Use table in physical units. Using the Make and Use tables, the direct and total requirement matrices were computed that provided energy intensities of each commodity in energy units per US dollar output, which circumvented the use of unreliable energy prices. A detailed description of this IOH model development and data

sources has been given in Dixit and Singh (2018). Using the commodity-specific energy intensities, the IEE, REE, and DE were calculated as follows:

Initial Embodied Energy

IOH aggregated approach: For the Aggregated IOH-approach, we used the total building cost ($C_{bldg.}$), energy source-specific energy intensities of the Educational and Vocational Structures commodity (TE_i), and fuel-specific primary-energy factors (P_i). When computing total requirements, all inputs of the five energy commodities were adjusted to zero to circumvent counting energy inputs twice (Dixit et al., 2015; Treloar, 1998). These removed inputs were compensated through energy source-specific PEFs. The IEE of the study buildings in this approach ($E_{bldg.}$) was calculated using Equation IV-1, shown below:

$$E_{bldg}(MBtu) = \sum_{i}^{n=6} TE_i \ (MBtu/\$) \times P_i \times C_{bldg} \ (\$) \qquad \text{Equation IV-1}$$

where:

'C_{bldg}' is the total building cost,

'TE_i' is the energy source-specific energy intensities of the Educational and Vocational Structures commodity,

'P_i' is the fuel-specific primary-energy factors,

'*n*' is the total number of energy source types,

'i' is the type of energy source (i.e., electricity, natural gas, coal, petroleum, human energy, and capital energy).

IOH disaggregated approach: For the Disaggregated IOH-approach, the total building cost was disaggregated into 23 categories of different materials and services. To avoid double-counting, the commodity inputs of these 23 materials for the aggregated *Educational and Vocational Structures* commodity were removed before re-computing the energy intensity of the aggregated commodity. For all costs other than the 23 categories, we used the energy intensity of the aggregated commodity. A detailed description of the methodology can be found in Dixit and Singh (2018). Equation IV-2 and Equation IV-3 were applied as per (Treloar, 1998) to calculate IEE in this disaggregated approach:

$$E_j = \sum_{i}^{n=6} TE_{i,j} \times P_i \times C_j \qquad \text{Equation IV-2}$$

$$E_{bldg,dis} = \sum_{j}^{m=23} E_j + \sum_{i}^{n=6} TE_{i,other} \times P_i \times C_{other} \qquad \text{Equation IV-3}$$

In the above equations, E_j, $TE_{i,j}$, C_j, and "*n"* denote the EE of a category of material or service "*j"*, total energy intensity (energy source "*i*") of a commodity representing a category of material or service "*j"*, the cost of material or service in category "*j"*, and the number of energy sources or fuel, respectively. In Equation IV-3, "*m*", $TE_{i,other}$, and C_{other} represent the number of material and service categories, total energy intensity (energy source "*i*") of the aggregated *Educational and Vocational Structures* commodity, and cost components other than the 23 categories, respectively. The EE associated with human labor and capital inputs is computed as follows based on Treloar (1998):

$$H_j = C_{h,j} \times \sum_{i}^{n=4} TE_{i,j}$$ Equation IV-4

$$C_j = C_{c,j} \times \sum_{i}^{n=4} TE_{i,j}$$ Equation IV-5

In Equation IV-4 and Equation IV-5, H_j, C_j, $C_{h,j}$, $C_{c,j}$, and $TE_{i,j}$ denote the energy of labor for commodities of category "*j*", the energy capital inputs of commodities of category "*j*", human labor coefficient, capital input coefficient, and total energy intensities of fossil fuel-based energy source "*i*", respectively. There are five energy-producing commodities within the United States' economy: (1) *Oil and gas extraction*, (2) *Coal mining*, (3) *Electric power generation, transmission, and distribution*, (4) *Natural gas distribution*, and (5) *Petroleum refineries*. In this paper, we combined the *Oil and gas extraction* and *Petroleum refineries* commodities. The total number of energy sources ("*n*") in EE calculations, therefore, are four: (1) electricity, (2) natural gas, (3) coal, and (4) petroleum. The cost values were obtained from the contractor's Schedule of Values (SOV), Whitestone Cost Report (Abate et al., 2009), and demolition cost calculator. The PEFs for different energy commodities were sourced from Dixit et al. (2014).

Recurrent Embodied Energy

To compute REE of the study buildings, the total cost of maintenance and replacement for educational buildings was sourced from the White Stone Cost Report (Abate et al., 2009). The total annual maintenance and replacement costs were calculated under three categories: (1) Unscheduled maintenance; (2) renewal and replacement; and

(3) Preventive maintenance and minor repair. Using the energy intensity of the *Nonresidential Maintenance and Repair* commodity of the developed IOH model, the total REE (R_{bldg}) was computed using Equation IV-6 considering a 60-year life cycle

$$R_{bldg} = \sum_{i}^{n=6} E_{i,mr} \times P_i \times C_{mr} \qquad \text{Equation IV-6}$$

where, $E_{i,mr}$; P_i; and C_{mr} denote the energy-source specific energy intensity of the *Nonresidential Maintenance and Repair* commodity, PEF of the energy source "*i*", and total maintenance and replacement cost, respectively. The Demolition Energy (DE) of the study buildings was computed using a demolition cost estimator source from (Building journal, 2020). The IO accounts do not include a separate commodity for building demolition activities. Each of the 12 construction commodities of the IO model is adjusted to include transactions of demolition-related activities. To compute the energy intensity of demolition, we added a demolition commodity by disaggregating each of the 12 construction commodities. In the US economy, the demolition activities are represented by the North American Industry Classification System (NAICS) 238910: *Site Preparation Contractors* under NAICS 238: *Specialty Trade Contractors* (USCB, 2017). We sourced the amounts of *Value of Sales, Shipments, Receipts, Revenue or Business Done* and *Annual Payroll* from the economic census data, US Census Bureau (USCB, 2012), for the whole construction sector as well as for the *Site Preparation Contractors* sector. Using the average percentage (~5.2%) of *Value of Sales, Shipments, Receipts, Revenue or Business Done* (~4.8%) and *Annual Payroll* (~5.7%) of the *Site Preparation Contractors* sector, we extracted 5.2% of the direct requirements of the 12 construction commodities and

combined them into one to represent the demolition commodity. Using the modified direct requirement matrix, the total requirements were computed for demolition activities. The energy source-specific intensity of the newly created demolition commodity was multiplied with the calculated demolition cost to determine the total DE of the study buildings.

4.2.2.3 Operating energy

We first modeled the exterior wall types of the study cases in compliance with the ASHRAE 90.1-2016 standards (ASHRAE 90.1, 2016) for climate zone 2A (College Station). According to the prescriptive codes, the exterior wall assembly needs to have a maximum U-value of 0.50 W/(m²·K) and the insulation needs a minimum R-value of 2.29 (m²·K)/W. These standards were used as a guide to develop the four wall types that would be suitable for climate zone 2A. Figure 2 lists the U-values of the wall types considered in our study. Assembly 1 with U = 0.46 W/(m²·K) is the least insulated wall type, whereas assembly 4 with U = 0.21 W/(m²·K) is the most insulated wall type. After creating these wall types, we modeled the existing building and the study cases on Autodesk® Revit® – a building information modeling tool. These models were then exported from Autodesk® Revit® in the gbxml format and uploaded to Autodesk® Green Building Studio® to compute the OE. The study considers heating, cooling, and lighting loads in its system boundary of OE calculation. Weather data for College Station, Texas, was used to simulate the energy use of the building. The annual OE consumption varies based on the level of insulation, exterior finish, and building location. The climate characteristics of College

Station are in Table IV-2. Autodesk® Green Building Studio® calculates the annual OE consumption of the building in terms of electricity and natural gas. The values of operating energy are converted to primary energy using the PEFs of 4.64 and 1.9 for electricity and natural gas, respectively (Dixit et al., 2014). Furthermore, this study did not consider the impact of the surrounding buildings for OE calculations and assumed the baseline building and its variations to be stand-alone buildings.

Table IV-2 Climate characteristics of College Station, Texas (Reprinted with permission from Venkatraj et al., 2020)

Climate Characteristics	
Location	College Station, Texas, United States
Heating degree days	847
Cooling degree days	1607
Köppen-Geiger classification	Cfa
ASHRAE climate zone	2A

4.2.3 Evaluating and interpreting the results

To quantify the EE-OE trade-offs, this study used EE factor – "the amount of EE spent to save one unit of operating energy", as a trade-off indicator (Venkatraj et al., 2020). The existing designs of building 1 and building 2 are considered the baseline buildings for the newly constructed and renovated educational buildings, respectively. The EE factors for the two case study buildings were calculated in a two-step process. In the first step, we calculated the difference in EE and OE between the baseline and each study case as ΔEE

and ΔOE as per Equation IV-7 and Equation IV-8, respectively. In the second step, we calculated the EE factor as the ratio of ΔEE to ΔOE as seen in Equation IV-9. These EE factors represent the EE spent in MJ to save 1 MJ of OE. Like the embodied energy values calculated earlier, these EE factors are also representative of the entire building life cycle (i.e., IEE, REE, and DE). Furthermore, we also determined and compared the variation in EE factors based on the LCI method used to compute EE.

$$\Delta EE = EE_{\text{baseline}} - EE_{\text{study case}} \qquad \text{Equation IV-7}$$

$$\Delta OE = OE_{\text{baseline}} - OE_{\text{study case}} \qquad \text{Equation IV-8}$$

$$\text{EE factor} = |\Delta EE / \Delta OE| \qquad \text{Equation IV-9}$$

4.3 Results

The LCEE values per unit area for the two case study buildings (baseline) and the variations computed using the three different LCI methods (process-based, aggregated IOH, and disaggregated IOH methods) are listed in Table IV-3 and Table IV-5, respectively. The first row of Table IV-3 and Table IV-5 list the LCEE values that were calculated using Athena IE – a process-based LCEE calculator tool; the second and third rows list the LCEE values quantified using the IOH-aggregated and IOH-disaggregated methods, respectively. The IOH-aggregated method used energy intensities of the aggregated Educational and Vocational Structures commodity, whereas the IOH-disaggregated method utilized the fuel-specific energy intensities of 23 disaggregated material groups. Table IV-3 and Table IV-5 emphasize the variation in LCEE values caused by using different LCI methods to compute EE. The column next to the LCEE

values shows the variation in LCEE values between the process-based and IOH-based approaches in terms of percentage difference.

This study also calculated EE factors to indicate the EE-OE trade-off for all five cases of the two case study buildings. The baseline column of Table IV-4 and Table IV-6 lists the annualized EE and OE values per unit area for the baseline case. The remaining columns list the values of ΔEE (LCEE difference from the baseline), ΔOE (OE difference from the baseline), and EE factors for each assembly. Positive values of ΔEE and ΔOE in Table IV-4 and Table IV-6 denote that the corresponding EE and OE values of the study case are less than that of the baseline, whereas negative values indicate otherwise. The values of ΔEE and ΔOE are calculated in MJ/m^2/year as per Equation IV-7 and Equation IV-8, whereas the EE factor is calculated as per Equation IV-9. Across the rows, we notice that the use of different LCI methods cause the EE values to vary, while the OE remains constant. Although the OE values are constant for each variation, the values of EE factor vary due to the change in EE values.

Table IV-3 LCEE values and % difference for Building 1: New construction (Reprinted with permission from Venkatraj and Dixit, 2021)

LCI method	Baseline		Assembly 1		Assembly 2		Assembly 3		Assembly 4	
	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference
Process-based	18748.55	0%	18262.73	0%	26536.95	0%	26608.01	0%	25807.08	0%
IOH-Aggregated	45053.04	140%	33088.63	81%	47392.72	79%	31165.02	17%	30370.41	18%
IOH-disaggregated	75973.95	305%	73299.76	301%	82756.74	212%	83494.51	214%	82107.87	218%

Table IV-4 EE factors for Building 1: New construction (Reprinted with permission from Venkatraj and Dixit, 2021)

LCI method	Baseline		Assembly 1			Assembly 2			Assembly 3			Assembly 4		
	EE MJ/m^2/year	OE MJ/m^2/year	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor
Process-based	312.48	3334.5	8.10	4.04	2.01	-129.8	115.66	1.12	-130.9	37.64	3.48	-117.6	69.28	1.70
IOH-Aggregated	750.88	3334.5	199.41	4.04	49.42	-38.99	115.66	0.34	231.47	37.64	6.15	244.71	69.28	3.53
IOH-disaggregated	1266.2	3334.5	44.57	4.04	11.04	-113.0	115.66	0.98	-125.3	37.64	3.33	-102.2	69.28	1.48

Table IV-5 LCEE values and % difference for Building 2: Renovation (Reprinted with permission from Venkatraj and Dixit, 2021)

LCI method	Baseline		Assembly 1		Assembly 2		Assembly 3		Assembly 4	
	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference	LCEE (MJ/m2)	% difference
Process-based	**6083.57**	**0%**	**6251.00**	**0%**	**8919.04**	**0%**	**8935.71**	**0%**	**8429.49**	**0%**
IOH-Aggregated	**10866.00**	**79%**	**8882.51**	**42%**	**11304.53**	**27%**	**8383.02**	**-6%**	**8155.47**	**-3%**
IOH-disaggregated	**13710.50**	**125%**	**13884.27**	**122%**	**14954.10**	**68%**	**15014.72**	**68%**	**14850.45**	**76%**

Table IV-6 EE factors for Building 2: Renovation (Reprinted with permission from Venkatraj and Dixit, 2021)

LCI method	Baseline		Assembly 1			Assembly 2			Assembly 3			Assembly 4		
	EE MJ/m^2/year	OE MJ/m^2/year	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor	ΔEE MJ/m^2/year	ΔOE MJ/m^2/year	EE factor
Process-based	101.39	3214.9	-2.79	-60.86	0.05	-47.26	63.09	0.75	-47.54	-27.25	1.74	-39.10	90.86	0.43
IOH-Aggregated	181.10	3214.9	33.06	-60.86	0.54	-7.31	63.09	0.12	41.38	-27.25	1.52	45.18	90.86	0.50
IOH-disaggregated	228.51	3214.9	-2.90	-60.86	0.05	-20.73	63.09	0.33	-21.74	72.75	0.30	-19.00	90.86	0.21

4.3.1 Experiment 1: New construction

Table IV-3 shows that the LCEE values for the baseline calculated using the process-based, IOH-aggregated, and IOH-disaggregated methods are 18748.55 MJ/m^2, 45053.04 MJ/m^2, and 75973.95 MJ/m^2, respectively. Here, the percentage difference in Table IV-3 indicates that the LCEE values calculated using the IOH-aggregated and IOH-disaggregated approaches are higher by 140% and 305%, respectively. Similarly, for assembly 1, the LCEE value calculated using the process-based approach is 18262.73 MJ/m^2, whereas the LCEE values for the IOH-based approaches are much higher (Table IV-3). In comparison with the process-based approach, the IOH-aggregated and IOH-disaggregated approaches show an increase in LCEE values by 81% and 301%, respectively. But for assembly 4, we observe a relatively smaller percentage difference of 18% and 218% for the IOH-aggregated and IOH-disaggregated approaches, respectively. Typically, these results indicate that the LCEE values computed using the IOH-aggregated and IOH-disaggregated methods are much higher than the process-based approach.

Figure IV-3 shows the proportion of LCEE components that were calculated using three different LCI methods for the baseline of building 1. The proportions of IEE in building LCEE calculated using the process-based, IOH-aggregated, and IOH-disaggregated approaches are 80%, 64%, and 74%, respectively. Similarly, REE across the three methods are 11%, 36%, and 26%, respectively. The proportion of DE is nearly zero in the IOH-based approaches. Figure IV-4, Figure IV-5, and Figure IV-6 illustrate the relative proportions of IEE, REE, and DE in building LCEE across the three different LCI methods for the four wall assemblies analyzed in this study. For assembly 1, the IEE

(MJ) per unit of area (m^2) varies from 15163 MJ/m^2, 16697 MJ/m^2, and 44532 MJ/m^2 across the process-based, IOH-aggregated, and IOH-disaggregated approaches, respectively (Figure IV-4). Similarly, Figure IV-5 and Figure IV-6 show the values of REE and DE, respectively, across the three different LCI methods. In Figure IV-5, we observe that the REE does not vary based on the assembly while using the IOH-aggregated and IOH-disaggregated approach. This is mainly because the IOH-aggregated and IOH-disaggregated method utilize maintenance and replacement cost based on the type of building, building lifecycle (number of years the building is functional – in this case we assume building lifecycle to be 60 years) and the building area (square foot). This cost is calculated under three different categories, they are: (i) unscheduled maintenance; (ii) renewal and replacement; and (iii) preventative maintenance and repair. Eventually, the REE was calculated using equation IV-6 mentioned earlier in this report. However, the REE calculations in the process-based software tool (Athena IE) shows some variation in REE across different assemblies.

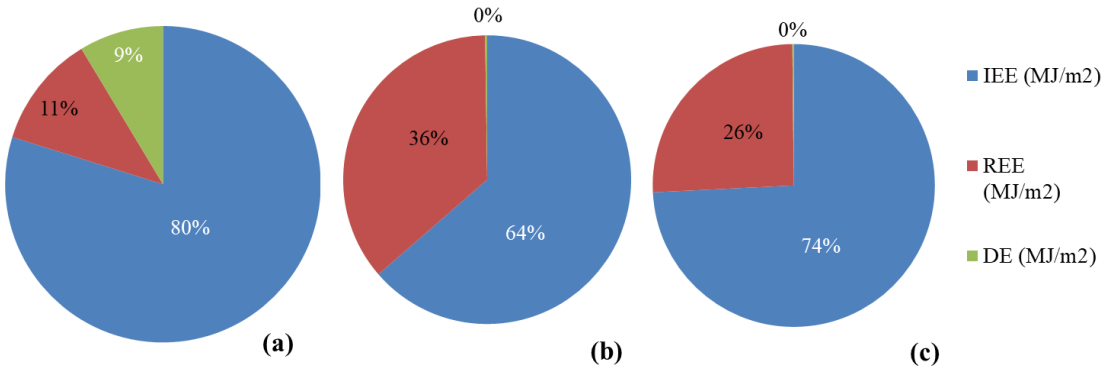
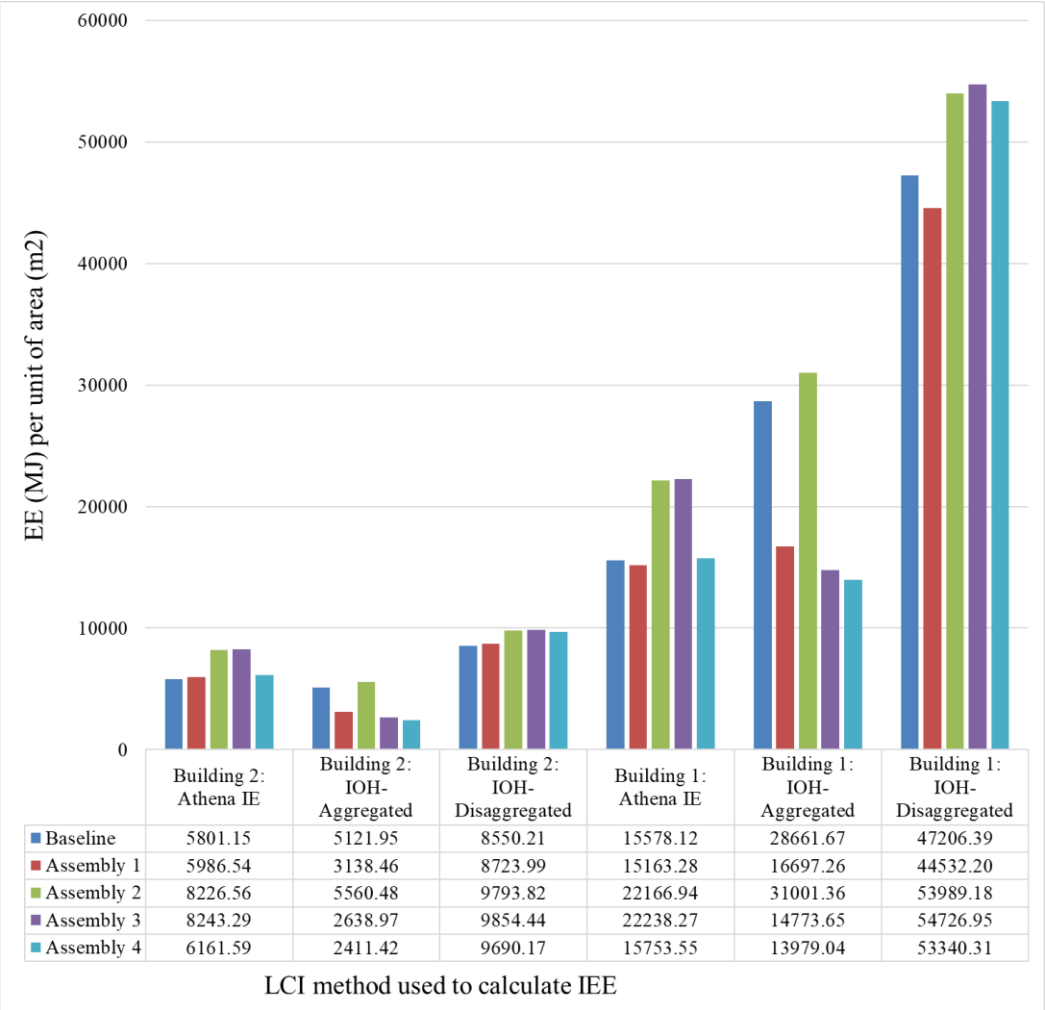


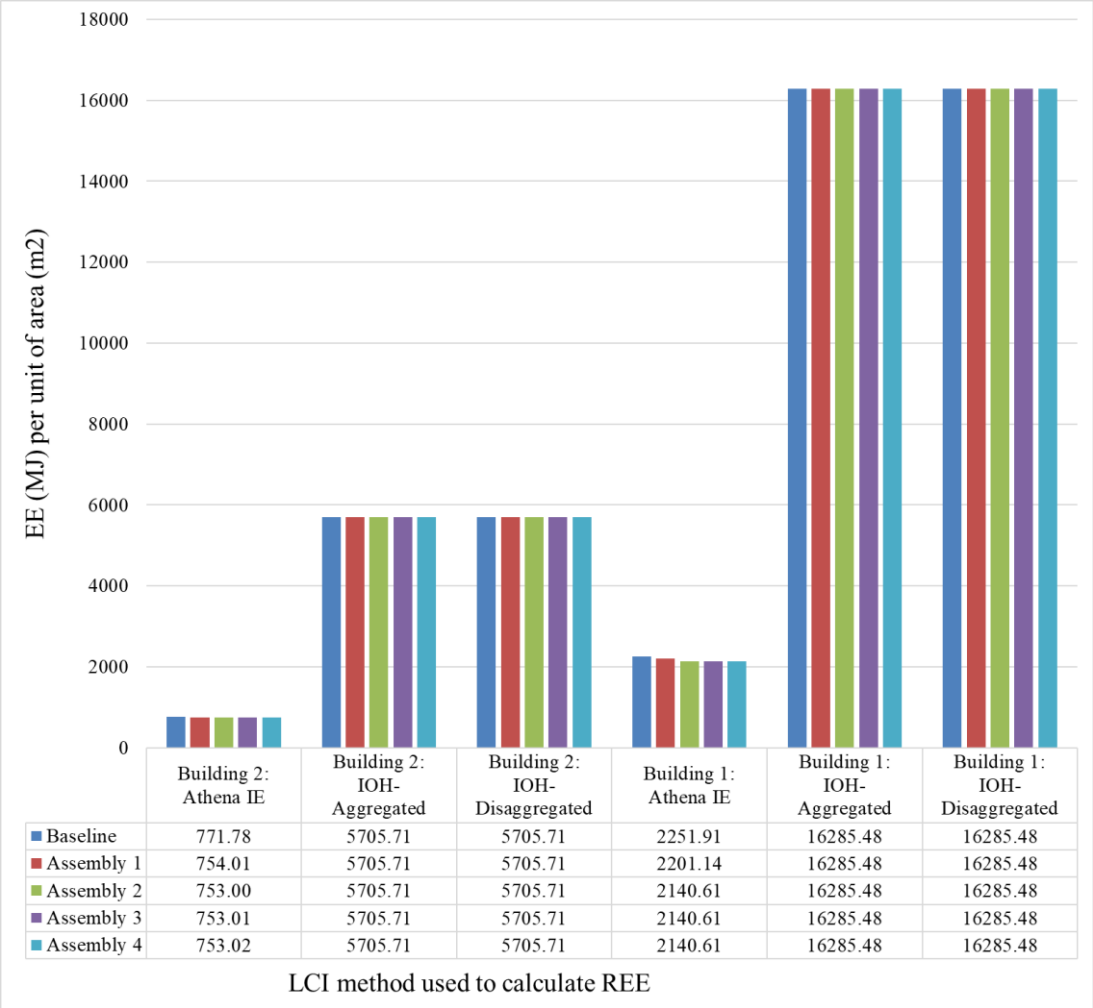
Figure IV-3 Proportion of building LCEE components calculated using (a) process-based approach, (b) IOH-aggregated approach, and (c) IOH-disaggregated approach for the baseline of building 1 (Reprinted with permission from Venkatraj and Dixit, 2021)

Table IV-4 shows that assembly 3 has the highest EE factor of 3.48, while assembly 2 has the lowest EE factor of 1.12 for the process-based approach. For the IOH-aggregated approach, assembly 1 has the highest EE factor of 49.42, while assembly 2 has the least EE factor of 0.34. Similarly, for the IOH-disaggregated method, we observe that assembly 1 has the highest EE factor of 11.04, while assembly 2 has the lowest EE factor of 0.98. The IOH-based approaches show that assembly 1 has the highest EE factor, whereas the process-based approach indicates that assembly 3 has the highest EE factor. Across all the LCI methods we notice that assembly 2 has the lowest EE factor, indicating that assembly 2 has the highest ΔOE of 115.66 MJ/m^2/year for a small increase in EE.



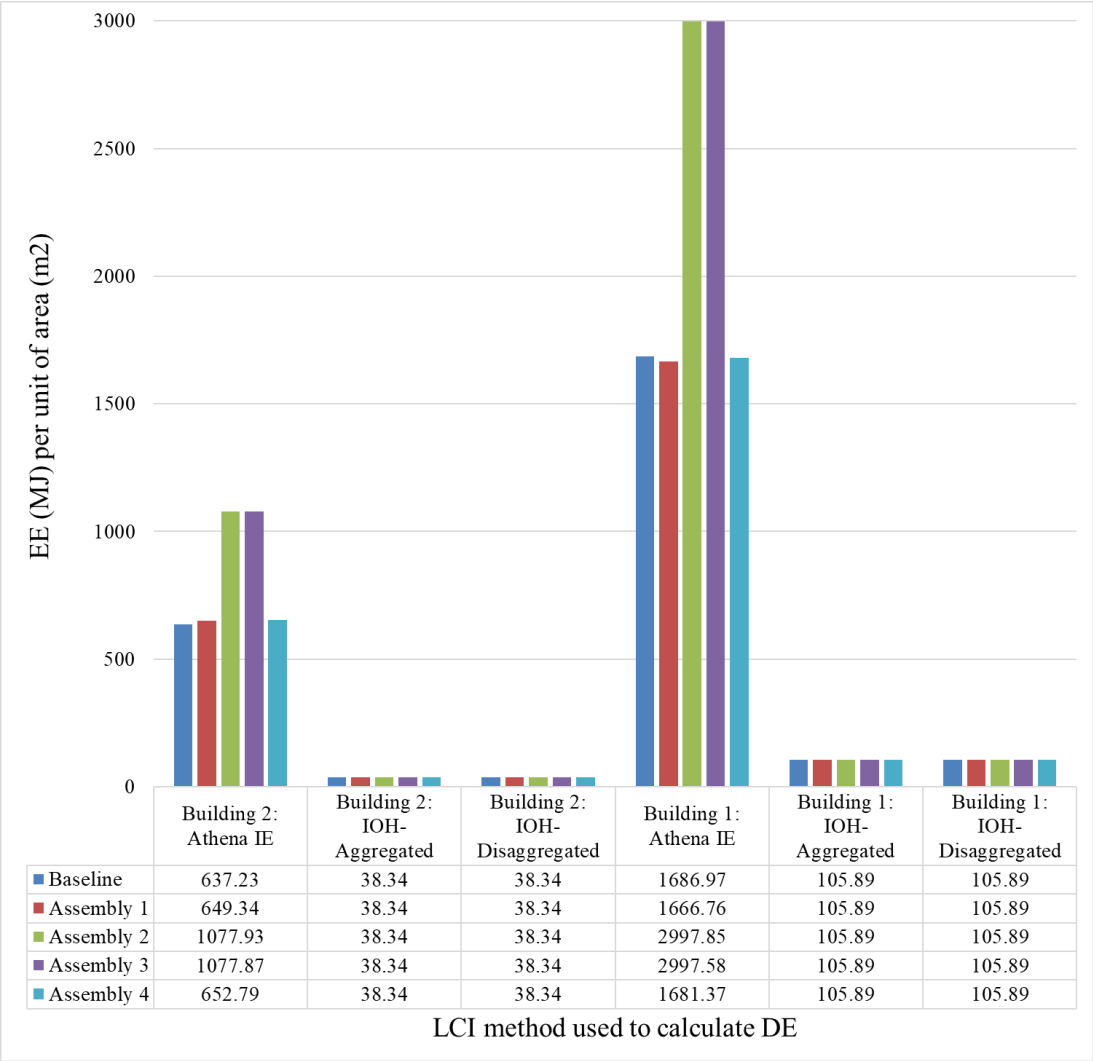
	Building 2: Athena IE	Building 2: IOH-Aggregated	Building 2: IOH-Disaggregated	Building 1: Athena IE	Building 1: IOH-Aggregated	Building 1: IOH-Disaggregated
Baseline	5801.15	5121.95	8550.21	15578.12	28661.67	47206.39
Assembly 1	5986.54	3138.46	8723.99	15163.28	16697.26	44532.20
Assembly 2	8226.56	5560.48	9793.82	22166.94	31001.36	53989.18
Assembly 3	8243.29	2638.97	9854.44	22238.27	14773.65	54726.95
Assembly 4	6161.59	2411.42	9690.17	15753.55	13979.04	53340.31

Figure IV-4 Proportion of IEE across different LCI methods and assemblies (Reprinted with permission from Venkatraj and Dixit, 2021)



	Building 2: Athena IE	Building 2: IOH-Aggregated	Building 2: IOH-Disaggregated	Building 1: Athena IE	Building 1: IOH-Aggregated	Building 1: IOH-Disaggregated
Baseline	771.78	5705.71	5705.71	2251.91	16285.48	16285.48
Assembly 1	754.01	5705.71	5705.71	2201.14	16285.48	16285.48
Assembly 2	753.00	5705.71	5705.71	2140.61	16285.48	16285.48
Assembly 3	753.01	5705.71	5705.71	2140.61	16285.48	16285.48
Assembly 4	753.02	5705.71	5705.71	2140.61	16285.48	16285.48

Figure IV-5 Proportion of REE across different LCI methods and assemblies (Reprinted with permission from Venkatraj and Dixit, 2021)



	Building 2: Athena IE	Building 2: IOH-Aggregated	Building 2: IOH-Disaggregated	Building 1: Athena IE	Building 1: IOH-Aggregated	Building 1: IOH-Disaggregated
Baseline	637.23	38.34	38.34	1686.97	105.89	105.89
Assembly 1	649.34	38.34	38.34	1666.76	105.89	105.89
Assembly 2	1077.93	38.34	38.34	2997.85	105.89	105.89
Assembly 3	1077.87	38.34	38.34	2997.58	105.89	105.89
Assembly 4	652.79	38.34	38.34	1681.37	105.89	105.89

Figure IV-6 Proportion of DE across different LCI methods and assemblies (Reprinted with permission from Venkatraj and Dixit, 2021)

4.3.2 Experiment 2: Renovation

Table IV-5 shows the LCEE values for the baseline and the study cases using the three different LCI methods. For the baseline building, the LCEE values computed using

the process-based, IOH-aggregated and IOH-disaggregated methods are 6083.67 MJ/m2, 10866 MJ/m2, and 13710.5 MJ/m2, respectively. Compared to the process-based approach, the LCEE computed using the IOH-aggregated and IOH-disaggregated methods increased by 79% and 125%, respectively (Table IV-5). Similarly, for assemblies 1 and 2, the IOH- aggregated method increased LCEE values by 42% and 27%, respectively, while the IOH-disaggregated method increased LCEE values by 122% and 68%, respectively. However, for assemblies 3 and 4, we notice that the LCEE values computed using the IOH-aggregated method decreased by 6% and 3%, respectively.

Figure IV-7 demonstrates that across the three LCI methods for the baseline of building 2, the process-based approach has the highest proportion of IEE (80%). The IOH-aggregated and IOH-disaggregated approaches have a high proportion of REE of nearly 53% and 40%, respectively. The DE for the process-based approach is nearly 9%. Figure IV-4, Figure IV-5, and Figure IV-6 show the relative proportions of IEE, REE, and DE in building LCEE across different LCI methods and assemblies. We observe that for LCEE calculated using the IOH-aggregated approach, the proportion of IEE is the lowest.

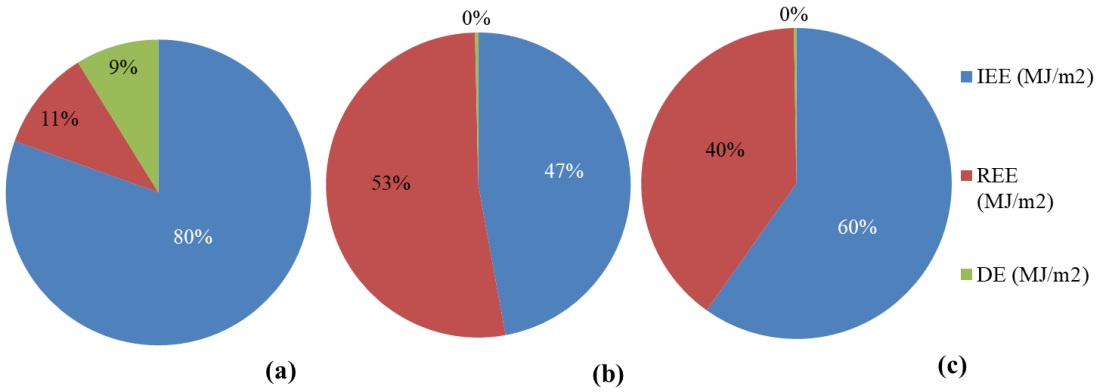


Figure IV-7 Proportion of building LCEE components calculated using (a) process-based approach, (b) IOH-aggregated approach, and (c) IOH-disaggregated approach for the baseline of building 2 (Reprinted with permission from Venkatraj and Dixit, 2021)

Table IV-6 lists the annualized EE and OE values calculated using the three different LCI methods for the baseline building. The annualized EE computed using the process-based, IOH-aggregated, and IOH-disaggregated methods are 101.39 MJ/m2/year, 181.10 MJ/m2/year, and 228.51 MJ/m2/year, respectively. The annualized OE for the baseline building is 3214.9 MJ/m2/year. The process-based and IOH-disaggregated approach shows that assembly 1 has the least EE factor of 0.05. However, the IOH-aggregated approach indicates that assembly 2 has the lowest EE factor of 0.12. Similarly, the process-based and IOH-aggregated approaches show that assembly 3 has the highest EE factor of 1.74 and 1.52, respectively, whereas the IOH-disaggregated approach suggests that assembly 2 has the highest EE factor of 0.33.

4.4 Discussion

Calculating EE is a complicated task due to issues of data accuracy, inconsistent system boundary definitions, and LCI calculation methods (Dixit et al., 2015). This study presents the differences in LCEE values caused by using three different LCI techniques on two case study buildings and their variations. Furthermore, we also calculated the EE factor to quantify EE-OE trade-offs. The variation in EE factors also highlights the uncertainty associated with different LCI methods and data sources used to calculate EE. The results of this study underscore the significance of using the IOH-disaggregated approach to improve the reliability and accuracy of EE calculations. In addition, the use of fuel-specific IEE provides more precise LCEE values, therefore helping us effectively optimize the buildings' carbon footprints.

The variations in the LCEE values across the different LCI methods can be mainly attributed to the differences in system boundary definitions and the use of aggregated/disaggregated energy intensities of commodities. For instance, the system boundary of a process-based approach includes only material inputs, whereas an IOH approach includes material inputs, building systems, and construction services relating to architectural, engineering, and project management tasks. It is important to note that the IOH approach utilizes macro-economic data for its calculations and, therefore, covers a wider system boundary. A process-based approach may also exclude human and capital energy intensities in its calculations, which are covered by an IOH approach. Expanding the system boundary of the process-based approach is associated with high levels of uncertainty since it is difficult to collect supply chain information or data for each major

and minor construction activity. This issue of data unavailability leads to the truncation of system boundary causing 'truncation error' which is a commonly known limitation of the process-based approach. As a result, process-based LCEE values are often underestimated, as substantiated by other studies (Crawford et al., 2018; Crawford, 2008; Lenzen, 2000). The results of our study also closely align with these observations. Table IV-3 and Table IV-5show that the process-based approach underestimates EE and has the least LCEE values for the base case as well as its variations with the hypothetical wall assemblies. Overall, we also observed that the IOH-disaggregated approach has the highest LCEE values across all cases since it uses fuel-specific energy intensities along with more comprehensive and reliable data for its calculations. The large divergence of LCEE values caused by using the IOH-disaggregated method can be attributed to the higher energy intensities of material commodities such as gypsum, steel, concrete blocks, etc. These results are further corroborated by studies conducted by (Acquaye et al., 2010; Crawford, 2004). Furthermore, we notice that as we approach more insulated assemblies (i.e., assembly 4), the variation in percentage difference decreases. For instance, the percentage difference for the LCEE values calculated using the IOH-aggregated and IOH-disaggregated methods for assembly 4 show slight increases of 18% and 218%, respectively, whereas these values increased by 140% and 305% for assembly 1. This may indicate that the energy intensities associated with specific material categories such as limestone and polyurethane may be underestimated in the IOH-approach. The results of our study signify the importance of enforcing a globally accepted EE calculation method. Although global efforts to standardize LCEE calculation are already underway (e.g.

Annex 57 and 72 of the International Energy Agency (Balouktsi et al., 2020), EE computation can be streamlined by developing: (i) place-based consistent LCI databases that are temporally, geographically, and technologically relevant (Azari et al., 2018); (ii) methodological guidelines and EE quantification protocols (Dixit, 2017); (iii) uncertainty quantification frameworks to estimate the extent of uncertainty associated with each LCI method (Azari et al., 2018); and (iv) metrics to determine the quality of data (ex: data quality index) (Ardente et al., 2008). These measures would help the researcher community calculate EE more reliably and accurately.

For building 1, both the IOH-based methods indicate that the low values of ΔOE for assembly 1 contribute to the high EE factor (Table IV-4). We notice that the ΔOE difference decreases for assemblies 3 and 4, indicating that it is difficult for the heat trapped inside the building to escape because of higher levels of insulation, as also concluded by (Rodrigues and Freire, 2017). Even though all the LCI methods suggest that assembly 2 has the best performance, it is important to understand that the EE factor varies significantly based on the method used to compute EE. Moreover, the process-based approach shows that assembly 3 has the worst performance, however, the IOH-approaches suggest that assembly 1 might be the worst. Such differences are also observed in the EE factors calculated for building 2. These results, therefore, signify the importance of understanding the advantages and drawbacks of each LCI technique prior to evaluating EE-OE tradeoffs. In other words, a particular assembly may show different embodied impacts based on which EE inventory is applied to the calculation.

Figure IV-3 and Figure IV-7 illustrate the variation in the proportions of IEE, REE, and DE across different LCI methods for the baseline of the case study buildings. We observe that the process-based approach has the highest proportion of IEE in comparison with the IOH-based approaches, whereas the IOH-based approaches have relatively higher proportions of REE. Figure IV-4, Figure IV-5, and Figure IV-6 indicate that the IOH-disaggregated approach has the highest values of EE per unit of area across all the assemblies and LCI methods. These high values may occur due to the use of material-specific intensities and larger system boundary coverage commonly considered in this approach. Because all building- and system-related preventive and corrective maintenance, and replacement costs were included in maintenance cost calculations, the magnitude of REE is much higher in the case of both IOH-based approaches. The disaggregation of total LCEE also shows that the IOH-based approaches have low proportions of DE in their calculations. Moreover, the proportion of REE and DE is constant across all assemblies since the energy intensity of the aggregated maintenance and replacement sector is considered along with fixed costs, which remain the same for all assemblies. The DE calculation is based on the generic demolition cost calculation, which is also assumed to be the same across the four assemblies. Overall, we observe that the process-based approach has a larger proportion of the building LCEE as IEE, whereas the IOH-based approaches distribute the values of LCEE between IEE and REE.

Table IV-4 and Table IV-6 lists the values of EE factors for the two case study buildings. We observe that the values of the EE factor for the newly constructed building are much higher than that of the renovated building. For instance, the EE factors of the

new construction and renovation building calculated using the process-based approach for assembly 1 are 2.01 and 0.05, respectively. These variations in EE factors become more pronounced in the IOH-based approaches. In this case, the EE factor for assembly 1 of the new construction is 49.42, whereas the renovated building has a low EE factor of 0.54. The low values of EE factors for the renovated building suggest that reusing/refurbishing the existing building structure and its components will have high building LCE savings. This further signifies the importance of reusing a building instead of demolishing it and constructing a new one.

4.5 Summary

In chapter IV, we evaluated the EE-OE relationship for one newly constructed and one renovated educational building by using the EE factor, an indicator to capture EE-OE trade-offs. The extensive review of existing literature suggests that the IOH-disaggregated method to calculate EE provides more complete results. To verify the findings of the literature review, we calculated the EE of the case study buildings and their variations using both process-based and IOH-based approaches. The results illustrate that the EE values calculated using the IOH-disaggregated approach are much higher than the process-based values since it covers a larger system boundary. The results further show that the variation in EE values due to the use of different LCI methods adversely influences the magnitude of EE factors, which represent the EE expense of saving a unit of OE. This means that the magnitude of EE factors may be different for different LCI methods, which further signifies the importance of creating a globally accepted EE calculation method.

This chapter quantified the EE-OE relationship through EE factors, which would significantly help designers and researchers understand the interdependencies between different LCE components and make informed decisions. They could identify and prioritize OE-saving design measures based on lower EE factor values.

CHAPTER V

DEVELOPING AN INTEGRATED FRAMEWORK TO GENERATE A SYNTHETIC BUILDING ENERGY DATASET*

5.1 Introduction

Our review of literature shows that the lack of building LCE data restricts the development of ML-driven solutions, specifically for building LCEA. Moreover, the process of collecting real-world building energy data, especially EE data is extremely time-consuming and nearly impossible (Hollberg and Ruth, 2016; Stephan and Stephan, 2016). Several studies show that techniques of parametric modeling and data augmentation can also be used to generate synthetic data which would enable the development of more robust energy prediction models. Therefore, this chapter seeks to explore the feasibility of generating a synthetic building LCE dataset using a parametric approach.

---

* Part of this chapter is reprinted with permission from:

"Challenges in implementing data-driven approaches for building life cycle energy assessment: A review" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 160, 112327, Copyright (2022) by Elsevier

"Life cycle embodied energy analysis of higher education buildings: A comparison between different LCI methodologies" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 144, 110957, Copyright (2021) by Elsevier

5.2 Synthetic data generation: Basic framework

5.2.1 Description of building characteristics

The medium office building (a simple rectangular building) from DOE's commercial reference building models was selected as the prototype model for this framework (Commercial Reference Buildings | Department of Energy, 2022). We believe that more complex building shapes such as the I, L, U, or H can be derived from a simple rectangle and may be considered in future studies. The building has steel frame walls (2x4 16IN OC) with various exterior finishes. The thermal properties of the building envelope (wall, roofs, windows, and floor) as well as other features such as the HVAC specifications, occupancy schedules, thermostat settings, etc., are in accordance with the ASHRAE 90.-2019 standards (as seen in Table V-1).

Table V-1 Building characteristics

Load parameters	Setting
Thermostat setpoint	75°F Cooling/70°F Heating
Thermal zoning	Perimeter zone depth: 15'
HVAC system	Packaged air-conditioning unit
Lighting power density	0.79 W/ft^2
Equipment power density	0.75 W/ft^2

5.2.2 Define design space and input features

As we already know, it is impractical to generate data that will represent the entire building stock in the United States. Therefore, prior to generating the building dataset, we defined certain design constraints for the geometry and construction parameters based on recommendations from similar studies.

Table V-2 lists the range values for the building length and width parameters obtained from literature. Building dimensions of 13 ft (4 m) might be too narrow and render the space non-functional. Therefore, we selected building dimensions that range from 30 ft (10 m) to 265 ft (80 m) for this study. This implies that the maximum floor area of the buildings generated in this study will be 70,000 ft^2 (6500 m^2). According to the commercial buildings energy consumption survey (CBECS) conducted by EIA in 2012, nearly 93% of the commercial buildings in the United States have an area that is less than 50,000 ft^2 (4625 m^2).

The studies conducted by Ali et al. (2017), Elbeltagi et al. (2017), and Li et al. (2019) have considered the number of stories to vary between 1 and 3. In addition, according to the CBECS, 97% of the buildings located in the midwestern and southern regions of the United States are less than 3-stories high (EIA, 2012). The building length, width, and height parameters are automatically in compliance with the slenderness ratio (height: width = 10: 1). This ratio is usually measured to ensure that the building is structurally stable.

Table V-2 Range values for building dimensions obtained from literature

Study	Range values for building length/width (m)
Singaravel et al. (2018)	4 to 80 / 1 to 80
Fang and Cho (2019)	9 to 17.3
Ali et al. (2017)	10 to 30
Elbeltagi et al. (2017)	10 to 30
Attia et al. (2012)	10 to 12
Li et al. (2019)	20 to 80

For our study, we selected climate zone 2A (Houston, Texas) and climate zone 5A (Chicago, Illinois) due to the differences in the number of heating degree days and cooling degree days. The construction assemblies of all the building envelope components are in accordance with the ASHRAE 90.1-2019 standards as seen in Table V-3. Table V-4 lists the input and output features with their corresponding range values.

Table V-3 ASHRAE 90.1-2019 building envelope requirements

Climate zone	2A - Houston	5A - Chicago
	Assembly U_{max} (Btu/hr.ft^2.^{o}F**)**	**Assembly U_{max} (**Btu/hr.ft^2.^{o}F**)**
Walls	0.077	0.052
Windows	0.500	0.380
Roof	0.039	0.032
Floor	0.107	0.074

Table V-4 Input features with potential range values

Number	I/P feature	Range values	Parameter type
1	Climate zone	2A - Houston 5A - Chicago	Categorical
2	Orientation	0^o – 360^o	Continuous
3	Length	30 ft – 265 ft	Continuous
4	Width	30 ft – 265 ft	Continuous
5	Number of stories	1 - 3	Continuous
6	Height of each floor	10 ft to 15 ft	Continuous
7	Window to wall ratio	10% - 90%	Continuous
8	Glazing system	Climate zone 2A – Single pane glass (U=0.050 Btu/hr.ft^2.^{o}F) – Double pane glass (U=0.40 Btu/hr.ft^2.^{o}F) – Triple pane glass(U=0.35 Btu/hr.ft^2.^{o}F) Climate zone 5A – Double pane glass (U=0.40 Btu/hr.ft^2.^{o}F) – Triple pane glass(U=0.35 Btu/hr.ft^2.^{o}F)	Discrete
9	Wall type	Climate zone 2A – Stucco + Fiberglass (U=0.075 Btu/hr.ft^2.^{o}F) – Brick + Mineral wool (U=0.066 Btu/hr.ft^2.^{o}F) – Concrete panel + Mineral wool (U=0.052 Btu/hr.ft^2.^{o}F) – Stucco + Fiberglass (U=0.051 Btu/hr.ft^2.^{o}F) – Concrete panel + Expanded polystyrene (U=0.047 Btu/hr.ft^2.^{o}F) Climate zone 5A – Stucco + Fiberglass (U=0.051 Btu/hr.ft^2.^{o}F) – Concrete panel + Expanded polystyrene (U=0.047 Btu/hr.ft^2.^{o}F) – Brick + Cellulose (U=0.040 Btu/hr.ft^2.^{o}F) – Brick + Extruded polystyrene (0.039 Btu/hr.ft^2.^{o}F) – Stone veneer + Extruded polystyrene (0.036 Btu/hr.ft^2.^{o}F)	Discrete
Number	**O/P target**		
1	Operating energy intensity		
2	Embodied energy intensity		

5.2.3 Granularity of the simulated dataset

The time-steps considered in ML-based building energy prediction fluctuate between studies as seen in Table V-5. We identified several time-steps that were commonly used from literature, e.g., 1-minute (Mena et al., 2014), 5-minutes (Ahmed et al., 2017; Setiawan et al., 2009), 15 minutes (Fan et al., 2014), 30 minutes (Wang et al., 2020), hourly (Lei et al., 2021; Luo et al., 2020), daily (Lee et al., 2019; Amber et al., 2018), monthly (Turhan et al., 2014) and yearly (D'Amico et al., 2019).

Table V-5 shows that studies using historic datasets consider short-term modeling (sub-hourly, hourly, daily, monthly), whereas simulation-based datasets use long-term (yearly) intervals. The highlighted rows in Table V-5 show that weather-based parameters are included as input features only for short-term prediction intervals while using simulated datasets. Utilizing daily/monthly prediction intervals for OE will make the process of calculating LCE computationally expensive and time-consuming. Therefore, we use yearly timesteps for OE.

Table V-5 Time steps used for simulated datasets (Reprinted with permission from Venkatraj and Dixit, 2022)

Study	ML model	Input variables	Time	Output
Yan and Yao, 2010	BPNN	Envelope	Yearly	OE
Wong et al., 2010	ANN	Weather, Envelope	Daily	OE
Sun and Han, 2013	BPNN	Envelope	Yearly	HL
Turhan et al., 2014	BPNN	Envelope	Yearly	HL
Azari et al., 2016	ANN	Envelope	Yearly	LCA indicators
Li et al., 2019	ANN	Envelope, Climate zone, Geometry	Yearly	OE
Elbeltagi et al., 2017	ANN	Envelope, Geometry	Yearly	OE
Ibeigi et al., 2020	ANN	Envelope, Occupancy	Yearly	OE
Feng et al., 2019	ANN	Envelope, HVAC	Yearly	GWP
Singarvel et al., 2018	ANN	Envelope, HVAC	Yearly	HL; CL
Li et al., 2009	SVM, ANN	Weather	Hourly	CL
Chou and Bui, 2014	BPNN, SVM	Envelope	Hourly	HL; CL
Zhao and Magoules, 2010	SVM	Weather, Envelope	Hourly	OE
Neto and Fiorelli, 2008	ANN	Weather	Daily	OE
Ali et al., 2018	ANN	Envelope, Geometry	Yearly	OE
Kerdan and Galvez, 2020	ANN	Envelope, Climate zone, HVAC	Hourly	OE; LCC
Kamel et al., 2020	XGBoost	Envelope	Hourly	OE
Ekici and Aksoy, 2012	BPNN	Envelope	Yearly	HL
Luo et al., 2020	ANN, SVM, LSTM	Weather, Occupancy	Hourly	HL; CL; LL
Martellotta et al., 2017	ANN	Weather, Lighting, Equipment	Hourly	HL

5.2.4 Generate simulation-based dataset

We generated the building dataset using a parametric approach that allows us to change the building geometry and construction parameters to create several variations of the prototype model as illustrated in Figure V-1. These variations were generated using Rhino 3D (design platform), along with the Grasshopper (GH) extension (visual programming tool).

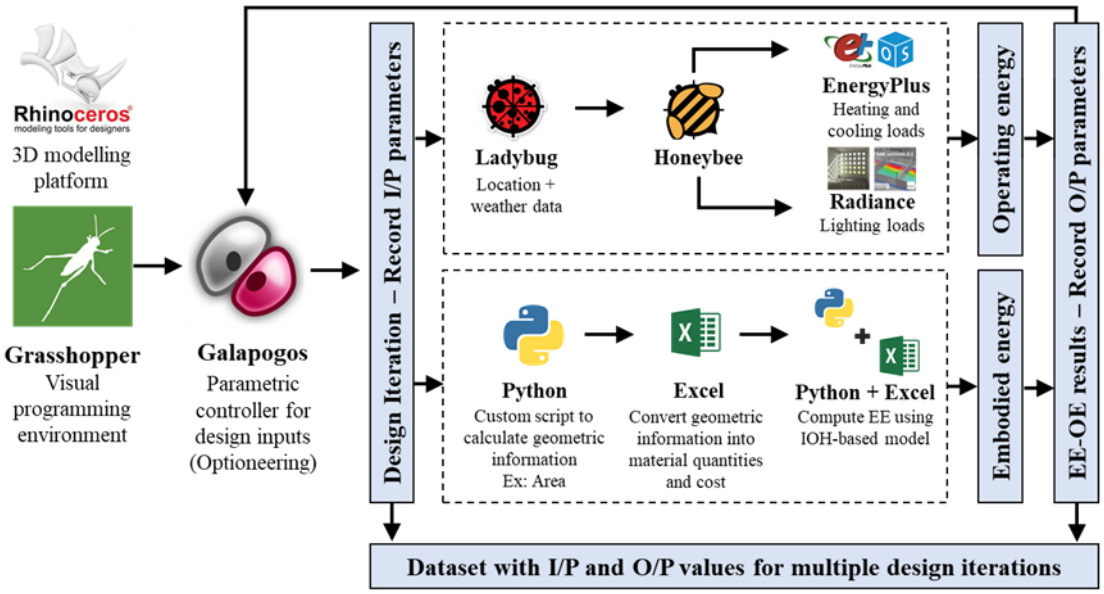


Figure V-1 Framework to generate simulation-based building dataset

The Galapagos add-on within the GH environment acts as a parametric controller for the design inputs. This evolutionary solver was used to manipulate the input parameter sliders based on the previously defined constraints to generate several design options of the prototype model within the design space

5.2.4.1 Operating energy

For OE calculations, each iteration was sent to the Ladybug and Honeybee plugins – these plug-ins use EnergyPlus as a simulation engine for calculations. The OE results for each iteration will be saved into an MS Excel file using the TT Toolbox plugin available in GH.

5.2.4.2 Embodied energy

Similarly, for EE calculations, we computed the surface area of different building components using mathematical formulas and custom scripting in GH ex: walls, floors, windows, etc., Next, the surface area information for each iteration will be exported into a Microsoft Excel file using the TT Toolbox plugin available in GH. From the surface area information, we derived material quantities and converted them into cost information using National Construction Estimator (2017). The total cost of repair and maintenance for a 60-year building lifespan was sourced from the White Stone Cost Report (2009-2010). This cost information is then disaggregated based on the material groups of the IOH model. The material-specific energy intensity and fuel-specific primary energy factors are applied to the disaggregated cost to estimate EE as shown in Figure V-2. A more detailed description of the EE calculation method is available in section 4.2.2.2.

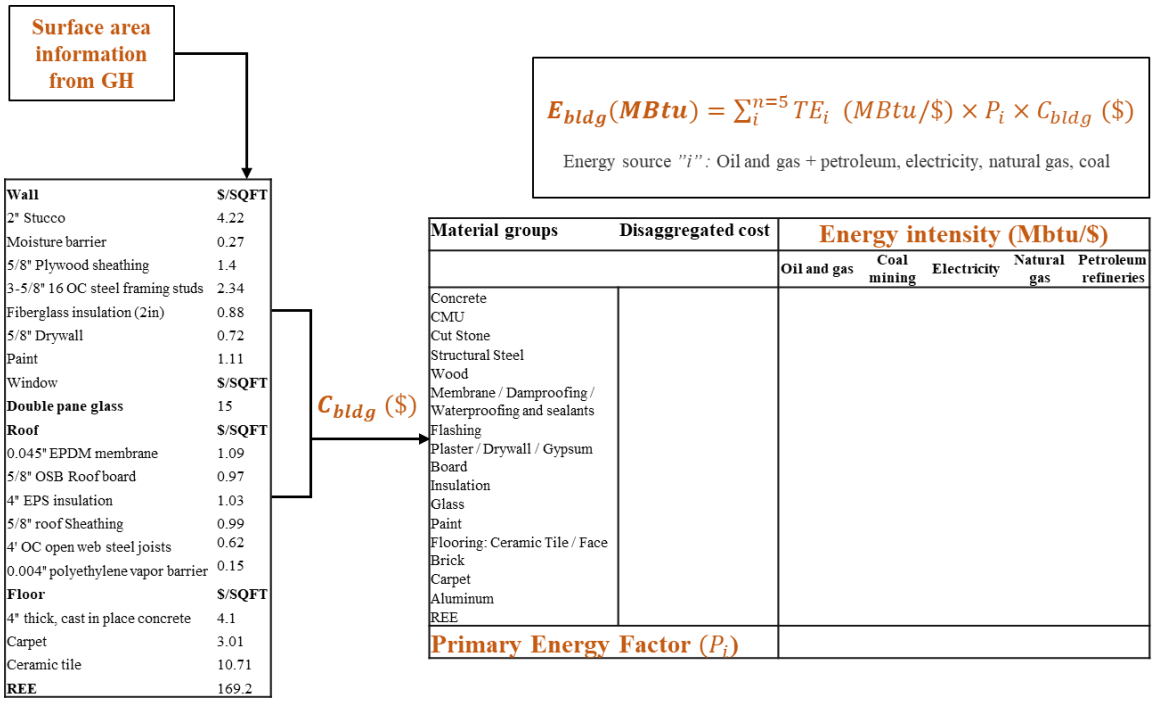


Figure V-2 IOH disaggregated approach for EE assessment

5.2.5 Size of the dataset

Ultimately, we obtained a simulation-based building dataset by recording the input and output values of each design iteration generated using the parametric approach (Figure V-1). The parameter values for each design iteration were exported to a Microsoft Excel database using the TT toolbox plug-in available in GH. Next, we had to determine the approximate size of the simulation-based dataset. Generating a large dataset increases computational cost, processing, and resources, while a small dataset might reduce the representation of the search space (Amasyali and El-Gohary, 2018).

As a rule of thumb, some studies recommend using at least a factor of 10 to 100 times the number of input parameters in the model (Alwosheel et al. 2018; Magnier and Haghighat, 2010). Another rule of thumb is to train the model on at least an order of magnitude with more examples than the input parameters (Google LLC, 2021). Therefore, we decided to utilize analogies from similar parametric simulation-based studies to determine the size of our dataset.

Table V-6 shows the number of input features and the corresponding size of the sample dataset used for training obtained from similar literature. Ultimately, we generated a dataset with 6000 samples.

Table V-6 Correlation between the number of input features and dataset size (Reprinted with permission from Venkatraj and Dixit, 2022)

Study	Number of input features	Size of the sample dataset
Ngo, 2019	14	243
Khalil et al., 2019	8	768
Feng et al., 2019	16	1152
Sharif and Hammad, 2019	12	463
Chou and Bui, 2014	8	768
Ilbeigi et al., 2020	6	1602
Magnier and Haghighat, 2010	20	450
Ye et al., 2018	5	60
Turhan et al., 2014	5	148
Seyrfar et al., 2021	17	1325
Zhong et al., 2019	8	1248
Lei et al., 2021	20	8176
Amber et al., 2018	5	1825
Luo et al., 2020	22	8760
Martellotta et al., 2017	15	3288
Cuilla et al., 2019	12	2184
Cheng-wen and Jian, 2010	20	132
Lee et al., 2019	6	5192
Sharif and Hammad, 2019	10	463

5.3 Implementation of the simulation-based parametric framework

Figure V-3 shows the parametric workflow created in GH to generate the simulation-based building energy dataset. This implementation of this workflow includes the following steps: (i) defining the constraints of the input parameters (sliders), (ii) generating different building geometries using the length, width, number of floors, floor height, orientation, construction assemblies, WWR, and type of glazing system, (iii) assigning appropriate thermal zones and scheduling conditions, (iv) connecting the thermal zones to the energy simulation tool (Honeybee) and (v) extracting surface area information for EE assessment, (vi) performing EE calculations on Microsoft Excel, and

(vii) recording the input parameters and output parameters on a Microsoft Excel file. This entire workflow is then connected to a solver (Galapagos) to continually generate design iterations of the prototype model based on the input constraints and record their corresponding output values.

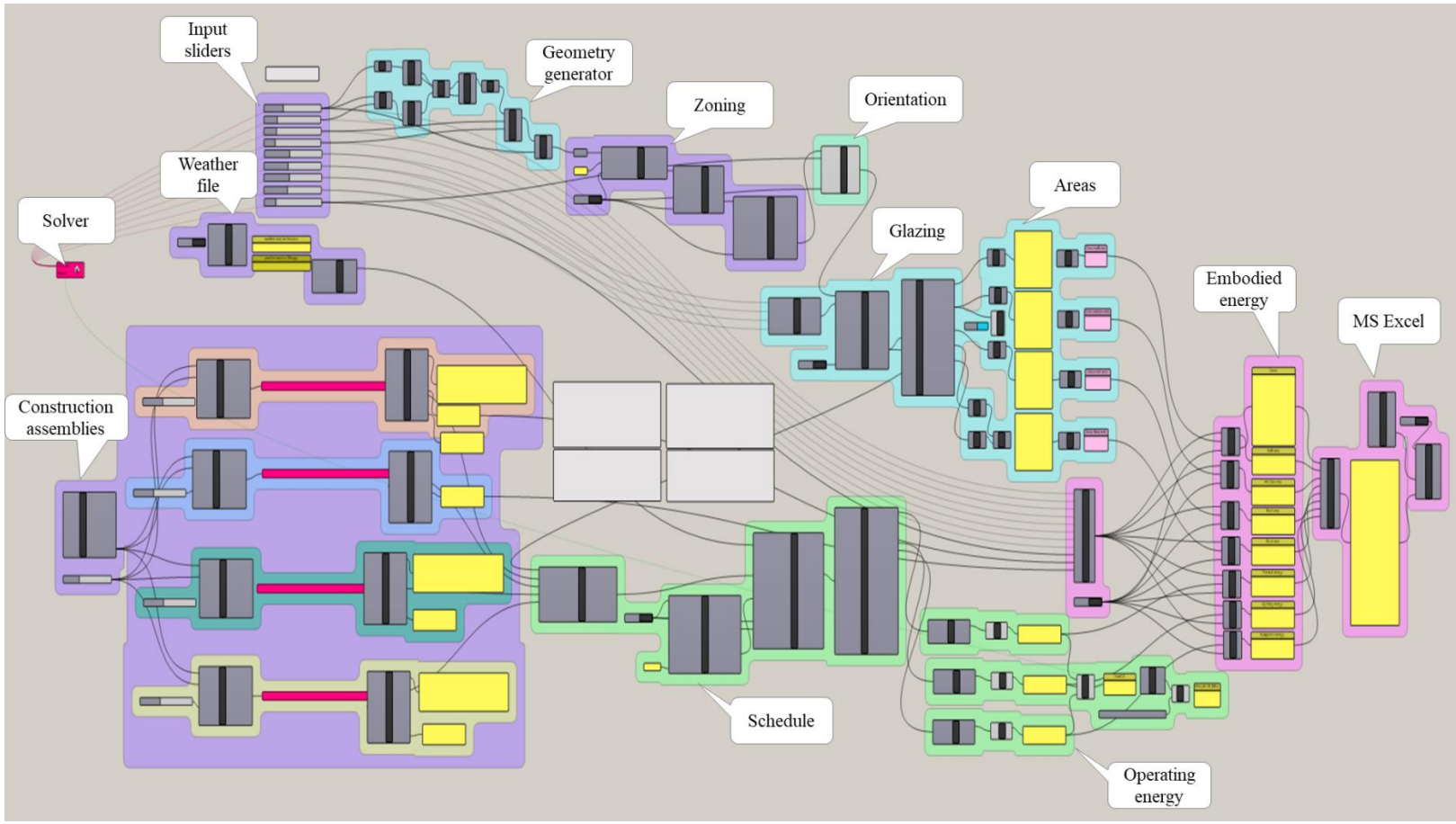


Figure V-3 Parametric workflow to generate the simulation-based dataset

CHAPTER VI

DEVELOPING BUILDING LIFE CYCLE ENERGY PREDICTION MODEL

6.1 Introduction

As mentioned in the previous sections, design space exploration using existing simulation and optimization methodologies is extremely time-consuming and complicated. Moreover, simulation tools used for building LCEA suffer from issues of interoperability. To address these issues, this chapter explores the feasibility of developing a building LCE prediction model, using the synthetic data generated in the previous chapter. This chapter (i) describes the process of developing the supervised ANN model for multi-output regression (OE and EE intensity), and (ii) demonstrates the application and validation of the developed ANN model using a case study building. The Jupyter Notebook application was utilized in this study for model development.

6.2 ANN model development: Basic framework

6.2.1 Feature selection and engineering

To prepare the sample dataset for the ML algorithm, we will have to select suitable input features and discard the remaining features (Khalil et al., 2019). We notice that each study has selected input features that have a significant correlation with the output target (Table 2). This process of choosing suitable inputs that are representative of the sample population is known as feature selection (Ayodel 2010). Furthermore, the number of input features determines the complexity of the model structure (Liu et al., 2019). Decreasing

the number of input features reduces the computational cost, whereas, increasing the input features may improve model performance (Zeng et al., 2018). It is, therefore, important to select the type and number of input features after careful consideration. To determine the input features of the ML model we conducted an initial study (Venkatraj et al., 2020) to evaluate the LCE implications of using OE reduction measures on EE. For this, we modeled 23 variations of an educational building and quantified the EE-OE trade-offs in four different climate zones across the United States. The findings of this study showed that all the input features play an important role in predicting the output. Therefore, we chose to retain all the 12 input features.

The previously generated sample dataset consists of data in its raw form i.e., it has not been prepared for the ML algorithm, and therefore, cannot be used. Discrete categorical variables need to be transformed into a representative numerical format using an encoding scheme before using them in the ML model. This process is commonly referred to as feature engineering. The discrete categorical variable (climate zone) has two distinct levels (2 and 5), each of these climate zones are mapped to binary (0 or 1) values using functions available in the Pandas library (Figure VI 2). We, therefore, create two columns named climate zone 2A and climate zone 5A. Since the variable only has two choices: climate zone 2A or climate zone 5A, it can be accurately represented using a single column of zeros and ones. We, therefore, drop the redundant climate zone 2 columns from the dataset.

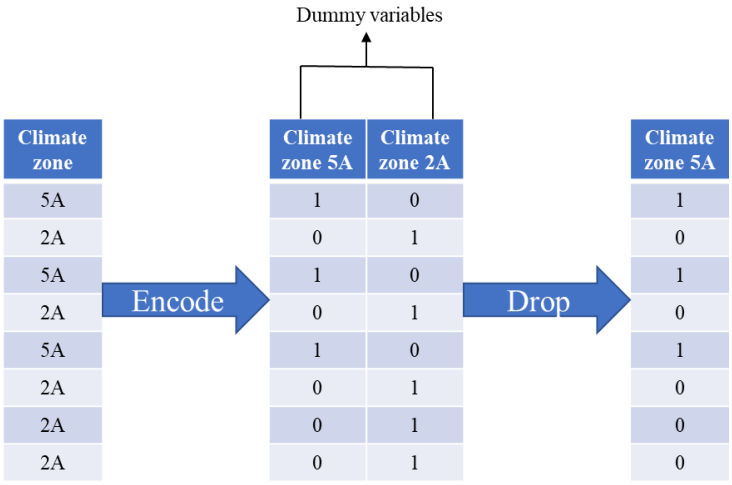


Figure VI-1 Feature engineering of categorical variables

6.2.2 Data pre-processing

After feature engineering, we made sure that missing, repeated, or invalid values are eliminated from the dataset (Amasyali and El-Gohary, 2018; Sharif and Hammad, 2019). Next, the data is transformed into a suitable format for the ML algorithm. Typically, data transformation includes normalization, smoothing, aggregation/disaggregation, or generalization of data (Amasyali and El-Gohary, 2018). Data normalization is essential since some of the input features included in our study do not have units (ex: location, number of stories), some have different units (ex: orientation, length, width, etc.) while a few features have percentages (ex: WWR). In addition, data normalization also prevents certain features with a larger range overcompensate for features with a smaller range (ex: length has a range between 32ft – 265ft, whereas the height of each floor has a range

between 10ft – 16ft) (Sharif and Hammad, 2019). Data normalization unifies and scales all the input features, therefore, preventing an overflow error in the input value. During the process of normalization, the original values are rescaled and shifted to a new range between 0 and 1. This also helps in reducing the dependency on feature selection and improves training performance.

Pre-processing raw data is of utmost importance, especially, while using a data-driven approach, since any discrepancies in the dataset will cause profound errors in training the ML model. The MinMaxScaler function from the Sklearn library was used to normalize the data. Scaling the data into the [0,1] range ensures that all the features contribute equally to fit the ML model. The underlying formula used by the MinMaxScaler function obtained from the Scikit Learn documentation is represented using Equation VI-1 and Equation VI-2.

$$X_{std} = \frac{(X - X_{min})}{(X_{max} - X_{min})} \qquad \text{Equation VI-1}$$

$$X_{scaled} = X_{std} * (max - min) + min \qquad \text{Equation VI-2}$$

In the above equations, *'X'* is used to denote the original value of variable *X*, while *'X_{max}'* and *'X_{min}'* are used to denote maximal value and minimum value of the column, respectively. The feature range to which we want the variables scaled is represented using *'max'* and *'min'*. In this study, the features are scaled between 0 and 1, therefore, the *max* and *min* values are 0 and 1, respectively.

6.2.3 Train-test split

The prepared dataset will be split into two subsets – the training dataset and the testing dataset. The training dataset will be used to train/fit the ML model, while the testing dataset will evaluate the fit of the ML model. The most common ratios used to split the dataset are: (i) 70% training and 30% testing, (ii) 80% training and 20% testing, or (iii) 67% training and 33% testing (Table II-2). Few studies also recommend splitting the data into three subsets – training dataset, validation dataset, and testing dataset. While splitting the dataset it is important to ensure that each subset (i) is large enough to yield statistically meaningful results, and (ii) represents the key characteristics of the entire dataset holistically. Another key factor is to never train on the test dataset (Google LLC, 2021). The test dataset acts as a proxy for new data and allows us to test the ML model.

We used the train_test_split function available in the Scikit Learn package to split the dataset. In this study, we utilized 70% of the data for training, 15% for validation, and the remaining 15% for testing.

6.2.3 Setting up the multi-output ANN model

The ANN model is preferred for solving interdependencies between multiple parameters (Asl et al., 2017; Seyedzadeh et al., 2018; Fan et al., 2017; Krarti, 2003). As mentioned earlier, the basic processing unit in an ANN model is referred to as a neuron/node. Each neuron holds the real-value representation of an attribute and is interconnected to neurons in the next layer of the network through weighted association lines (Ali et al., 2017; As et al., 2018). To further explain, each node in the hidden layer

is a weighted sum of the input nodes, while the output node is the weighted sum of the hidden nodes (Google LLC, 2021). A simple mathematical representation of the summation process is shown using Equation VI-3 and Figure VI-2, where x, w, b, and y refer to the raw input feature value, weight, bias, and model output, respectively.

$$y = \sum_{i=1}^{n} x_i w_i + b_i \qquad \text{Equation VI-3}$$

$$R(z) = \max(0, z) \qquad \text{Equation VI-4}$$

The output value from each hidden layer neuron is then transformed using an activation function to limit its range and set boundaries. In this study, the rectified linear unit (ReLu) was used as the hidden layer activation function. The ReLu activation function is defined using Equation VI-4, where z is the weighted sum of the input node.

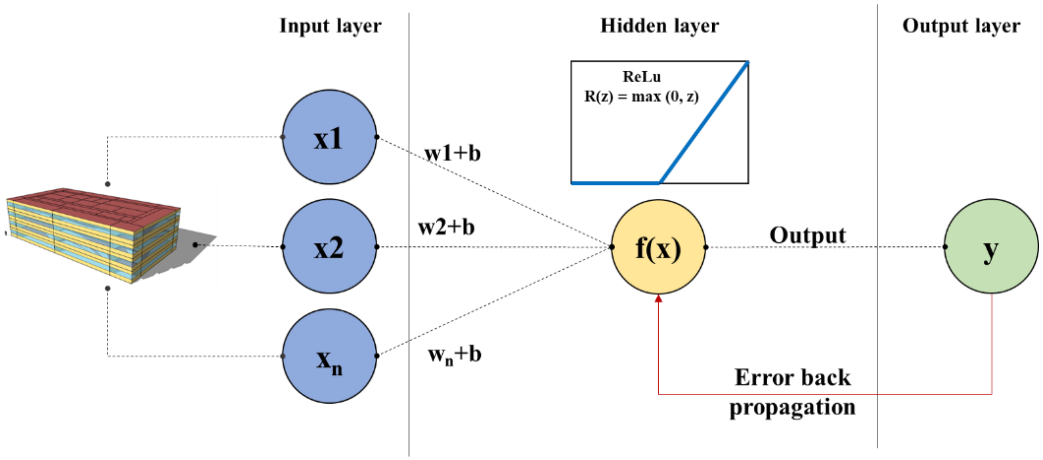


Figure VI-2 Mathematical representation of the hidden layer neuron

Determining the number of hidden layers and hidden layer nodes are the next steps in the process of model development. Current methods to determine the number of nodes are mostly ad-hoc or based on trial-error (Sharif and Hammad, 2019). Therefore, we began the training process, with 7 nodes in the hidden layer based on Equation VI-5, one of the most common rules of thumb in neural network design (Feng et al., 2019; Runge and Zmeureanu, 2019, Fan et al., 2017). Here, n_h, n_i, and n_o represent the number of hidden layer nodes, input layer nodes, and output layer nodes, respectively.

$$n_h = \frac{n_i + n_o}{2} \qquad \text{Equation VI-5}$$

Figure VI-3 illustrates the network architecture of the ANN model that was used during the initial stages of model development. The initial network architecture had 12 nodes in the input layer, 7 nodes in the hidden layer, and 2 nodes in the output layer.

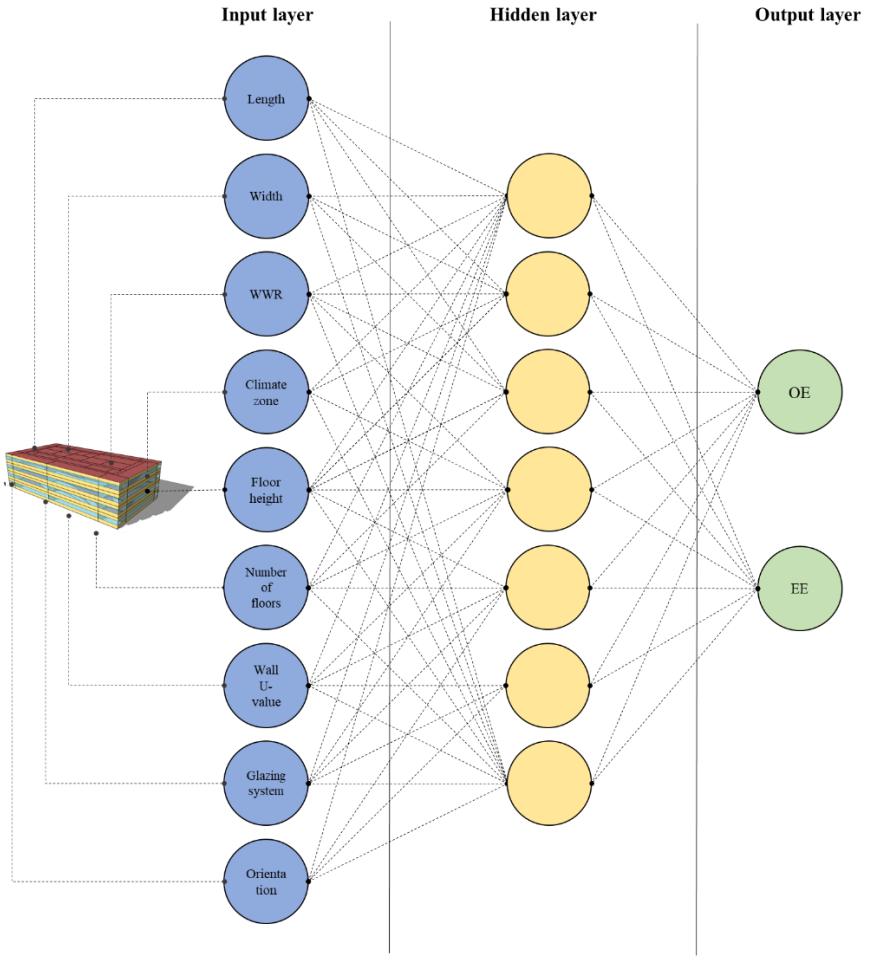


Figure VI-3 Schematic network architecture of the initial model

At the beginning of the training process, random weights are generated for each of these weighted connections. During the training process, the network compares the true output value of the original dataset with the network output to compute the error of the output using a cost (loss) function. The loss function is a metric used to measure the overall error and is continuously monitored for each epoch. This error is backpropagated as

negative feedback to adjust and refine the weights and biases to reduce the difference (Ali et al., 2017). The weights are adjusted using the gradient descent algorithm which essentially helps the model minimize its loss. Ultimately, the goal of the training process is to minimize the loss function and reach the point of convergence, thereby ensuring that the model fits the problem correctly. In this study, we utilized mean squared error (MSE) as the loss function and the Adam optimizer to adaptively adjust the learning rate of the model. The MSE is calculated using Equation VI-6, where n is the number of samples, $\hat{y}$ is the prediction output, and y is the true value.

$$Cost\ function = \ MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y} - y)^2 \qquad \text{Equation VI-6}$$

This iterative process of training continues until it is terminated by a certain criterion. Generally, studies define the stopping criterion based on the maximum number of iterations, loss, or root mean squared error (RSME). Determining an appropriate stopping criterion is very important since it ensures that the model does not overfit the training data and generalizes well. Therefore, in this study, we utilized early stopping as a regularization technique to prevent the model from overfitting. Early stopping terminates the process of model training when the loss starts increasing on the test dataset as seen in Figure VI-4. The patience criterion for early stopping was set to 20 in this study i.e., if the loss increases for 20 consecutive iterations, the training process would be terminated (as seen in Figure VI-5).

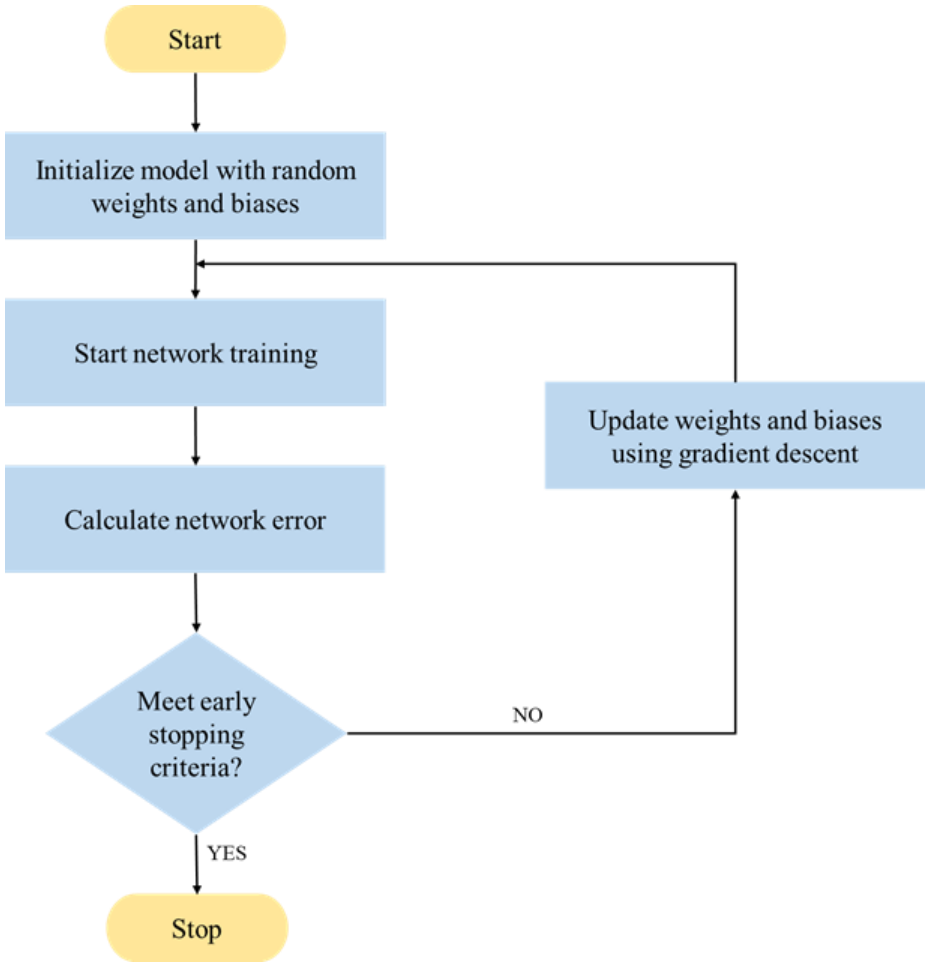


Figure VI-4 Flowchart of the iterative training process

Subsequently, we performed several experiments to improve the performance of the ML model. Several parameters of the initial ML model such as the number of hidden layer nodes, network architecture, learning rate, and activation function, were modified to improve the accuracy of the model. The ML model was created using the TensorFlow, Keras, and Scikit Learn packages.

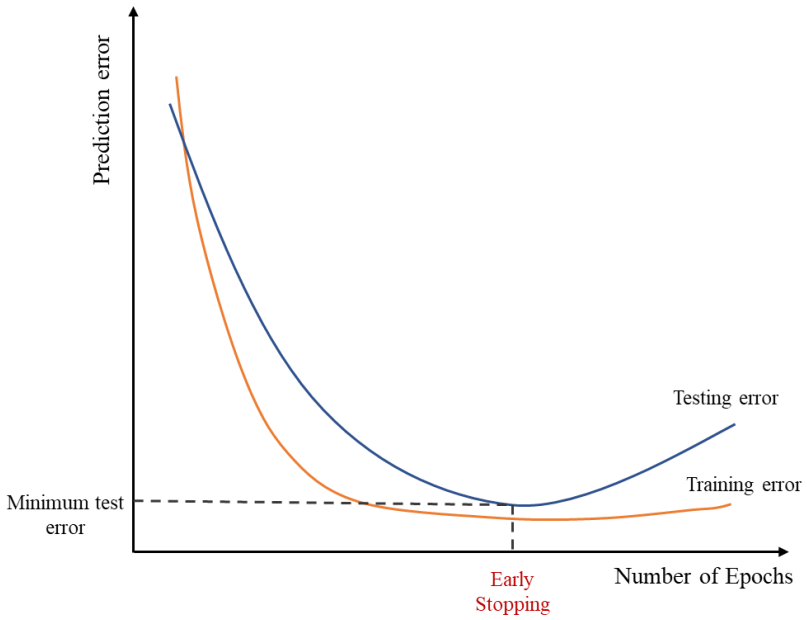


Figure VI-5 Loss curve for early stopping

6.2.4 Performance evaluation of the ANN model

To evaluate the performance of the developed model, we compared the output targets generated from the model and the true value from the test dataset. Some of the standard performance evaluation methods found in literature include the coefficient of determination (R-squared), mean absolute percentage error (MAPE), root mean square error (RMSE), and mean absolute error (MAE). These performance metrics are calculated using Equation VI-7 to Equation VI-10 shown below:

$$R^2 = 1 - \frac{\sum(\hat{y} - y)^2}{\sum(y - \bar{y})^2} \qquad \text{Equation VI-7}$$

$$MAPE\ (\%) = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}-y}{y}\right| \times 100 \qquad \text{Equation VI-8}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}-y)^2} \qquad \text{Equation VI-9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}-y| \qquad \text{Equation VI-10}$$

where:

'n' is the total number of data points,

'$\hat{y}$' is the predicted output,

'y' is the actual output,

'$\bar{y}$' is the mean of the actual output.

6.3 Test case

We tested the generalization ability of the developed ML model on two test cases. Test case 1 is a prototype model of a midrise apartment building located in climate zone 5A retrieved from DOE's commercial reference building models (Commercial Reference Buildings | Department of Energy, 2022) (Figure VI-6). The 4-storey rectangular building has a gross floor area of 33,700 ft^2. The exterior envelope of the steel-framed building is finished with 4" stucco and fiberglass insulation (U= 0.052 Btu/hr.ft^2.^{o}F). The WWR on each side is 20% and the U-value of the windows is 0.40 Btu/hr.ft^2.^{o}F. Every floor has a central corridor with four apartments on each side. Each apartment is 25'x 38' and is considered a seperate thermal zone. The remaining energy simulation settings for the

prototype building is in accordance with the ASHRAE 90.1-2018 prescriptive codes for climate zone 5A. We then estimated the bill of quantities using surface area information for the prototype model and calculated EE intensity using the IOH-disaggregated approach.

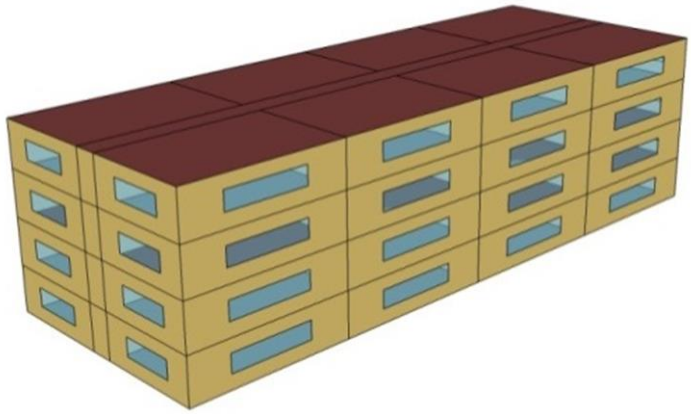


Figure VI-6 Prototype model of a midrise apartment building (Retrieved from DOE's commercial reference building models)

Test case 2 is an academic building located in Texas A&M University, College Station. The 2-story academic building was constructed in 2013 and has a gross floor area of 114,366 ft^2. The exterior finishes of the steel-framed building essentially consist of brick veneer and stone. The building is mainly used for educational and recreational

purposes. Figure VI-7 illustrates the exterior view of the case study building obtained from the street view in Google Maps. For EE information of the building, data was collected from the university architect's office. OE information regarding the building dimensions, WWR, orientation, etc. were gathered from Google Earth. For this case study, we assume that the wall and window U-values for the academic building are 0.066 Btu/hr.ft^2.^{o}F and 0.40 Btu/hr.ft^2.^{o}F, respectively. These assumptions were made in accordance with the ASHRAE 90.1-2018 prescriptive codes for climate zone 2A.

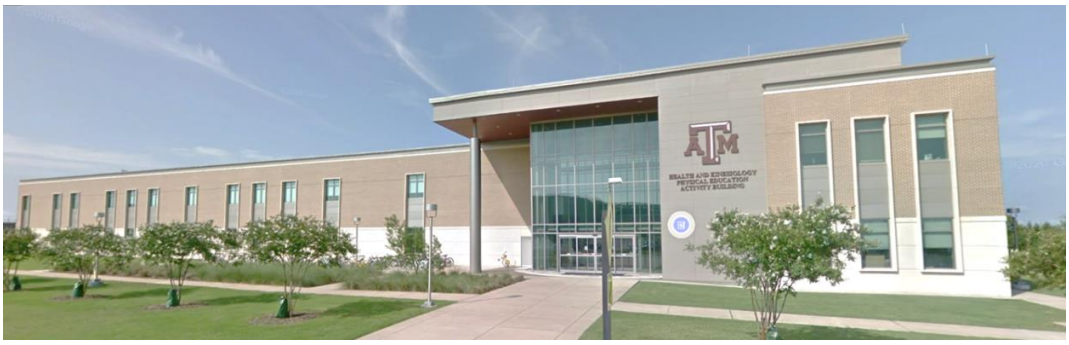


Figure VI-7 Exterior view of the case study building (Retrieved from Google Images)

The OE intensity and EE intensity of the test cases were evaluated using both the traditional and data-driven approaches. In the traditional approach, the geometry of the test cases was modeled on Rhino. Then the Ladybug and Honeybee plugins available in

the Grasshopper environment were used to simulate OE intensity. For EE, we first disaggregated the information obtained from the contractor's schedule of values into different material groups. After that, the IOH-disaggregated approach described in section 4.2.2.2 was utilized to calculate EE intensity.

In the data-driven approach, we deployed the ML model developed in this study to estimate building energy. Here, we provided information for the previously defined input fields. After entering the input parameters, the predict method was called to estimate OE intensity and EE intensity. Finally, we determined and compared the variation in results between both approaches.

CHAPTER VII

RESULTS

7.1 Overview of the synthetic dataset

In this research, we utilized the explorative nature of the parametric design technology to generate the synthetic building LCE dataset. Each design iteration of the prototype building was generated by varying input parameters within the constraints defined earlier. The OE simulations were performed using the EnergyPlus engine provided by the Ladybug and Honeybee plugins, while the EE assessment was conducted using the IOH-disaggregated approach on Microsoft Excel. All the input features along with their corresponding results were recorded for each design iteration (as seen in Table VII-1). This information was stored in a separate Microsoft Excel file which was later used to train and test the ANN model. We continued this iterative process of data generation until we obtained 6000 data points.

Figure VII-1 shows a small sample of the 3-dimensional models created using the parametric workflow. The relationship between the input and output parameters for all the design iterations in the dataset are graphically represented using a parallel coordinate plot, as seen in Figure VII-2. The floor height, number of floors, length, width, WWR-N, WWR-S, WWR-E, WWR-W, orientation, wall U-value, and window U-value are included as the input features, whereas the annual OE and EE intensities are recorded as output targets. To visualize the parallel coordinate plot better we included the sum of EE and OE as another parameter labeled LCE. These parameters are represented as a vertical

axis on the parallel coordinate plot. Each vertical axis on this plot is intersected by a polyline which represents the normalized parameter value in the design space.

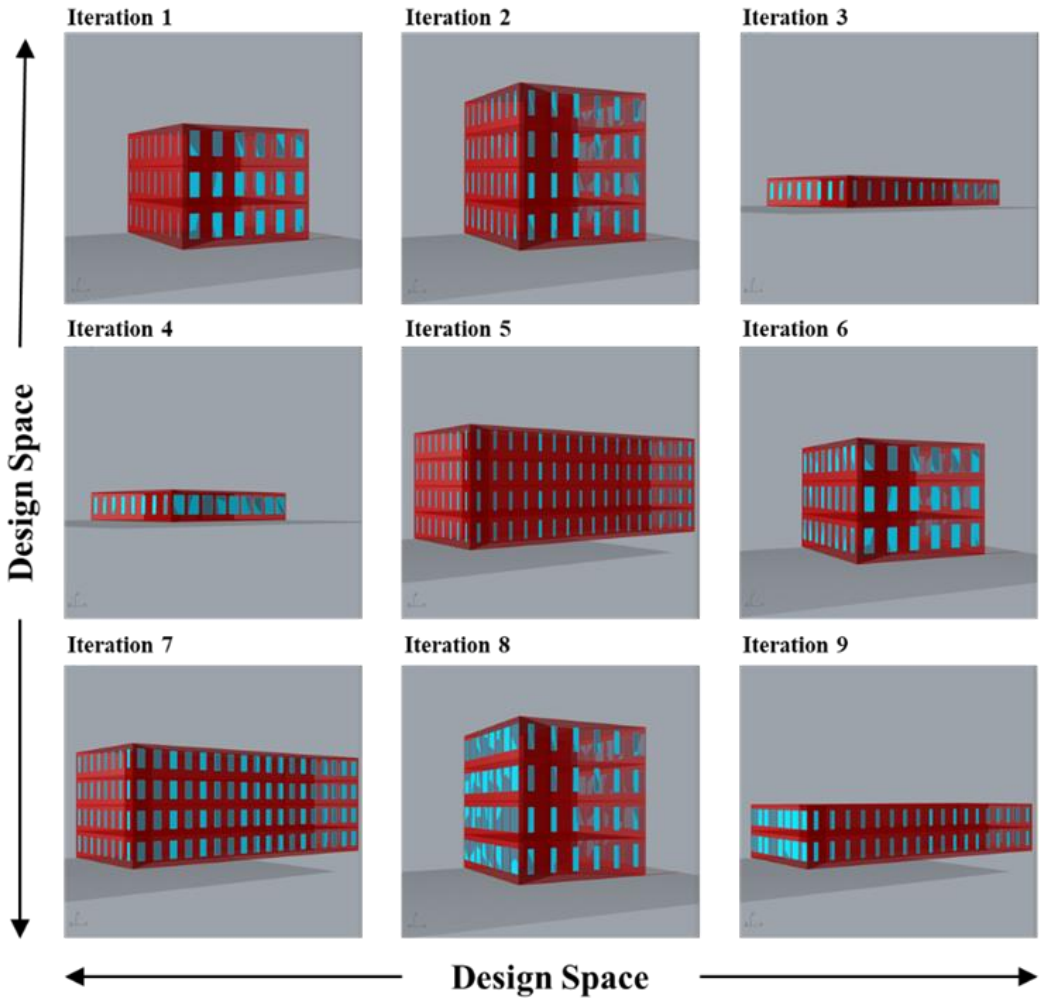


Figure VII-1 Sample of design iterations created using the parametric script

Table VII-1 Excerpt of the dataset generated using the parametric framework

Input parameters												Output parameters	
Climate zone	Floor height	Number of floors	Length	Width	WWR-N	WWR-W	WWR-S	WWR-E	Orientation	Wall U-value	Window U-value	OE (kBtu/ft^2.year)	EE (kBtu/ft^2.year)
2	13	1	249	194	0.26	0.84	0.11	0.16	297	0.051	0.35	57.6	31.48
5	15	1	117	118	0.88	0.3	0.02	0	260	0.051	0.35	55.32	33.62
2	14	1	106	176	0.38	0.59	0.55	0.13	172	0.052	0.4	62.17	31.75
2	15	3	170	157	0.14	0.07	0.52	0.16	90	0.047	0.35	54.78	33.13
2	11	2	253	176	0.85	0.1	0.54	0.15	201	0.052	0.4	54.52	30.5
5	9	1	222	127	0.8	0.46	0.69	0.06	107	0.044	0.4	50.69	31.39
5	9	3	170	231	0.22	0.72	0.24	0.35	52	0.039	0.35	41.83	30.55
5	12	1	170	210	0.85	0.23	0.51	0.6	347	0.036	0.4	52.47	34.22
2	13	1	89	35	0.23	0.45	0.8	0.24	270	0.066	0.4	76.3	38.94
2	13	1	63	197	0.14	0.52	0.3	0.83	346	0.075	0.55	70.52	31.48
2	13	2	76	195	0.55	0.32	0.31	0.59	73	0.047	0.35	59.87	35.01
2	11	3	250	141	0.33	0.41	0.79	0.35	207	0.047	0.55	55.74	30.11
5	9	3	249	85	0.73	0.7	0.94	0.07	63	0.036	0.35	45.22	35.13
2	11	2	108	124	0.49	0.01	0.11	0.78	161	0.066	0.35	57.43	33.31
2	15	1	85	116	0.84	0.38	0.42	0.77	1	0.075	0.4	70.67	32.74
2	11	2	95	228	0.84	0.8	0.04	0.15	194	0.051	0.4	58.16	30.3
2	14	3	160	116	0.58	0	0.25	0.02	161	0.052	0.35	54.9	33.77
5	14	2	127	36	0.93	0.84	0.43	0.76	181	0.047	0.4	73.31	36.17
5	15	3	178	112	0.81	0.48	0.59	0.77	120	0.044	0.35	51.57	34.1
5	11	1	199	153	0.24	0.53	0.79	0.78	229	0.047	0.35	52.82	32.92
5	10	1	55	221	0.05	0.36	0.43	0.47	16	0.044	0.35	55.24	35.21
2	15	2	52	115	0.52	0.68	0.27	0.19	196	0.066	0.35	69.55	39.26
2	9	1	118	194	0.45	0.92	0.76	0.04	194	0.066	0.35	59.19	32.6
5	10	1	213	163	0.82	0.2	0.81	0.08	209	0.039	0.4	46.76	31.23
5	10	3	167	61	0.67	0.54	0.77	0.06	167	0.047	0.4	50.81	31.97
2	11	3	241	76	0.38	0.83	0.03	0.22	111	0.051	0.4	59.17	30.29
2	12	2	47	201	0.76	0.45	0.03	0.52	163	0.066	0.4	66.23	34
2	15	3	68	63	0.62	0.34	0.63	0.3	320	0.052	0.55	72.7	35.56

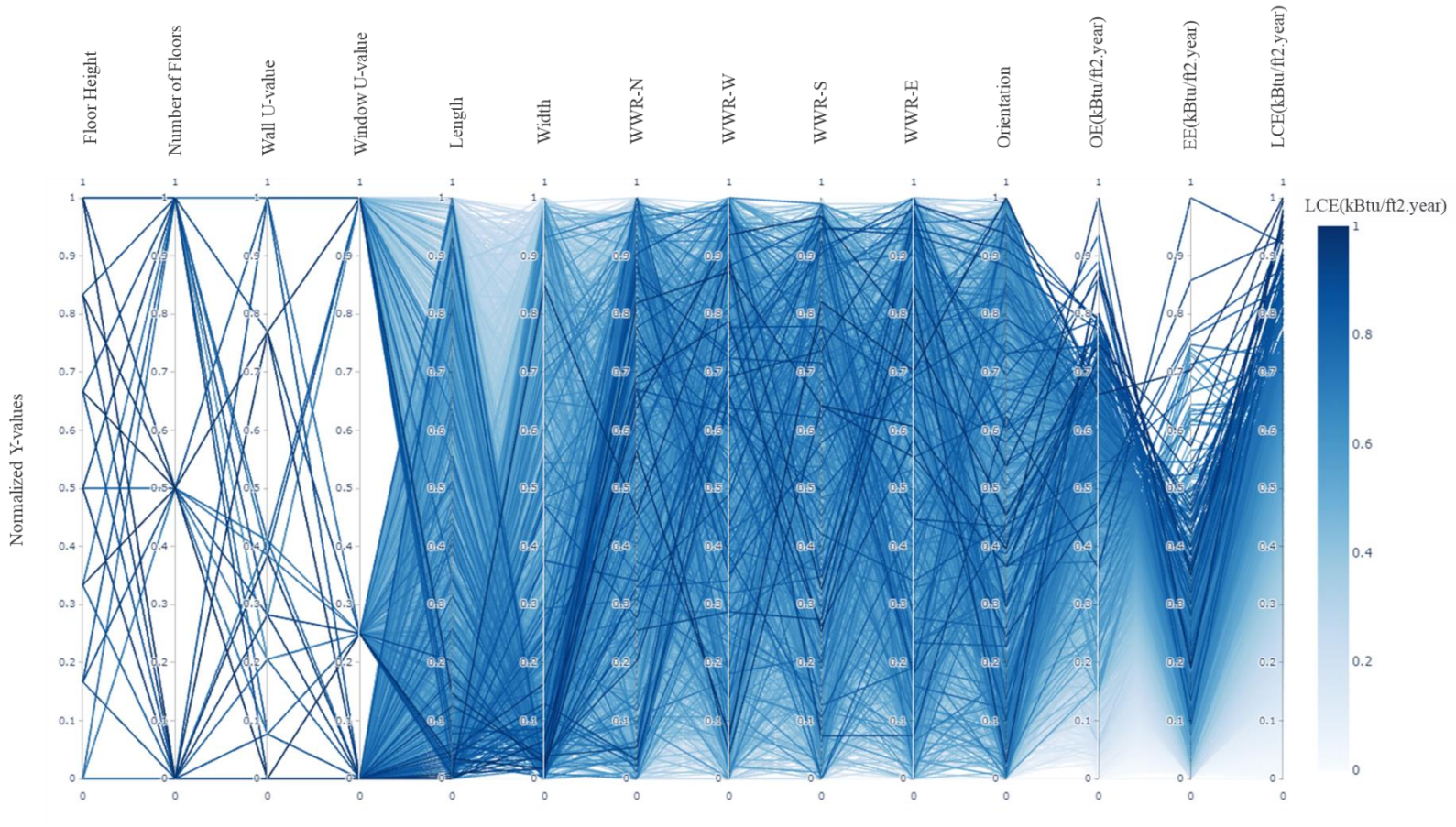


Figure VII-2 Parallel coordinate plot of the entire building life cycle energy dataset

7.2 ANN model evaluation

7.2.1 Basic statistical analysis of the synthetic dataset

Initially, an exploratory data analysis was performed on the synthetic dataset to identify anomalies in the data. The dataset consists of 12 independent variables and 2 dependent variables. Table VII-2 shows the descriptive statistics of the synthetically generated building energy dataset. These statistics were obtained using the describe() function in the Pandas library. This function describes and summarizes the mean, standard deviation, minimum value (min), 25%tile, 50%tile (median), 75%tile, and maximum value (max). The measure of central tendency and dispersion of the sample data is represented using the mean and standard deviation, respectively.

Table VII-2 Descriptive statistics of the synthetic dataset

Variable	Mean	Standard Deviation	Min	25%	50%	75%	Max
Floor height	11.93	1.99	9.00	10.00	12.00	14.00	15.00
Number of floors	2.12	0.82	1.00	1.00	2.00	3.00	3.00
Length	144.45	66.71	30.00	87.00	142.00	202.00	262.00
Width	143.26	66.49	30.00	86.00	142.00	201.00	262.00
WWR-N	0.48	0.27	0.00	0.25	0.48	0.71	0.94
WWR-W	0.47	0.27	0.00	0.23	0.47	0.70	0.94
WWR-S	0.48	0.27	0.00	0.25	0.48	0.70	0.95
WWR-E	0.47	0.27	0.00	0.23	0.47	0.70	0.94
Orientation	177.53	102.30	0.00	89.00	176.00	265.00	359.00
Wall U-value	0.05	0.01	0.04	0.04	0.05	0.05	0.08
Window U-value	0.40	0.07	0.35	0.35	0.40	0.40	0.55
OE (kBtu/ft^2.year)	56.86	7.99	41.44	51.37	56.61	61.26	97.38
EE (kBtu/ft^2.year)	33.49	3.73	28.80	31.08	32.50	34.73	68.36

The kernel distribution estimate plot is seen in Figure VII-3 and Figure VII-4 shows the probability density of the OE intensity and EE intensity, respectively. The y-axis in both the figures is used to represent the kernel density estimate of the probability density function, whereas the x-axis is used to represent the energy intensity. We notice that the shape of both the output target curves are (approx.) normally distributed for each climate zone.

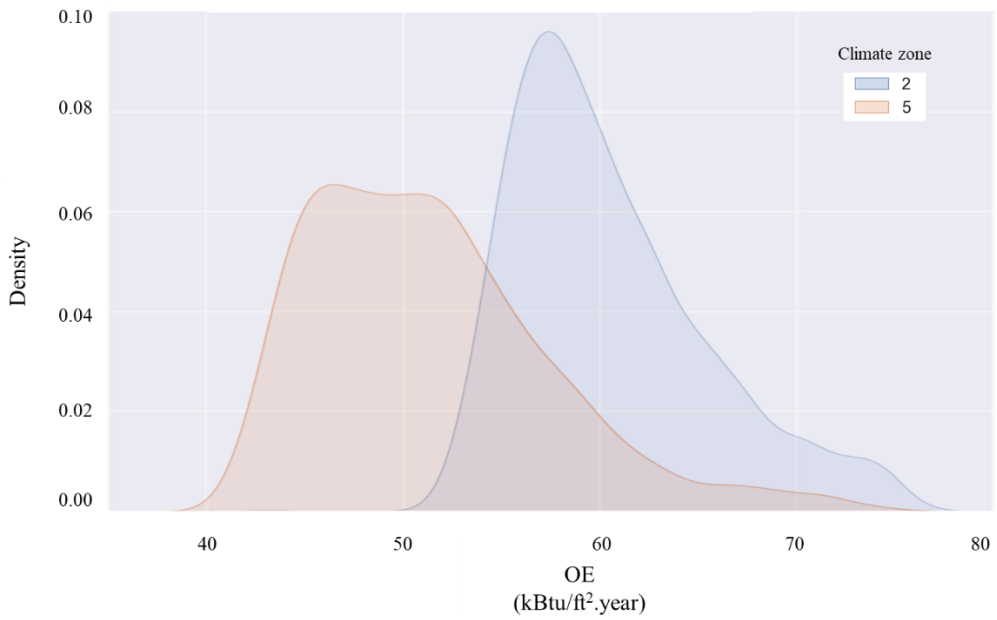


Figure VII-3 Kernel density estimate plot of OE intensity

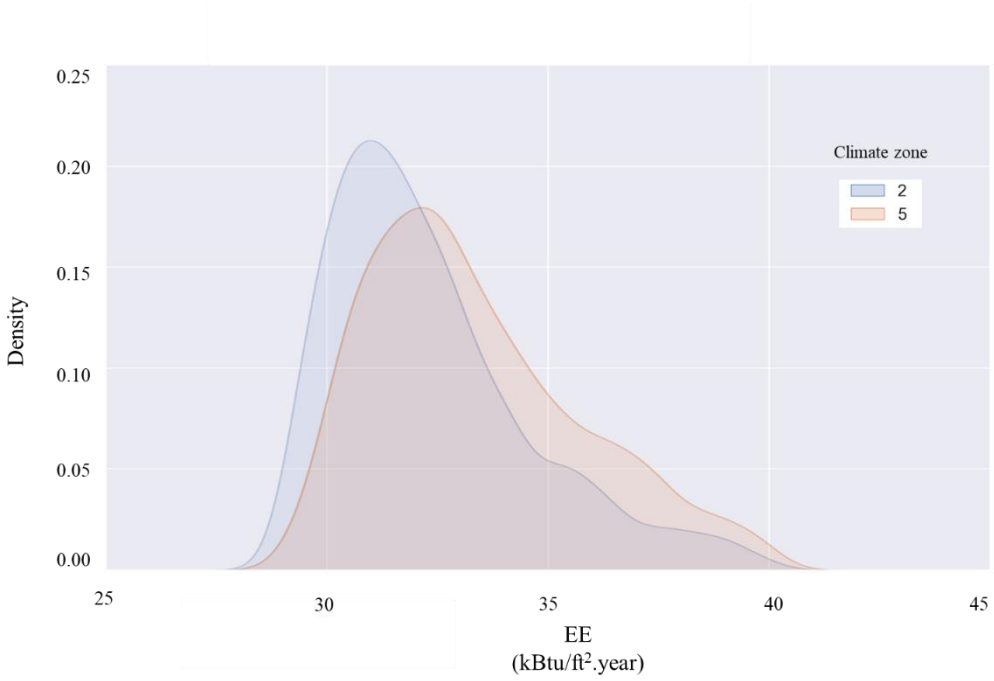


Figure VII-4 Kernel density estimate plot of EE intensity

7.2.2 Feature selection

In this study, we utilize the correlation matrix for feature selection. Figure VII-5 shows the correlation of various input parameters with each other and with the output parameters, in the building energy dataset. The plotted matrix has 14 rows and 14 columns. Each row and column in the figure represent a parameter, while the value within the grid is the correlation coefficient. The Pearson correlation coefficient *'ρ'* was used to determine whether the input parameters are associated with the output parameters. The value of ρ ranges between -1 and +1. A perfectly positive correlation is indicated by an ρ of +1, while a perfectly negative correlation is denoted by -1. A ρ value of 0 indicates that

there is no correlation between the variables. In the figure below, the diagonal cells show the correlation of each variable with itself.

The heatmap indicates that the input features such as length and width have a moderate positive correlation with the EE intensity and OE intensity ($0.5<\rho<0.7$). This indicates that the energy consumption increases if the square footage of the building increases. The number of floors parameter has a low positive correlation ($0.3<\rho<0.5$), while the remaining variables except climate zone have negligible correlation ($0.0<\rho<0.3$) with EE intensity and OE intensity. The climate zone is the only feature that has a negative correlation with the outputs. This is mainly because the climate zone feature is included as a nominal categorical data attribute in this correlation matrix (i.e., climate zones cannot be ordered based on their label). To further explain, climate zone '2' does not occur before climate zone '5' nor is it considered bigger or smaller than '5'. Therefore, it would be incorrect to order the values of the climate zone data attribute.

The correlation matrix does not indicate any other strong correlations between the independent parameters that could potentially lead to the issue of multicollinearity. The combination of this interpretation along with our domain knowledge on the subject matter obtained from the previous study (Venkatraj et al., 2020) dissuaded us from dropping any feature.

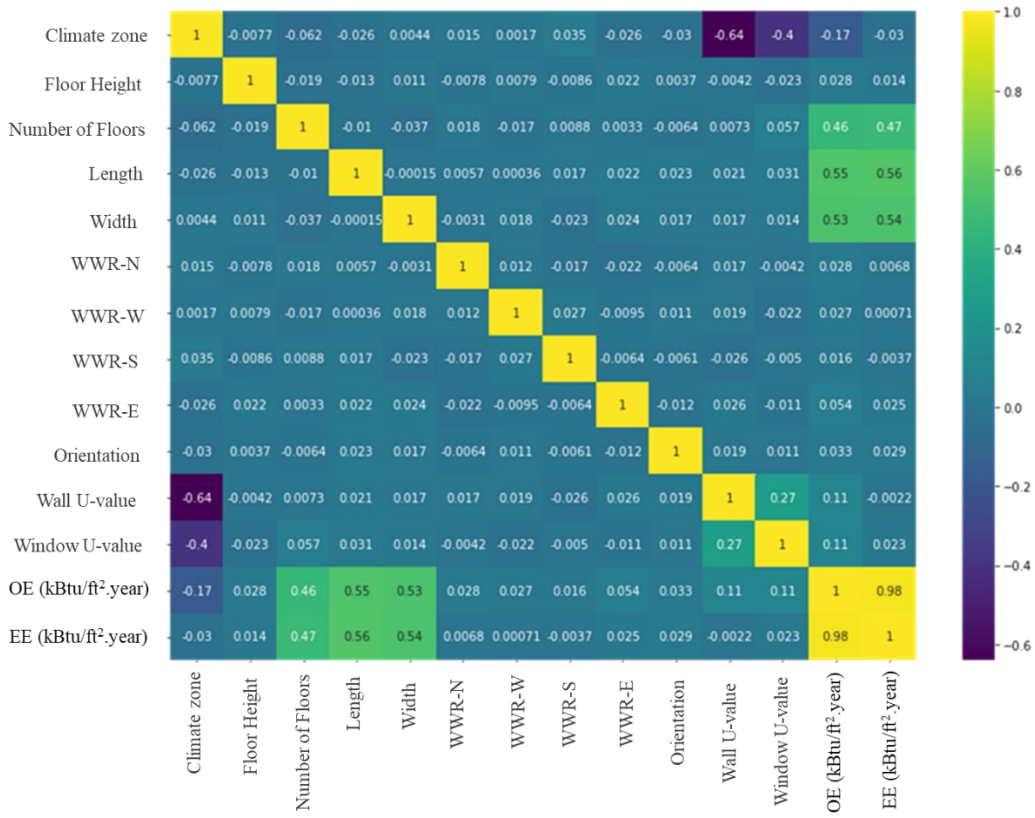


Figure VII-5 Correlation matrix of the building energy dataset

The Pearson correlation coefficient only shows a linear correlation between different variables, therefore, implying that the input parameters may have a non-linear relationship with the output parameters. This implies that a linear regression model may not accurately capture the complex relationships between different building parameters and may, therefore, fail to accurately predict building energy. As a result, we were persuaded to utilize an ANN algorithm that has the capability of modeling non-linear relationships to develop the prediction model.

7.2.3 Experiments to determine optimum network architecture

As highlighted earlier, there is no generic rule to determine the optimum network architecture. Existing methods are mostly ad-hoc or based on trial and error. Therefore, we conducted a series of systematic experiments to determine the most suitable network architecture for our specific problem. These experiments were focused on evaluating the performance of the model by changing the number of nodes in each hidden layer and the number of hidden layers itself (depth of the model).

Table VII-3 shows the list of experiments that were conducted with 1-, 2-, and 3-hidden layers. The experimentation process began with one hidden layer which had 7 nodes. For each experiment, we incremented the number of nodes in the hidden layer by one and evaluated the performance of the model using metrics such as MAPE and R^2. MAPE is a measure of prediction accuracy, whereas R^2 determines the goodness of fit (predicted values fit closely to the true values). While using a single hidden layer, the network architecture of the best performing model had 10 hidden layer nodes. Upon adding another hidden layer, the performance of the model further improves. The highlighted row in

Table VII-3 (Experiment 18), shows the architecture of the model with the best MAPE and R^2 performance.

Table VII-3 Experiments to determine optimum network architecture

Experiment	Model architecture	Output 1 (OE intensity)		Output 2 (EE intensity)		Epochs
		MAPE (%)	R^2	MAPE (%)	R^2	
1	12, 7, 2	3.64	0.86	3.83	0.73	175
2	12, 8, 2	3.92	0.85	4.22	0.64	89
3	12, 9, 2	3.38	0.84	4.91	0.58	113
4	12, 10, 2	2.82	0.89	3.43	0.75	61
5	12, 11, 2	3.87	0.86	4.45	0.64	50
6	12, 12, 2	2.80	0.91	4.36	0.58	54
7	12, 13, 2	3.68	0.86	3.97	0.69	225
8	12, 14, 2	3.71	0.86	4.19	0.65	84
9	12, 15, 2	3.91	0.85	4.19	0.63	63
10	12, 9, 5, 2	2.92	0.89	4.73	0.58	90
11	12, 9, 7, 2	1.73	0.96	3.85	0.73	134
12	12, 9, 9, 2	2.37	0.94	4.14	0.70	206
13	12, 11, 5, 2	2.50	0.93	2.87	0.83	116
14	12, 11, 7, 2	1.82	0.95	2.55	0.87	131
15	12, 11, 9, 2	2.95	0.90	4.30	0.68	58
16	12, 14, 5, 2	1.48	0.97	5.05	0.52	156
17	12, 14, 7, 2	2.75	0.90	3.89	0.58	48
18	**12, 17, 10, 2**	**1.95**	**0.96**	**2.47**	**0.90**	**113**
19	12, 9, 6, 3, 2	2.57	0.93	5.43	0.50	195
20	12, 10, 5, 10, 2	1.59	0.97	3.72	0.76	54
21	12, 10, 15, 10, 2	1.67	0.96	3.49	0.78	100
22	12, 12, 15, 10, 2	1.71	0.96	3.20	0.74	88
23	12, 12, 15, 12, 2	2.21	0.95	3.29	0.72	80
24	12, 14, 7, 10, 2	1.88	0.96	3.11	0.71	70
25	12, 14, 10, 7, 2	1.77	0.96	2.84	0.84	107

The early stopping callback was utilized in this study to stop the training process instead of using a fixed number of epochs as the stopping criterion. This callback will prevent overfitting by terminating the training process when the performance of the model does not improve for 20 consecutive epochs. Therefore, the number of epochs varies for each experiment (as seen in

Table VII-3). In addition, this study also conducted a few experiments by tweaking other hyperparameters such as the activation function, batch size, and learning rate. However, we noticed that these adjustments did not improve the performance of the model.

7.2.4 Performance evaluation of the developed multi-output ANN model

The final network architecture of the ML model developed in this study is illustrated in Figure VII-6. This model has 12 nodes in the input layer, 17 nodes in the first hidden layer, 10 nodes in the second hidden layer, and 2 nodes in the output layer. The model summary shows that it consists of three dense layers (i.e., nodes of the dense layer are connected to all the nodes of the preceding layer). Figure VII-7 shows the output shape and number of parameters (weights) in each layer. Each dense layer transforms the input it receives based on the network's weights and biases. In total, the model has 423 trainable parameters. Table VII-4 shows the setting of the model hyperparameters used during the training process.

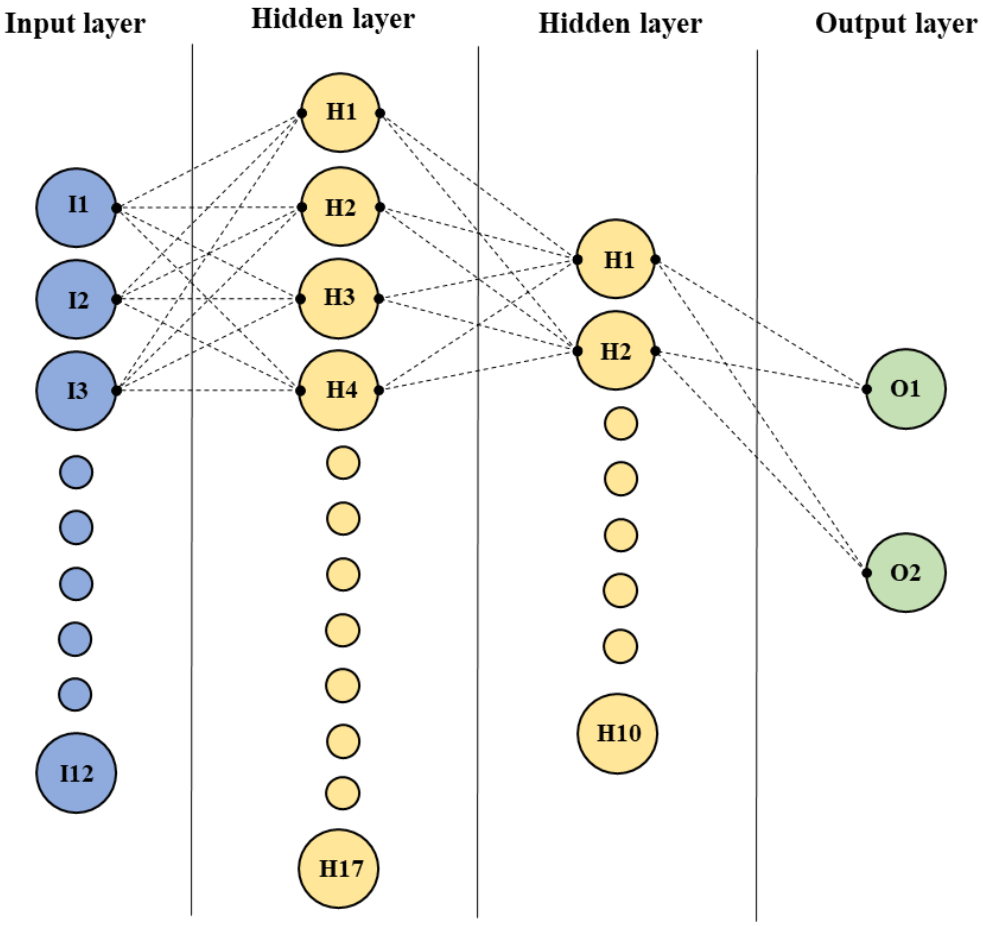


Figure VII-6 Network architecture of the energy prediction model

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 17)	221
dense_1 (Dense)	(None, 10)	180
dense_2 (Dense)	(None, 2)	22

Total params: 423
Trainable params: 423
Non-trainable params: 0

Figure VII-7 Model summary

Table VII-4 Summary of hyperparameters used for training

Model hyperparameters	Setting
Batch size	32
Activation function	ReLu
Optimizer	Adam
Learning rate	0.01
Patience criteria for early stopping	20 epochs

Figure VII-8 shows the history of the model's learning performance over 113 epochs. During an epoch, the model evaluates its performance on the training dataset and validation dataset using the loss function. This measured performance (loss/cost) of the model is plotted in the form of a learning curve. The loss curve provides crucial information regarding the fit of the model. If the validation loss decreases to a point and then begins increasing again, then the model is overfitting i.e., it will not generalize well. In the opposite scenario, if the model is not trained for sufficient epochs it will not learn from the data and is said to have the problem of underfitting. During the first epoch, we notice that the model loss (MSE) is extremely high due to random weight initializations. This loss progressively decreases during the training process which adjusts the weights of the network. Figure VII-8 shows that the training process continued for 113 epochs. We observe that the loss curve decreases to a point of stability during which the training process is stopped by using the early stopping callback (Figure VII-8). At the end of the training process, the overall loss was 0.028. The MAE for output 1 (OE intensity) was 0.92, whereas the MAE for output 2 (EE intensity) was 0.90.

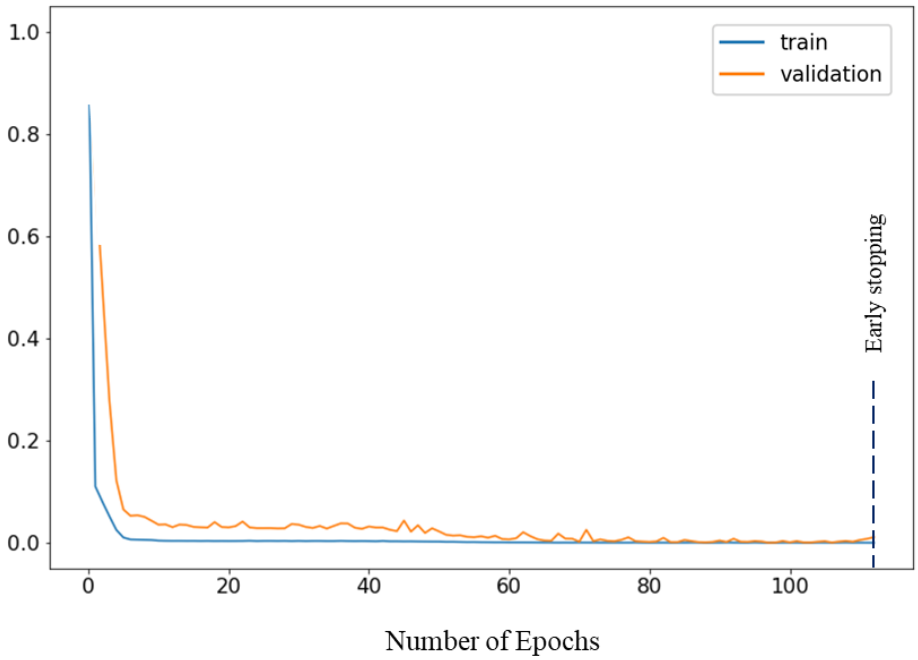


Figure VII-8 Training loss curve

Figure VII-9 illustrates the correlation analysis between the actual energy intensity obtained using the parametric workflow and the model output. We notice that the predicted values (represented by the points) are very close to the regressed diagonal line for both the outputs (OE intensity and EE intensity). This demonstrates a very strong correlation between the actual and predicted values. The R^2 values are also very close to 1, suggesting that the model has a good fit. It is important to note that R^2 values do not account for bias. For that reason, we also examine the residual plots to ensure that our model predictions are not biased. The x-axis and y-axis of Figure VII-10 represent the predicted values and standardized residual values, respectively. The residual value is essentially calculated as

the difference between the actual value and the predicted value for each data point. Positive residual values indicate that the prediction was low, whereas negative values indicate that the prediction was too high. Residual values of zero represent a perfect prediction. We do not notice any specific patterns or residual outlier values in the residual plot shown in Figure VII-10. In addition, we observe that the residual values are clustered and evenly distributed around the lower single digits of the y-axis. Therefore, indicating that the developed model does indeed fit the data well. However, in Figure VII-10b we notice that the plot exhibits some characteristics of heteroscedasticity. This means that the variance in the residual values is unequally distributed. From the plot we notice that the residual value increases as we progress from lower to higher values of EE intensity, indicating that further improvements can be made to the developed model. This heteroscedasticity may arise due to the right skewness of the EE intensity data that we saw earlier in Figure VII-4. Since the model is otherwise performing well, we accept the current model as is.

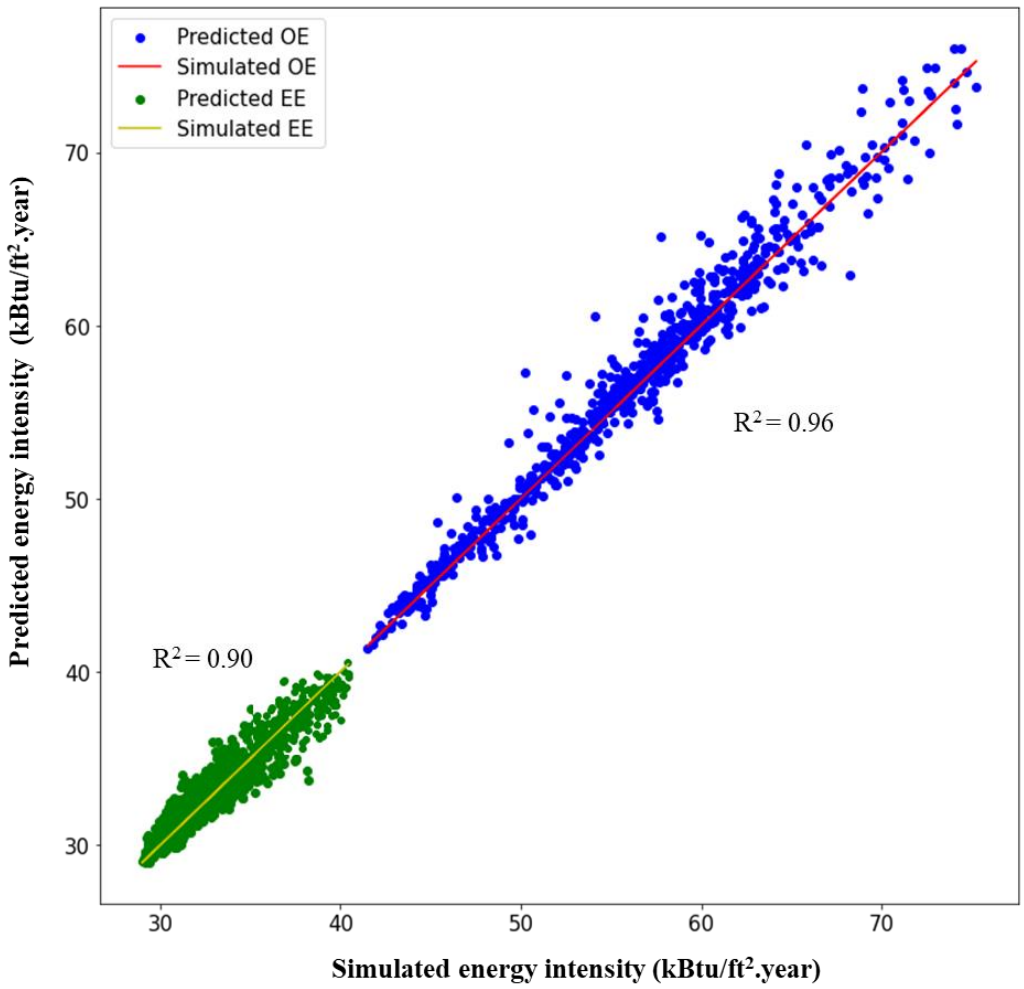
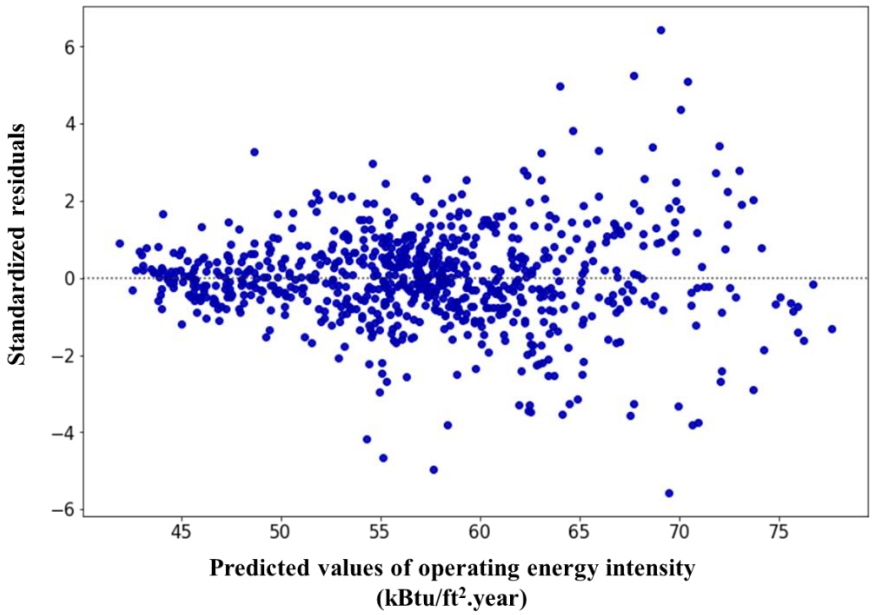
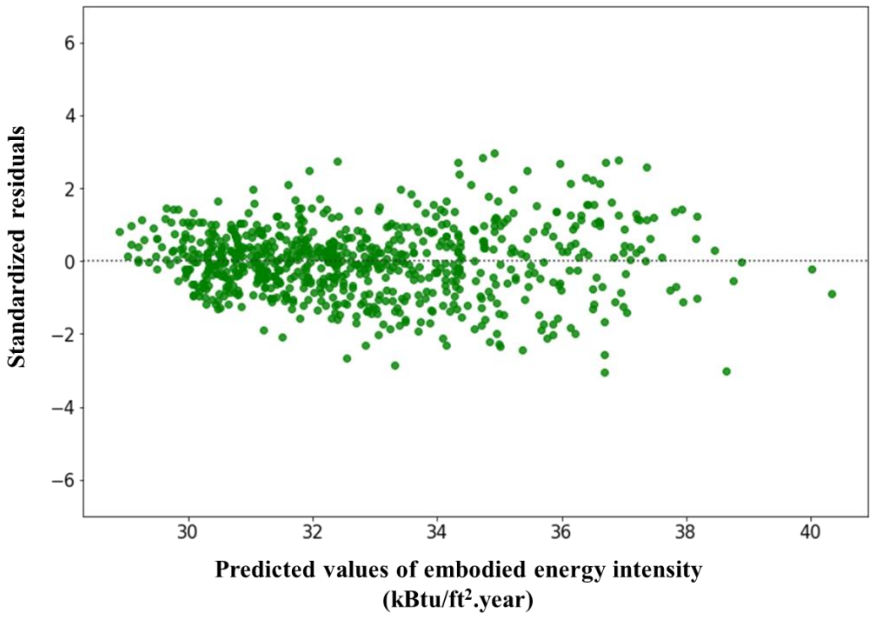


Figure VII-9 Correlation between actual and predicted energy intensity



(a)



(b)

Figure VII-10 Residual plot for predicted values of (a) operating energy intensity, and (b) embodied energy intensity

Figure VII-11 shows the comparison between the predicted values and true (simulated) values of OE intensity and EE intensity for all the data points in the test dataset. In Figure VII-11, the data point is plotted on the x-axis while the model predictions and true values of energy intensities are plotted on the y-axis. Each point in the scatter plot represents the predicted value, whereas the continuous line represents the true value in Figure VII-11(a). A more detailed version of the same plot is illustrated in Figure VII-11(b). Here, the true values are represented using continuous lines, whereas the predicted values are shown using dashed lines. This graph along with Table VII-5 shows that the over or underestimated values of energy intensities are negligible and very close to the true value. Therefore, indicating that the developed ML model has high predictive accuracy.

Table VII-5 Predicted vs True values for a small sample of the testing dataset

OE intensity (kBtu/ft^2.year)			EE intensity (kBtu/ft^2.year)		
Simulated	**Predicted**	**Residual**	**Simulated**	**Predicted**	**Residual**
55.85	56.44	-0.59	38.99	38.87	0.12
63.29	60.93	2.36	34.23	34.34	-0.11
48.38	48.53	-0.15	33.16	33.18	-0.02
55.61	54.59	1.02	33.53	32.29	1.24
62.17	59.51	2.66	31.63	32.77	-1.14

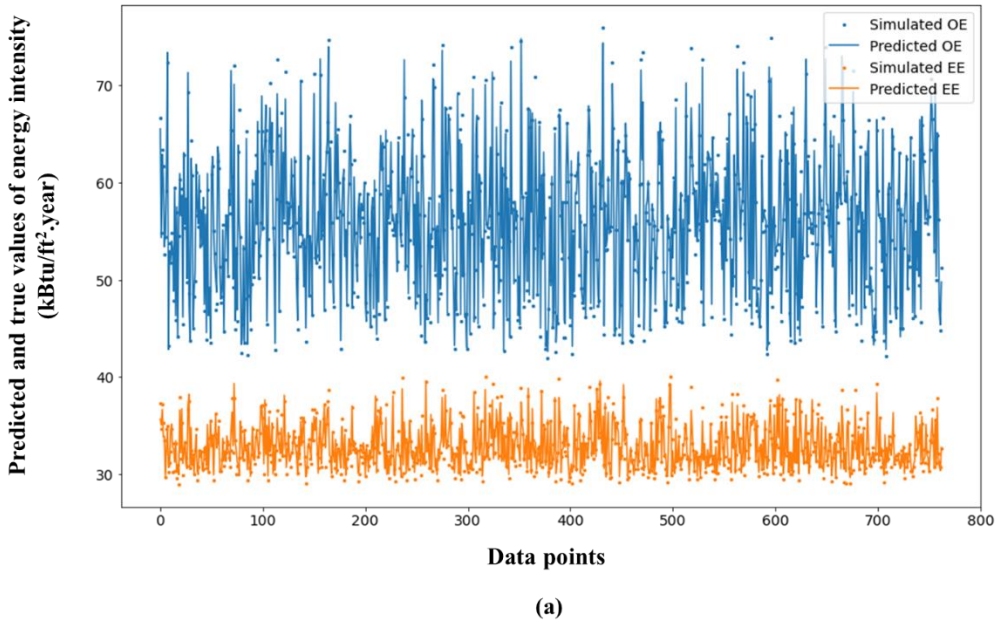


Figure VII-11 Scatter plot comparison between predicted and simulated values of energy intensities (a) for all the data points in the testing dataset, and (b) small sample of the testing dataset

7.3 Test case and results

Test case 1 and 2 were modelled in the Rhino/Grasshopper environment (as seen in Figure VII-12 and Figure VII-13). The OE intensity for both the case study buildings were simulated using the Ladybug and Honeybee plugins within the GH environment. These plugins utilize the EnergyPlus which is considered the most robust energy simulation engine in the building industry.

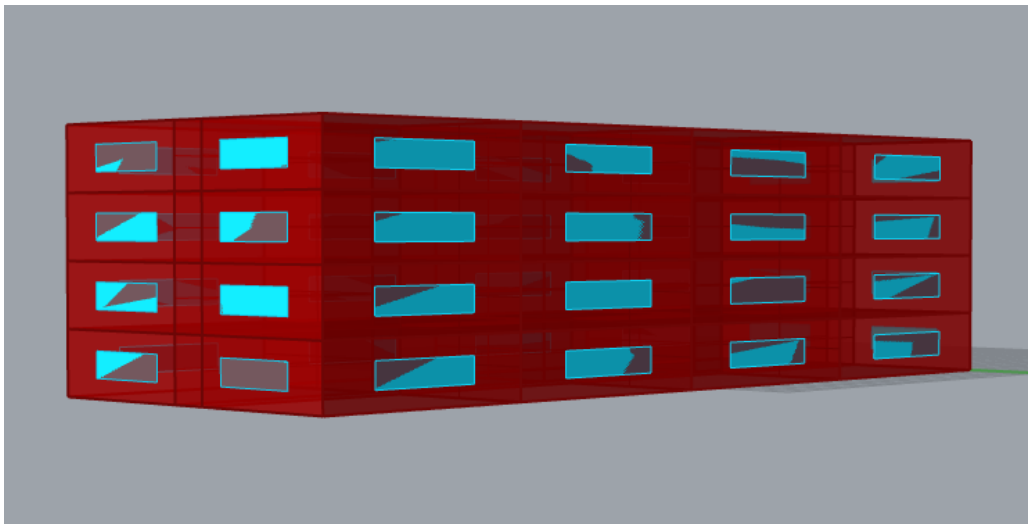


Figure VII-12 Energy model of test case 1 located in climate zone 5A

For test case 1 located in climate zone 5A, we calculated the EE using the surface area information of different building components obtained from the prototype model. The

simulation results of OE and EE intensity were 44.02 kBtu/ft^2.year and 31.02 kBtu/ft^2.year, respectively. The input parameter values shown in Table VII-6 were entered into the developed ML model to predict the OE and EE intensities of Building 1. The prediction results of OE and EE intensities were 45.92 kBtu/ft^2.year and 31.70 kBtu/ft^2.year, respectively. The percentage error between the simulated and predicted value of OE and EE intensities are 4.13% and 2.14%, respectively.

Table VII-6 Characteristics of case study building in climate zone 5A

Input Parameter	Value
Location	5A
Length	152ft
Width	55ft
Number of floors	4
Floor height	10ft
Wall U-value	0.051 Btu/hr.ft^2.^{o}F
WWR-N	0.20
WWR-W	0.20
WWR-S	0.20
WWR-E	0.20
Window U-value	0.40 Btu/hr.ft^2.^{o}F
Orientation	0 ^{o}

For test case 2 located in climate zone 2A, the simulation output of OE intensity was 52.40 kBtu/ft^2.year. The contractor's schedule of values was used to calculate EE intensity using the IOH-disaggregated approach. The EE intensity calculated in this manner was 28.63 kBtu/ft^2.year.

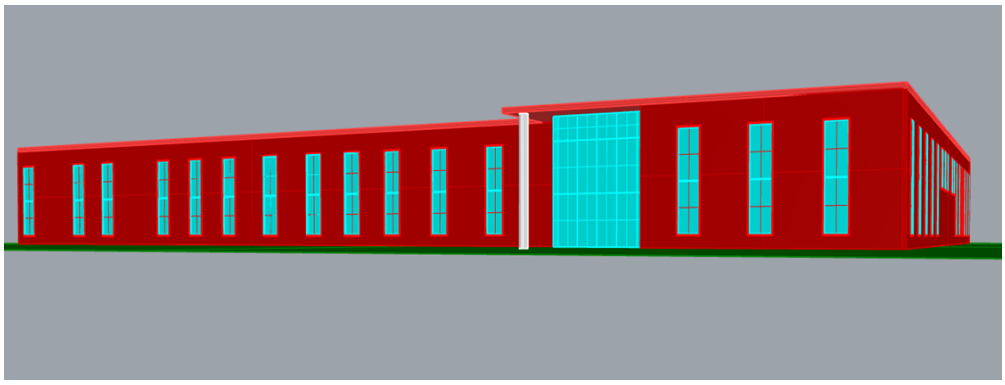


Figure VII-13 Energy model of building 2 located in climate zone 2A

Next, we entered the building parameter values (Table VII-7) into the developed ML model to predict its energy intensities. The predicted values of OE and EE intensities were 53.51 kBtu/ft^2.year and 30.78 kBtu/ft^2.year, respectively. The percentage error between the simulated output and predicted output is 2.07% and 6.9% for OE and EE intensity, respectively.

Table VII-7 Characteristics of test case 2 in climate zone 2A

Input Parameter	Value
Location	2A
Length	255ft
Width	360ft
Number of floors	2
Floor height	15ft
Wall U-value	0.066 Btu/hr.ft^2.^{o}F
WWR-N	0.500

Continued

Input Parameter	Value
WWR-W	0.010
WWR-S	0.276
WWR-E	0.15
Window U-value	0.40 Btu/hr.ft^2.^{o}F
Orientation	315^o

CHAPTER VIII

DISCUSSION*

8.1 Overview

In this study, we developed an ML model for building energy prediction by comprehensively integrating building LCE components. The findings of this study show that ML models can indeed simplify the process of building LCE analysis. More specifically, ANN models are capable of handling complex nonlinear relationships between multiple building parameters. This study also has significant time and cost implications. The machine learning approach drastically reduces the amount of time and expertise required for modeling. The computer used to generate the parametric building energy dataset and develop the ML model has an Intel® Core™ i9-9980HK CPU @ 2.4GHz processor and 64 GB RAM. We are all aware that running a single building energy simulation takes several hours to a day and is computationally expensive. After setting up the parametric workflow, each run took approximately 60-90 seconds for the 1-storey and 2-storey building models whereas, the computational time taken for the 3-storey building models nearly doubled (i.e., ~120-180 seconds). The time taken to model the 3D geometry of the case study building on Rhino was approximately 3-4 hours. After which, the energy simulation took 2-3 minutes to run. As already mentioned, the traditional energy

* Part of this chapter is reprinted with permission from:

"Challenges in implementing data-driven approaches for building life cycle energy assessment: A review" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 160, 112327, Copyright (2022) by Elsevier

simulation workflows are plagued by issues of interoperability. Therefore, the EE calculations for both the studies were conducted on Microsoft Excel using the IOH-disaggregated approach. These EE calculations took an additional 10 minutes to compute. Our study overcomes these challenges by utilizing the ML approach that provides instantaneous results. The computational time was approximately 1 microsecond for each step during the training process. The developed model requires very few numerical inputs to perform energy assessments that are almost as accurate as the traditional simulation results. It is important to mention that while it took approximately 12-13 minutes to simulate the OE and EE results of the case study building, the developed ANN provided results in less than 1 second. Undoubtedly, the speed of the ML approach is advantageous for generative design and optimization where millions of design options need to be simulated.

Most of the issues related to ML are associated with the lack of data. Most of the existing open source/publicly available databases contain information regarding the building type, construction materials, floor area, climate zone, energy sources, electricity consumption, etc., However, data required for EE calculations such as bill of quantities, the service life of different building systems, system boundary definitions, etc., is missing. Our findings suggest that the robustness of the ML model can be improved by adding more diversity to the data. Expanding the database to include energy and material information consumed during the entire building life cycle for different building types, shapes, materials, locations, etc. will make it more diverse and robust. Techniques of parametric modeling and data augmentation can be used to generate synthetic data which

would expand the database. These findings are in corroboration with studies conducted by Singaravel et al. (2018), Wang et al. (2021), D'Amico et al. (2019). Furthermore, we also recommend combining real-world data with synthetic data to create a holistic dataset.

Based on the set of experiments we conducted to identify optimum network architecture, we observe that the model performance does not increase significantly by adding more nodes or depth to the network (Table VII-3). This may be due to the limited size of our synthetic dataset. Moreover, Table VII-3 also shows that the model's MAPE and R^2 performance for OE intensity are much better in comparison to EE intensity. This can be explained by the right skewness and variation in EE data (Figure VII-4) which implies that most of the data points are distributed to the left side. The density of data on the left side may be much higher for both EE and OE intensities since we used the Galapagos (an optimization-based plug-in) to generate data. In the future, studies may consider using other simulation approaches to generate a more balanced dataset. We also conducted ML trainings using the same network architecture to understand the performance difference between a (i) single output prediction model (separate trainings for OE and EE) and (ii) multi-output prediction model (combined training for OE and EE).

Table VIII-1 shows that the MAPE values for the single output model is slightly higher in comparison to the multi-output model. Similarly, Figure VIII-1 illustrates that the R^2 value of the single output model is lower in comparison to the model that simultaneously trained both outputs. These results indicate that the combined ML trainings leverage the correlation between the two prediction tasks to improve the model performance. Multi-output prediction models are, therefore, beneficial in terms of training

time and predictive performance. Training disjoint models for each output can become computationally expensive and time-consuming, particularly for models that utilize large datasets or have complex network architecture.

Table VIII-1 Single output vs multi-output prediction model

	Single output prediction model			Multi-output prediction model		
	Epochs	**MAPE (%)**	**R^2**	**Epochs**	**MAPE (%)**	**R^2**
OE intensity	63	2.01	0.94	113	1.95	0.96
EE intensity	141	2.80	0.86		2.47	0.90

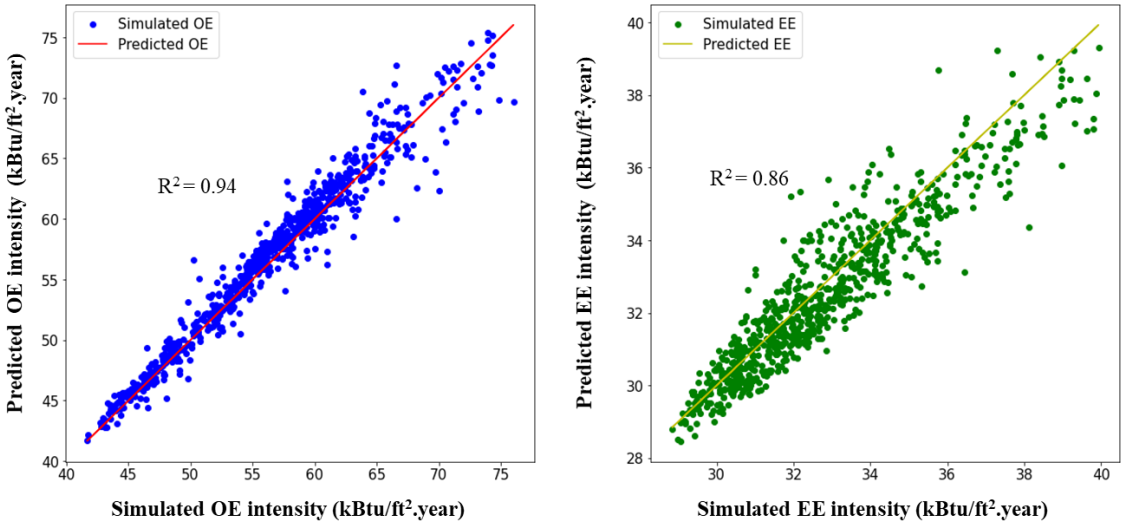


Figure VIII-1 Correlation between simulated and predicted energy intensity for single output prediction model

8.2 Generalization ability of the developed model

Generalization is defined as the ability of the ML model to handle unseen data. The performance of the ML model is quantified using the goodness of fit – how well the approximation matches the target function. When we train a model too well on training data it leads to the problem of overfitting, whereas if the model is not trained enough, it may cause the problem of underfitting. Ideally, the performance of the ML model on unseen data should be similar to its performance on the training dataset. In most cases, we assume that the distributions for the training and test data are the same. However, in real-world scenarios, the ML model is deployed in areas that are very different from the distribution of the original training dataset. Most often, if the data collected in the training dataset is not an accurate representation of the population, the ML model will not perform/generalize well. We know that most datasets used for training are biased since some parts of the population are either missing or under-represented.

To reduce sample bias in this study, we carefully selected parameter constraints to ensure that the model space is large enough to represent most of the building stock and derive energy consumption patterns. Despite these considerations, the generated sample dataset is still missing certain parts of the population. Epistemic uncertainty arises in the regions of the population that lack data. In our case, it refers to input parameters that are outside the range of our parametric model. For example, we do not have data for instances of the parametric model that have length or width values of more than 265 ft (80 m). In such cases where the input parameter value is outside the range of the training dataset, the ANN model will use extrapolation to determine the output value. Extrapolation is defined

as the ‘prediction from a model that is a projection, extension, or expansion of an estimated model outside the range of data that was used for model calibration’.

Generally, an ANN model does not perform well in regions that have no training data and are most likely associated with the high extrapolation error costs (Bartley et al., 2019). For instance, in test case 1, the number of floors parameter is outside the training range. In addition, we assume that test case 1 has core and perimeter zoning while evaluating the OE intensity using the prediction model. Here, the percentage error for OE and EE intensities are 4.14% and 2.14%, respectively. Similarly, in test case 2, one input parameter (width = 360 ft) was out of the range of data used for training the model. Moreover, the shape of the case study building is not a perfect rectangle and therefore, certain assumptions were made to the input parameters of the prediction model. Upon comparing the simulation results with the model outputs, we notice that the percentage error for the case study building was estimated to be 2.07% for OE intensity and 6.90% for EE intensity. In both the test cases, the error costs are less than 10%, which is recommended for highly accurate forecasting (Moreno et al., 2013). However, we observe that these error costs are much higher than the MAPE values of the model for OE and EE intensities, which were 1.95% and 2.47%, respectively. These results show that the developed model has good generalization ability and the error cost for out-of-sample predictions are high. Since the training classifier is dependent on the function within the dataset, we can only be approximately correct for the out-of-range distribution. This data dependence on the finite dataset used for training limits the wide use and application of

ANN models. We, thereby, recommend using input parameters that are within the range used for sample data generation to ensure that we obtain an accurate prediction.

8.3 Challenges associated with the static prediction model

Building energy prediction models typically follow static learning e.g., they do not account for changes in climate, energy production, economy, etc., (Somu et al., 2021). Time-series forecasting techniques are used when real-time historic data is available to forecast future trends of energy consumption (Deb et al., 2017). For instance, Ekonomou (2010) utilized historic data from 1992 to 2004 to train a long-term energy consumption prediction model. This model was later tested using data from 2005 to 2008. However, collecting data for such long time periods, sensor deployment, and monitoring is very expensive (Deb et al., 2017). Moreover, the collected data is severely affected by non-linearity and high levels of uncertainty (Li et al., 2015; Jain et al., 2014; Amasyali and El-Gohary, 2018). Using time-series forecasting techniques on long-term simulated datasets is significantly more complicated.

Operational energy: The challenge with long-term prediction intervals is to account for changes in weather-based parameters over time. The general circulation model (GCM) a.k.a. general climate model is a forecasting tool that generates future weather files for building energy simulations (Zou et al., 2021). The generated weather files require further processing and cannot be used directly in an energy simulation program. First, GCM data is biased i.e., historic GCM data shows considerable deviation from actual weather data collected from meteorological stations. Second, the data has a coarse

resolution of daily average instead of hourly values that are required in simulation tools (Hosseini et al., 2021). Hosseini et al. (2021) applied bias-correction and developed a hybrid classification-regression model to resolve these issues. However, their study observed that (i) GCM data is static i.e., it does not address changes in statistical relations between weather parameters that take place over time, and (ii) generating GCM data is computationally intensive.

Embodied energy: The temporal representativeness and granularity of LCA data are also known to substantially reduce the accuracy of EE assessments (Dixit et al., 2012). LCI databases such as Ecoinvent, Simapro, Chinese life cycle database, etc., that are predominantly used for process-based analysis lack temporal information (Su et al., 2017; Weidema et al., 2013). Similarly, economic IO-data used in IO-based approaches is not published and updated promptly (Langston and Langston, 2008; Crawford 2004; Dixit, 2015). As a result, LCA data lacks temporal relevance in terms of economy, transportation, manufacturing technologies, or fuel mix (Optis and Wild, 2010; Reap et al., 2008; Kendall and Price, 2012). Typically, IO-data is published once every five years (2002, 2007, 2012, etc.), indicating that the granularity of LCA data is very low. For instance, studies use IO-data from 2012 for their current analysis, since it was most recently published by the USBEA. This data is adjusted to the present time using energy inflation factors. Adjusting outdated data to represent the present time may still cause discrepancies in EE analysis. Moreover, these adjustments are made using a yearly time-step, therefore, disregarding temporal variability that occurs within a year (Roux et al., 2016; Itten et al., 2012). Roux et al. (2016) studied the differences in LCA values between

using hourly electricity mix data and annual average mix for the year 2013. They found that using yearly mix data could result in large errors (over 30%) illustrating the importance of using hourly time-steps. These errors would inevitably amplify over the service life (~60 years) of a building. Traditional methods of LCA, therefore, fail to evaluate the influence of time on building environmental performance (Su et al., 2017; Pinsonnault et al., 2014). Several studies recommend using dynamic LCA to overcome the problems caused by traditional static methods. These studies use dynamic characterization factors and dynamic weighting factors to account for time variability. Otherwise, studies suggest developing a time varying LCI database that represents economic and technological progress will improve the accuracy of building LCA (Su et al., 2017).

Our study acknowledges that issues of temporal representativeness and granularity exist in both OE and EE assessments. Developing a long-term prediction model by considering the combined changes in EE (energy mix, economy, and technology) and OE (weather-based parameters) over 60 years is beyond the scope of this study. We, therefore, encoded the climate zone information into binary information to remove the effect of the climate zone on energy consumption. It is important to note that the results of this study should be interpreted with caution and a proper background understanding of the subject matter.

Overcoming these challenges in future studies would improve decision-making related to tasks that have long-term implications such as building retrofitting, energy supply strategy, demand management programs, and energy optimization. ML models can

also be trained to learn the temporal dependencies and variations from LCA and weather data. For instance, Somu et al. (2021) integrated data from real-time observations with historic data using a sliding window approach to improve the performance accuracy of the model.

CHAPTER IX

CONCLUSIONS*

The current state of research shows that several studies have been published in the field of data-driven building energy prediction. Most of these studies are focused on the operational energy component of building LCE. Studies show that the need to construct energy-efficient buildings has indeed increased the proportion of EE in the building. Therefore, this research investigated the feasibility of developing an ML framework to predict both EE intensity and OE intensity to address building energy from a life cycle perspective. For this, we conducted a preliminary study to identify the more complete, well-established, and robust method for EE assessment. We also defined standardized guidelines and protocols for parametric modeling in terms of LOD and system boundary definitions to resolve methodological inconsistencies Next, we developed a parametric framework to generate a simulation-based building energy dataset for different building typologies. This synthesized dataset was generated using the Rhino-Grasshopper environment. The data generated in this manner was used to develop, train, and test the building energy prediction ANN model. Finally, we also demonstrated the application of the prediction model on a case study building.

---

* Part of this chapter is reprinted with permission from:

"Challenges in implementing data-driven approaches for building life cycle energy assessment: A review" by Varusha Venkatraj and Manish Kumar Dixit, Renewable and Sustainable Energy Reviews, 160, 112327, Copyright (2022) by Elsevier

The findings of this study show that the developed ANN model does have high prediction accuracy in estimating both OE intensity and EE intensity. The analytical results show that the most optimum neural network architecture had a MAPE value of 1.95% and R^2 of 0.96 for OE intensity; and a MAPE value of 2.47% and R^2 of 0.90 for EE intensity. More notably, the results of the case study also demonstrate that the predicted and simulated values of OE and EE intensity are in close agreement with each other. These results clearly indicate the benefits of using ML models for BPA in terms of cost, time, and effort. More importantly, designers and engineers would not have to switch back and forth between several tools or deal with issues caused by interoperability. We strongly believe that these benefits would tremendously improve the design decision-making process from a building LCE perspective.

9.1 Research significance and contributions

Data-driven approaches require initial efforts to create and test computational algorithms. Once successfully tested, these algorithms offer more efficient and faster assessment and optimization. This research developed an ML-based energy analytical framework, especially focused on evaluating both EE and OE. The practical implementation of this research would enable designers to evaluate the LCE implications of their design decisions in real-time. Providing immediate feedback would enhance the user interaction and experience while using this model. This framework has the potential to eliminate the use of multiple energy simulation platforms during the early design phase. This would result in reducing the effort, time, and cost spent on running building

simulations. Our case study also demonstrated the implementation of the ML model to estimate building EE and OE using data obtained from Google Earth. We believe that this research would help design professionals evaluate several design options in terms of its LCE performance, thereby, streamlining and improving the existing design-decision making process.

9.2 Future work and recommendations

Despite the ubiquitous use of ML models for building energy prediction, there are still certain gaps and unanswered research problems that need to be addressed. In this section, we suggest potential research directions for future investigations.

- Establishing a transparent and reliable open-source platform where users can upload real world LCA data would significantly contribute to the creation of a large-scale database.
- Providing incentives to organizations or individuals will further promote data collection.
- Enforcing standards on the processes of data collection, data reporting, data management, and data validation is essential to ensure comparability between studies.
- Carrying out research on systematically standardizing the ML model development process specifically for building energy problems is essential.
- Most of the researchers focus their attention on developing the ML model. To promote the widespread use of ML models, other issues such as model generalizability and transferability need to be addressed.

- The impact of environmental factors such as global warming potential, GHG emissions. etc., has always been overlooked.
- Current methods to develop ML models are technologically advanced and computationally expensive; developing a singular cloud-based ML tool that evaluates building energy would help individuals tremendously.
- Existing studies fail to evaluate building LCE from the perspective of time; developing prediction models that account for long-term temporal aspects and climate change implications would play a crucial role in design decision-making.
- Future studies may consider integrating the ML model with optimization algorithms to further explore the entire design space and determine the optimal design option that achieves minimum building LCE usage.

REFERENCES

Abate, D., Towers, M., Dotz, R., & Romani, L. (2009). The Whitestone facility maintenance and repair cost reference 2009-2010. Whitestone Research, California.

Abbasi, S., & Noorzai, E. (2021). The BIM-Based multi-optimization approach in order to determine the trade-off between embodied and operation energy focused on renewable energy use. Journal of Cleaner Production, 281, 125359. https://doi.org/10.1016/J.JCLEPRO.2020.125359

Abediniangerabi, B., Makhmalbaf, A., & Shahandashti, M. (2021). Deep learning for estimating energy savings of early-stage facade design decisions. Energy and AI, 5, 100077. https://doi.org/10.1016/J.EGYAI.2021.100077

Acquaye, A. (2010). A stochastic hybrid embodied energy and CO2_eq intensity analysis of building and construction processes in Ireland.

Ahmad, A. S., Hassan, M. Y., Abdullah, M. P., Rahman, H. A., Hussin, F., Abdullah, H., & Saidur, R. (2014). A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renewable and Sustainable Energy Reviews, 33, 102–109. https://doi.org/10.1016/J.RSER.2014.01.069

Ahmad, M. W., Mourshed, M., & Rezgui, Y. (2017). Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption. Energy and Buildings, 147, 77–89. https://doi.org/10.1016/J.ENBUILD.2017.04.038

Almási, A. D., Woźniak, S., Cristea, V., Leblebici, Y., & Engbersen, T. (2016). Review of advances in neural networks: Neural design technology stack. Neurocomputing, 174, 31–41. https://doi.org/10.1016/J.NEUCOM.2015.02.092

Alobaidi, M. H., Chebana, F., & Meguid, M. A. (2018). Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. Applied Energy, 212, 997–1012. https://doi.org/10.1016/J.APENERGY.2017.12.054

Alwosheel, A., van Cranenburgh, S., & Chorus, C. G. (2018). Is your dataset big enough? Sample size requirements when using artificial neural networks for discrete choice analysis. Journal of Choice Modelling, 28, 167–182. https://doi.org/10.1016/J.JOCM.2018.07.002

Amarasinghe, K., Marino, D. L., & Manic, M. (2017). Deep neural networks for energy load forecasting. IEEE International Symposium on Industrial Electronics, 1483–1488. https://doi.org/10.1109/ISIE.2017.8001465

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews, 81, 1192–1205. https://doi.org/10.1016/J.RSER.2017.04.095

Amber, K. P., Ahmad, R., Aslam, M. W., Kousar, A., Usman, M., & Khan, M. S. (2018). Intelligent techniques for forecasting electricity consumption of buildings. Energy, 157, 886–893. https://doi.org/10.1016/J.ENERGY.2018.05.155

Ardente, F., Beccali, M., Cellura, M., & Mistretta, M. (2008). Building energy performance: a LCA case study of kenaf-fibres insulation board. Energy and Buildings, 40(1), 1-10.

ASHRAE, A. ASHRAE/IES Standard 90.1-2016: Energy Standard for Buildings Except Low-Rise Residential Buildings. 2016. American Society of Heating, Refrigerating and Air-Conditioning Engineers: Atlanta.

Asl, M. R., Das, S., Tsai, B., Molloy, I., & Hauck, A. (2017). Energy Model Machine (EMM) Instant Building EnergyPrediction using Machine Learning. Ecaade 2017: Sharing of Computable Knowledge! (Shock!), Vol 2, 2(May 2019).

Asl, M. R., Xu, W., Shang, J., Tsai, B., & Molloy, I. (2016). Regression-based building energy performance assessment using building information model (BIM). ASHRAE and IBPSA-USA Building Simulation Conference.

Atmaca, A., & Atmaca, N. (2015). Life cycle energy (LCEA) and carbon dioxide emissions (LCCO2A) assessment of two residential buildings in Gaziantep, Turkey. Energy and Buildings, 102, 417–431. https://doi.org/10.1016/J.ENBUILD.2015.06.008

Attia, S., Gratia, E., de Herde, A., & Hensen, J. L. M. (2012). Simulation-based decision support tool for early stages of zero-energy building design. In Energy and Buildings (Vol. 49). https://doi.org/10.1016/j.enbuild.2012.01.028

Aydinalp, M., Ugursal, V. I., & Fung, A. S. (2004). Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. Applied Energy, 79(2), 159–178. https://doi.org/10.1016/J.APENERGY.2003.12.006

Aygenç, M. (2019). Life Cycle Assessment (LCA) of a LEED-certified green building using two different LCA tools a thesis submitted to the graduate school of natural and applied sciences of middle east technical university.

Ayodele, T. O. (2010). Types of Machine Learning Algorithms. New Advances in Machine Learning.

Azari, R., & Abbasabadi, N. (2018). Embodied energy of buildings: A review of data, methods, challenges, and research trends. Energy and Buildings, 168, 225–235. https://doi.org/10.1016/J.ENBUILD.2018.03.003

Azari, R., Garshasbi, S., Amini, P., Rashed-Ali, H., & Mohammadi, Y. (2016). Multi-objective optimization of building envelope design for life cycle environmental performance. Energy and Buildings, 126, 524–534. https://doi.org/10.1016/J.ENBUILD.2016.05.054

Bagnasco, A., Fresi, F., Saviozzi, M., Silvestro, F., & Vinci, A. (2015). Electrical consumption forecasting in hospital facilities: An application case. Energy and Buildings, 103, 261–270. https://doi.org/10.1016/J.ENBUILD.2015.05.056

Balouktsi, M., Lützkendorf, T., Röck, M., Passer, A., Reisinger, T., & Frischknecht, R. (2020, November). Survey results on acceptance and use of Life Cycle Assessment among designers in world regions: IEA EBC Annex 72. In IOP Conference Series: Earth and Environmental Science (Vol. 588, No. 3, p. 032023). IOP Publishing.

Basbagill, J., Flager, F., Lepech, M., & Fischer, M. (2013). Application of life-cycle assessment to early stage building design for reduced embodied environmental impacts. Building and Environment, 60. https://doi.org/10.1016/j.buildenv.2012.11.009

Baum, M., & Council, U. G. B. (2007). Green building research funding: an assessment of current activity in the United States. Washington, DC: US Green Building Council.

Beccali, M., Cellura, M., Fontana, M., Longo, S., & Mistretta, M. (2013). Energy retrofit of a single-family house: Life cycle net energy saving and environmental benefits. Renewable and Sustainable Energy Reviews, 27, 283-293.

BEES | NIST. (n.d.). Retrieved December 22, 2021, from https://www.nist.gov/services-resources/software/bees

Ben-Nakhi, A. E., & Mahmoud, M. A. (2004). Cooling load prediction for buildings using general regression neural networks. Energy Conversion and Management, 45(13–14), 2127–2141. https://doi.org/10.1016/J.ENCONMAN.2003.10.009

Bin, G., & Parker, P. (2012). Measuring buildings for sustainability: Comparing the initial and retrofit ecological footprint of a century home–The REEP House. Applied Energy, 93, 24-32.

Birgisdottir, H., Moncaster, A., Wiberg, A. H., Chae, C., Yokoyama, K., Balouktsi, M., Seo, S., Oka, T., Lützkendorf, T., & Malmqvist, T. (2017). IEA EBC annex 57 'evaluation of embodied energy and CO2eq for building construction.' Energy and Buildings, 154, 72–80. https://doi.org/10.1016/J.ENBUILD.2017.08.030

Biswas, M. A. R., Robinson, M. D., & Fumo, N. (2016). Prediction of residential building energy consumption: A neural network approach. Energy, 117, 84–92. https://doi.org/10.1016/J.ENERGY.2016.10.066

Blengini, G. A., & Di Carlo, T. (2010). The changing role of life cycle phases, subsystems and materials in the LCA of low energy buildings. Energy and Buildings, 42(6), 869-880.

Bourdeau, M., Zhai, X. qiang, Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. Sustainable Cities and Society, 48, 101533. https://doi.org/10.1016/J.SCS.2019.101533

Brusaferri, A., Matteucci, M., Portolani, P., & Vitali, A. (2019). Bayesian deep learning based method for probabilistic forecast of day-ahead electricity prices. Applied Energy, 250, 1158–1175. https://doi.org/10.1016/J.APENERGY.2019.05.068

Budig, M., Heckmann, O., Ng Qi Boon, A., Hudert, M., Lork, C., & Cheah, L. (2020). DATA-DRIVEN EMBODIED CARBON EVALUATION OF EARLY BUILDING DESIGN ITERATIONS.

Building Journal.com (nd). Online construction demolition estimating. Quickly estimate the cost of residential and commercial demolition projects in over 160 US. Cities. Retrieved from http://www.buildingjournal.com/commercial-construction-estimating-demolition.html

Cabeza, L. F., Boquera, L., Chàfer, M., & Vérez, D. (2021). Embodied energy and embodied carbon of structural building materials: Worldwide progress and barriers through literature map analysis. Energy and Buildings, 231, 110612. https://doi.org/10.1016/J.ENBUILD.2020.110612

Cabeza, L. F., Rincón, L., Vilariño, V., Pérez, G., & Castell, A. (2014). Life cycle assessment (LCA) and life cycle energy analysis (LCEA) of buildings and the building sector: A review. In Renewable and Sustainable Energy Reviews (Vol. 29). https://doi.org/10.1016/j.rser.2013.08.037

Carbon Leadership Forum. (2019). EC3 Tool Methodology Beta - Retrieved December 22, 2021, from https://carbonleadershipforum.org/ec3-methodology/

Cavalliere, C., Habert, G., Dell'Osso, G. R., & Hollberg, A. (2019). Continuous BIM-based assessment of embodied environmental impacts throughout the design process. Journal of Cleaner Production, 211, 941–952. https://doi.org/10.1016/J.JCLEPRO.2018.11.247

Cellura, M., Guarino, F., Longo, S., & Mistretta, M. (2014). Energy life-cycle approach in Net zero energy buildings balance: Operation and embodied energy of an Italian case study. Energy and Buildings, 72, 371–381. https://doi.org/10.1016/J.ENBUILD.2013.12.046

Chalal, M. L., Benachir, M., White, M., & Shrahily, R. (2016). Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review. https://doi.org/10.1016/j.rser.2016.06.040

Chang, Y., Ries, R. J., Man, Q., & Wang, Y. (2014). Disaggregated IO LCA model for building product chain energy quantification: A case from China. Energy and buildings, 72, 212-221.

Chari, A., & Christodoulou, S. (2017). Building energy performance prediction using neural networks. Energy Efficiency 2017 10:5, 10(5), 1315–1327. https://doi.org/10.1007/S12053-017-9524-5

Chastas, P., Theodosiou, T., & Bikas, D. (2016). Embodied energy in residential buildings-towards the nearly zero energy building: A literature review. In Building and Environment (Vol. 105). https://doi.org/10.1016/j.buildenv.2016.05.040

Chen, B., Liu, Q., Chen, H., Wang, L., Deng, T., Zhang, L., & Wu, X. (2021). Multiobjective optimization of building energy consumption based on BIM-DB and LSSVM-NSGA-II. Journal of Cleaner Production, 294, 126153. https://doi.org/10.1016/J.JCLEPRO.2021.126153

Chong, H. Y., Lee, C. Y., & Wang, X. (2017). A mixed review of the adoption of Building Information Modelling (BIM) for sustainability. Journal of Cleaner Production, 142, 4114–4126. https://doi.org/10.1016/J.JCLEPRO.2016.09.222

Chou, J. S., & Bui, D. K. (2014). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. Energy and Buildings, 82, 437–446. https://doi.org/10.1016/J.ENBUILD.2014.07.036

Chung, Y., Haas, P. J., Upfal, E., & Kraska, T. (2018). Unknown Examples & Machine Learning Model Generalization. https://arxiv.org/abs/1808.08294v2

Citherlet, S., & Defaux, T. (2007). Energy and environmental comparison of three variants of a family house during its whole life span. Building and Environment, 42(2), 591–598. https://doi.org/10.1016/J.BUILDENV.2005.09.025

Ciulla, G., & D'Amico, A. (2019). Building energy performance forecasting: A multiple linear regression approach. Applied Energy, 253, 113500. https://doi.org/10.1016/J.APENERGY.2019.113500

Ciulla, G., D'Amico, A., lo Brano, V., & Traverso, M. (2019). Application of optimized artificial intelligence algorithm to evaluate the heating energy demand of non-residential buildings at European level. Energy, 176, 380–391. https://doi.org/10.1016/J.ENERGY.2019.03.168

Collinge, W. O., Landis, A. E., Jones, A. K., Schaefer, L. A., & Bilec, M. M. (2013). Dynamic life cycle assessment: framework and application to an institutional building. The International Journal of Life Cycle Assessment, 18(3), 538-552.

Commercial Reference Buildings | Department of Energy. (n.d.). Retrieved February 21, 2022, from https://www.energy.gov/eere/buildings/commercial-reference-buildings.

Copiello, S. (2016). Economic implications of the energy issue: Evidence for a positive non-linear relation between embodied energy and construction cost. Energy and buildings, 123, 59-70.

Cortes, C., Vapnik, V., & Saitta, L. (1995). Support-Vector Networks Editor. In Machine Leaming (Vol. 20). Kluwer Academic Publishers.

Crawford, R. H. (2004). Using input-output data in life cycle inventory analysis (Ph. D.). Deakin University.

Crawford, R.H. (2008) Validation of a hybrid life-cycle inventory analysis method. Journal of Environmental Management 88 (3):496-506.

Crawford, R.H., Bontinck, P.-A., Stephan, A., Wiedmann, T. and Yu, M. (2018) Hybrid life cycle inventory methods – a review. Journal of Cleaner Production 172:1273-1288.

Crawford, R.H., Stephan, A. and Prideaux, F. (2019) Environmental Performance in Construction (EPiC) database. Melbourne: The University of Melbourne.

D'Amico, A., Ciulla, G., Traverso, M., lo Brano, V., & Palumbo, E. (2019). Artificial Neural Networks to assess energy and environmental performance of buildings: An Italian case study. Journal of Cleaner Production, 239, 117993. https://doi.org/10.1016/J.JCLEPRO.2019.117993

D'Amico, B., Myers, R. J., Sykes, J., Voss, E., Cousins-Jenvey, B., Fawcett, W., Richardson, S., Kermani, A., & Pomponi, F. (2019). Machine Learning for Sustainable Structures: A Call for Data. Structures, 19, 1–4. https://doi.org/10.1016/J.ISTRUC.2018.11.013

Davies, P. J., Emmitt, S., & Firth, S. K. (2014). Challenges for capturing and assessing initial embodied energy: a contractor's perspective. Construction Management and Economics, 32(3), 290-308.

Deb, C., Zhang, F., Yang, J., Lee, S. E., & Shah, K. W. (2017). A review on time series forecasting techniques for building energy consumption. Renewable and Sustainable Energy Reviews, 74, 902–924. https://doi.org/10.1016/J.RSER.2017.02.085

Delgarm, N., Sajadi, B., Delgarm, S., & Kowsary, F. (2016). A novel approach for the simulation-based optimization of the building's energy consumption using NSGA-II: Case study in Iran. Energy and Buildings, 127, 552-560.

Deng, H., Fannon, D., & Eckelman, M. J. (2018). Predictive modeling for US commercial building energy use: A comparison of existing statistical and machine learning algorithms using CBECS microdata. Energy and Buildings, 163, 34–43. https://doi.org/10.1016/J.ENBUILD.2017.12.031

Deru, M. (n.d.). U.S. Life Cycle Inventory Database Roadmap (Brochure). Retrieved December 22, 2021, from www.nrel.gov/

Dixit, M. K. (2015). A Framework for an Improved Input-output-based Hybrid Method for Embodied Energy Calculation.

Dixit, M. K. (2017). Embodied energy analysis of building materials: An improved IO-based hybrid method using sectoral disaggregation. Energy, 124, 46–58. https://doi.org/10.1016/j.energy.2017.02.047

Dixit, M. K. (2017). Life cycle embodied energy analysis of residential buildings: A review of literature to investigate embodied energy parameters. Renewable and Sustainable Energy Reviews, 79, 390-413.

Dixit, M. K., & Singh, S. (2018). Embodied energy analysis of higher education buildings using an input-output-based hybrid method. Energy and Buildings, 161, 41-54.

Dixit, M. K., Culp, C. H., & Fernández-Solís, J. L. (2013). System boundary for embodied energy in buildings: A conceptual model for definition. Renewable and Sustainable Energy Reviews, 21, 153-164.

Dixit, M. K., Culp, C. H., & Fernandez-Solis, J. L. (2014). Calculating primary energy and carbon emission factors for the United States' energy sectors. RSC Advances, 4(97), 54200-54216.

Dixit, M. K., Culp, C. H., & Fernandez-Solis, J. L. (2015). Embodied energy of construction materials: Integrating human and capital energy into an IO-based hybrid model. Environmental science & technology, 49(3), 1936-1945.

Dixit, M. K., Fernández-Solís, J. L., Lavy, S., & Culp, C. H. (2010). Identification of parameters for embodied energy measurement: A literature review. Energy and Buildings, 42(8). https://doi.org/10.1016/j.enbuild.2010.02.016

Dixit, M. K., Fernández-Solís, J. L., Lavy, S., & Culp, C. H. (2012). Need for an embodied energy measurement protocol for buildings: A review paper. In Renewable and Sustainable Energy Reviews (Vol. 16, Issue 6, pp. 3730–3743). https://doi.org/10.1016/j.rser.2012.03.021

Doh, J. H., & Panuwatwanich, K. (2014). Variations in embodied energy and carbon emission intensities of construction materials. Environmental Impact Assessment Review, 49, 31-48.

Dong, B., Cao, C., & Lee, S. E. (2005). Applying support vector machines to predict building energy consumption in tropical region. Energy and Buildings, 37(5), 545–553. https://doi.org/10.1016/J.ENBUILD.2004.09.009

Dong, Q., Xing, K., & Zhang, H. (2017). Artificial Neural Network for Assessment of Energy Consumption and Cost for Cross Laminated Timber Office Building in Severe Cold Regions. https://doi.org/10.3390/su10010084

EC3 Tool Key Features. (n.d.). Retrieved December 22, 2021, from https://buildingtransparency.org

Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. Energy and Buildings, 49, 591–603. https://doi.org/10.1016/J.ENBUILD.2012.03.010

Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. Advances in Engineering Software, 40(5), 356–362. https://doi.org/10.1016/J.ADVENGSOFT.2008.05.003

Ekonomou, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. Energy, 35(2), 512–517. https://doi.org/10.1016/J.ENERGY.2009.10.018

Elbeltagi, E., & Wefki, H. (2021). Predicting energy consumption for residential buildings using ANN through parametric modeling. Energy Reports, 7, 2534–2545. https://doi.org/10.1016/J.EGYR.2021.04.053

Elnabawi, M. H. (2020). Building Information Modeling-Based Building Energy Modeling: Investigation of Interoperability and Simulation Results. Frontiers in Built Environment, 6, 193. https://doi.org/10.3389/FBUIL.2020.573971/BIBTEX

Estimator, A. I. (2017). Athena Impact Estimator for Buildings V4. 2 Software and Database Overview. Retrieved, 10(14), 2017.

Estimator, A. I. (n.d.). Athena Sustainable Materials Institute. (n.d.). Retrieved December 22, 2021, from http://www.athenasmi.org/our-software-data/impact-estimator/

Fan, C., Xiao, F., & Wang, S. (2014). Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Applied Energy, 127, 1–10. https://doi.org/10.1016/J.APENERGY.2014.04.016

Fan, C., Xiao, F., & Zhao, Y. (2017). A short-term building cooling load prediction method using deep learning algorithms. Applied Energy, 195, 222–233. https://doi.org/10.1016/J.APENERGY.2017.03.064

Fan, C., Xiao, F., Yan, C., Liu, C., Li, Z., & Wang, J. (2019). A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning. Applied Energy, 235, 1551–1560. https://doi.org/10.1016/J.APENERGY.2018.11.081

Fang, X., Gong, G., Li, G., Chun, L., Li, W., & Peng, P. (2021). A hybrid deep transfer learning strategy for short term cross-building energy prediction. Energy, 215, 119208. https://doi.org/10.1016/J.ENERGY.2020.119208

Fathi, S., Srinivasan, R., Fenner, A., & Fathi, S. (2020). Machine learning applications in urban building energy performance forecasting: A systematic review. Renewable and Sustainable Energy Reviews, 133, 110287. https://doi.org/10.1016/J.RSER.2020.110287

Fay, R., Treloar, G., and Iyer-Raniga, U. (2000). Life-cycle energy analysis of buildings: a case study. Building Research and Information, 28(1), 31-41.

Feng, K., Lu, W., & Wang, Y. (2019). Assessing environmental performance in early building design stage: An integrated parametric design and machine learning method. Sustainable Cities and Society, 50, 101596. https://doi.org/10.1016/J.SCS.2019.101596

Frey, P., Anderson, P., Andrews, M., & Wolf, C. (2010). Building reuse: Finding a place on American climate policy agendas.

Fumo, N. (2014). A review on the basics of building energy estimation. Renewable and Sustainable Energy Reviews, 31, 53–60. https://doi.org/10.1016/J.RSER.2013.11.040

García Kerdan, I., & Morillón Gálvez, D. (2020). Artificial neural network structure optimisation for accurately prediction of exergy, comfort and life cycle cost performance of a low energy building. Applied Energy, 280, 115862. https://doi.org/10.1016/j.apenergy.2020.115862

Ghattas, R., Gregory, J., Olivetti, E., Greene, S., & Hub, C. S. (2013). Life Cycle Assessment for Residential Buildings: a Literature Review and Gap Analysis, a Publication of Concrete Sustainability Hub.

Gilles, F., Bernard, S., Ioannis, A., & Simon, R. (2017). Decision-making based on network visualization applied to building life cycle optimization. Sustainable Cities and Society, 35, 565–573. https://doi.org/10.1016/j.scs.2017.09.006

Goia, F. (2016). Search for the optimal window-to-wall ratio in office buildings in different European climates and the implications on total energy saving potential. Solar Energy, 132, 467-492.

Guan, J., Zhang, Z., & Chu, C. (2016). Quantification of building embodied energy in China using an input–output-based hybrid LCA model. Energy and Buildings, 110, 443-452.

Hammond, G. P., & Jones, C. I. (2008). Embodied energy and carbon in construction materials. Proceedings of Institution of Civil Engineers: Energy, 161(2), 87–98. https://doi.org/10.1680/ener.2008.161.2.87

Han, H., Cui, X., Fan, Y., & Qing, H. (2019). Least squares support vector machine (LS-SVM)-based chiller fault diagnosis using fault indicative features. Applied Thermal Engineering, 154, 540–547. https://doi.org/10.1016/J.APPLTHERMALENG.2019.03.111

Harkouss, F., Fardoun, F., & Biwole, P. H. (2018). Multi-objective optimization methodology for net zero energy buildings. Journal of Building Engineering, 16, 57–71. https://doi.org/10.1016/J.JOBE.2017.12.003

Hernandez, P., and Kenny, P. (2011). Development of a methodology for life cycle building energy ratings. Energy policy, 39(6), 3779-3788.

Hollberg, A., Ruth, J., & Hollberg alexanderhollberg, A. (n.d.). BUILDING COMPONENTS AND BUILDINGS LCA in architectural design-a parametric approach. The International Journal of Life Cycle Assessment. https://doi.org/10.1007/s11367-016-1065-1

Hong, T., Kim, J., & Lee, M. (2019). A multi-objective optimization model for determining the building design and occupant behaviors based on energy, economic, and environmental performance. Energy, 174, 823–834. https://doi.org/10.1016/J.ENERGY.2019.02.035

Horowitz, K. J., & Planting, M. A. (2006). Concepts and Methods of the us input-Output Accounts (No. 0066). Bureau of Economic Analysis.

Hosseini, M., Bigtashi, A., & Lee, B. (2021). Generating future weather files under climate change scenarios to support building energy simulation – A machine learning approach. Energy and Buildings, 230, 110543. https://doi.org/10.1016/J.ENBUILD.2020.110543

Huang, S., Lin, Y., Chinde, V., Ma, X., & Lian, J. (2021). Simulation-based performance evaluation of model predictive control for building energy systems. Applied Energy, 281, 116027. https://doi.org/10.1016/J.APENERGY.2020.116027

Ibn-Mohammed, T., Greenough, R., Taylor, S., Ozawa-Meida, L., and Acquaye, A. (2013). Operational vs. embodied emissions in buildings—A review of current trends. Energy and Buildings, 66, 232-245.

IBPSA. (2019). Best Directory | Building Energy Software Tools. (n.d.). Retrieved December 22, 2021, from https://www.buildingenergysoftwaretools.com/

IEA (2019), Tracking Buildings, IEA, Paris https://www.iea.org/reports/tracking-buildings

IES Virtual Environment | The Leading Integrated Suite for Accurate Whole Building Performance Simulation. (n.d.). Retrieved December 22, 2021, from https://www.iesve.com/software/virtual-environment

Ilbeigi, M., Ghomeishi, M., & Dehghanbanadaki, A. (2020). Prediction and optimization of energy consumption in an office building using artificial neural network and a genetic algorithm. Sustainable Cities and Society, 61, 102325. https://doi.org/10.1016/J.SCS.2020.102325

Itten, R., Frischknecht, R., Stucki, M., Scherrer, P., & Psi, I. (2012). Life cycle inventories of electricity mixes and grid. http://esu-services.ch/fileadmin/download/publicLCI/itten-2012-electricity-mix.pdf

Jain, R. K., Smith, K. M., Culligan, P. J., & Taylor, J. E. (2014). 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, 123, 168–178. https://doi.org/10.1016/J.APENERGY.2014.02.057

Jovanović, R., Sretenović, A. A., & Živković, B. D. (2015). Ensemble of various neural networks for prediction of heating energy consumption. Energy and Buildings, 94, 189–199. https://doi.org/10.1016/J.ENBUILD.2015.02.052

Junnila, S., Horvath, A., & Guggemos, A. A. (2006). Life-cycle assessment of office buildings in Europe and the United States. Journal of Infrastructure systems, 12(1), 10-17.

KalogirouP, S. A., NeocleousP, C. C., & SchizasP, C. N. (n.d.). BUILDING HEATING LOAD ESTIMATION USING ARTIFICIAL NEURAL NETWORKS.

Kamel, E., Sheikh, S., & Huang, X. (2020). Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. Energy, 206, 118045. https://doi.org/10.1016/j.energy.2020.118045

Karimpour, M., Belusko, M., Xing, K., and Bruno, F. (2014). Minimising the life cycle energy of buildings: Review and analysis. Building and Environment, 73, 106-114.

Kavaklioglu, K. (2011). Modeling and prediction of Turkey's electricity consumption using Support Vector Regression. Applied Energy, 88(1), 368–375. https://doi.org/10.1016/J.APENERGY.2010.07.021

Kavitha, B., & Molykutty, M. v. (2021). Life cycle energy analysis of a glazed commercial building using building information modelling (BIM) tools. Materials Today: Proceedings, 37(Part 2), 940–946. https://doi.org/10.1016/J.MATPR.2020.06.148

Kaytez, F. (2020). A hybrid approach based on autoregressive integrated moving average and least-square support vector machine for long-term forecasting of net electricity consumption. Energy, 197, 117200. https://doi.org/10.1016/J.ENERGY.2020.117200

Kendall, A., & Price, L. (2012). Incorporating time-corrected life cycle greenhouse gas emissions in vehicle regulations. Environmental Science and Technology, 46(5), 2557–2563. https://doi.org/10.1021/es203098j

Khalil, A. J., Barhoom, A. M., Abu-Nasser, B. S., Musleh, M. M., & Abu-Naser, S. S. (2019). Energy Efficiency Prediction using Artificial Neural Network. International Journal of Academic Pedagogical Research (IJAPR), 3(9), 1–7. https://philpapers.org/rec/KHAEEP

Khasreen, M. M., Banfill, P. F. G., & Menzies, G. F. (2009). Life-cycle assessment and the environmental impact of buildings: A review. Sustainability, 1(3), 674–701. https://doi.org/10.3390/su1030674

Kiss, B., & Szalay, Z. (2020). Modular approach to multi-objective environmental optimization of buildings. Automation in Construction, 111, 103044. https://doi.org/10.1016/J.AUTCON.2019.103044

Kouw, W. M., & Loog, M. (2018). An introduction to domain adaptation and transfer learning. https://arxiv.org/abs/1812.11806v2

Krogmann, U., Minderman, N., Senick, J., & Andrews, C. J. (2008). Life-cycle assessment of New Jersey Meadowlands Commission Center for environmental and scientific education building. Rutgers Center for Green Building, New Brunswick, New Jersey, USA.

Kumar, R., Kumar Aggarwal Yashwant Singh Parmar, R., Dhar Sharma, J., Aggarwal, R., & Sharma, J. D. (2013). Estimation of Total Energy Load of Building Using Artificial Neural Network. Energy and Environmental Engineering, 1(2), 25–35. https://doi.org/10.13189/eee.2013.010201

Kumar, S., Pal, S. K., & Singh, R. P. (2018). A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. Energy and Buildings, 176, 275–286. https://doi.org/10.1016/J.ENBUILD.2018.06.056

Kunkel, K., Stevens, L., Stevens, S., Sun, L., Janssen, E., Wuebbles, D., Kruk, M., Thomas, D., Shulski, M., Umphlett, N., Hubbard, K., Robbins, K., Romolo, L., Akyuz, A., Pathak, T., Bergantino, T., & Dobson, J. (2013). Regional Climate Trends and Scenarios for the U.S. National Climate Assessment Part 4. Climate of the U.S. Great Plains. HPRCC Personnel Publications. https://digitalcommons.unl.edu/hprccpubs/39

Kwok, S. S. K., Yuen, R. K. K., & Lee, E. W. M. (2011). An intelligent approach to assessing the effect of building occupancy on building cooling load prediction. Building and Environment, 46(8), 1681–1690. https://doi.org/10.1016/J.BUILDENV.2011.02.008

Lai, F., Magoulès, F., & Lherminier, F. (2008). Vapnik's learning theory applied to energy consumption forecasts in residential buildings.

Https://Doi.Org/10.1080/00207160802033582.
https://doi.org/10.1080/00207160802033582

Langston, Y. L., & Langston, C. A. (2008). Reliability of building embodied energy modelling: An analysis of 30 Melbourne case studies. Construction Management and Economics, 26(2). https://doi.org/10.1080/01446190701716564

Lee, C. M., & Ko, C. N. (2009). Time series prediction using RBF neural networks with a nonlinear time-varying evolution PSO algorithm. Neurocomputing, 73(1–3), 449–460. https://doi.org/10.1016/J.NEUCOM.2009.07.005

Lee, S., Jung, S., & Lee, J. (n.d.). Prediction Model Based on an Artificial Neural Network for User-Based Building Energy Consumption in South Korea. https://doi.org/10.3390/en12040608

Lei, L., Chen, W., Wu, B., Chen, C., & Liu, W. (2021). A building energy consumption prediction model based on rough set theory and deep learning algorithms. Energy and Buildings, 240, 110886. https://doi.org/10.1016/J.ENBUILD.2021.110886

Lenzen, M. (2000) Errors in Conventional and Input-Output-based Life-Cycle Inventories. Journal of Industrial Ecology 4 (4):127-148.

Lewis, M. (2021). States Act to Reduce Embodied Carbon in Public Procurement.

Li, A., Xiao, F., Fan, C., & Hu, M. (n.d.). Development of an ANN-based building energy model for information-poor buildings using transfer learning Article History. https://doi.org/10.1007/s12273-020-0711-5

Li, C., Ding, Z., Zhao, D., Yi, J., & Zhang, G. (2017). Building energy consumption prediction: An extreme deep learning approach. Energies, 10(10). https://doi.org/10.3390/en10101525

Li, J., Wu, Z., & Zhang, H. C. (2008). Application of neural network on environmental impact assessment tools. International Journal of Sustainable Manufacturing, 1(1–2), 100–121. https://doi.org/10.1504/IJSM.2008.019229

Li, Q., Meng, Q., Cai, J., Yoshino, H., & Mochida, A. (2009). Applying support vector machine to predict hourly cooling load in the building. Applied Energy, 86(10), 2249–2256. https://doi.org/10.1016/J.APENERGY.2008.11.035

Li, X., Lǔ, J. H., Ding, L., Xu, G., & Li, J. (2009). Building cooling load forecasting model based on LS-SVM. Proceedings - 2009 Asia-Pacific Conference on Information Processing, APCIP 2009, 1, 55–58. https://doi.org/10.1109/APCIP.2009.22

Li, X., Yuyan, D., Lixing, D., & Liangzhong, J. (2010). Building cooling load forecasting using fuzzy support vector machine and fuzzy C-mean clustering. CCTAE 2010 - 2010

International Conference on Computer and Communication Technologies in Agriculture Engineering, 1, 438–441. https://doi.org/10.1109/CCTAE.2010.5543577

Lim, Y. W., Majid, H. A., Samah, A. A., Ahmad, M. H., and Ossen, D. R. (2018). Bim and Genetic Algorithm Optimisation for Sustainable Building Envelope Design. Building Information Systems in the Construction Industry, 159.

Lippiatt, B. (1998, June). Building for environmental and economic sustainability (BEES). In Construction and the environment, CIB World Congress on, in Gävle, Sweden June.

Liu, T., Tan, Z., Xu, C., Chen, H., & Li, Z. (2020). Study on deep reinforcement learning techniques for building energy consumption forecasting. Energy and Buildings, 208, 109675. https://doi.org/10.1016/j.enbuild.2019.109675

Liu, Z., Wu, D., Liu, Y., Han, Z., Lun, L., Gao, J., Jin, G., & Cao, G. (2019). Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction. Energy Exploration & Exploitation, 37(4), 1426–1451. https://doi.org/10.1177/0144598718822400

Lobaccaro, G., Wiberg, A. H., Ceci, G., Manni, M., Lolli, N., & Berardi, U. (2018). Parametric design to minimize the embodied GHG emissions in a ZEB. Energy and Buildings, 167, 106–123. https://doi.org/10.1016/j.enbuild.2018.02.025

Lu, Y., Wang, S., and Shan, K. (2015). Design optimization and optimal control of grid-connected and standalone nearly/net zero energy buildings. Applied Energy, 155, 463-477.

Luo, X. J., Oyedele, L. O., Ajayi, A. O., & Akinade, O. O. (2020). Comparative study of machine learning-based multi-objective prediction framework for multiple building energy loads. Sustainable Cities and Society, 61, 102283. https://doi.org/10.1016/J.SCS.2020.102283

Lützkendorf, T., Foliente, G., Balouktsi, M., & Wiberg, A. H. (2015). Net-zero buildings: Incorporating embodied impacts. In Building Research and Information (Vol. 43, Issue 1). https://doi.org/10.1080/09613218.2014.935575

Ma, Z., Ye, C., Li, H., & Ma, W. (2018). Applying support vector machines to predict building energy consumption in China. Energy Procedia, 152, 780–786. https://doi.org/10.1016/J.EGYPRO.2018.09.245

Machine Learning in MATLAB - MATLAB & Simulink. (n.d.). Retrieved October 10, 2021, from https://www.mathworks.com/help/stats/machine-learning-in-matlab.html

Magnier, L., & Haghighat, F. (2010). Multiobjective optimization of building design using TRNSYS simulations, genetic algorithm, and Artificial Neural Network. Building and Environment, 45(3), 739–746. https://doi.org/10.1016/J.BUILDENV.2009.08.016

Majeau-Bettez, G., Strømman, A.H. and Hertwich, E.G. (2011) Evaluation of process- and input-output-based life cycle inventory data with regard to truncation and aggregation issues. Environmental Science & Technology 45 (23):10170-7.

Marino, D. L., Amarasinghe, K., & Manic, M. (2016). Building energy load forecasting using Deep Neural Networks. IECON Proceedings (Industrial Electronics Conference), 7046–7051. https://doi.org/10.1109/IECON.2016.7793413

Martellotta, F., Ayr, U., Stefanizzi, P., Sacchetti, A., & Riganti, G. (2017). On the use of artificial neural networks to model household energy consumptions. Energy Procedia, 126, 250–257. https://doi.org/10.1016/J.EGYPRO.2017.08.149

Martínez-Rocamora, A., Solís-Guzmán, J., & Marrero, M. (2016). LCA databases focused on construction materials: A review. Renewable and Sustainable Energy Reviews, 58, 565-573.

Massana, J., Pous, C., Burgas, L., Melendez, J., & Colomer, J. (2016). Short-term load forecasting for non-residential buildings contrasting artificial occupancy attributes. Energy and Buildings, 130, 519–531. https://doi.org/10.1016/J.ENBUILD.2016.08.081

Mena, R., Rodríguez, F., Castilla, M., & Arahal, M. R. (2014). A prediction model based on neural networks for the energy consumption of a bioclimatic building. Energy and Buildings, 82, 142–155. https://doi.org/10.1016/J.ENBUILD.2014.06.052

Menzies, G. F., & Tsolaki, T. (2016). Think outside the box: 'Churn' and the environmental impact of office fit-out retrofit. In CIBSE Technical Symposium 2016. Chartered Institution of Building Services Engineers.

Menzies, G. F., Turan, S., & Banfill, P. F. G. (2007). Life-cycle assessment and embodied energy: A review. In Proceedings of Institution of Civil Engineers: Construction Materials (Vol. 160, Issue 4, pp. 135–143). ICE Publishing. https://doi.org/10.1680/coma.2007.160.4.135

Mihalakakou, G., Santamouris, M., & Tsangrassoulis, A. (2002). On the energy consumption in residential buildings. Energy and Buildings, 34(7), 727–736. https://doi.org/10.1016/S0378-7788(01)00137-2

Mocanu, E., Nguyen, P. H., Gibescu, M., & Kling, W. L. (2016). Deep learning for estimating building energy consumption. Sustainable Energy, Grids and Networks, 6, 91–99. https://doi.org/10.1016/J.SEGAN.2016.02.005

Mohandes, S. R., Zhang, X., & Mahdiyar, A. (2019). A comprehensive review on the application of artificial neural networks in building energy analysis. Neurocomputing, 340, 55–75. https://doi.org/10.1016/J.NEUCOM.2019.02.040

Monteiro, H., Fernández, J. E., & Freire, F. (2016). Comparative life-cycle energy analysis of a new and an existing house: The significance of occupant's habits, building systems and embodied energy. Sustainable Cities and Society, 26, 507-518.

Naji, S., Keivani, A., Shamshirband, S., Alengaram, U. J., Jumaat, M. Z., Mansor, Z., & Lee, M. (2016). Estimating building energy consumption using extreme learning machine method. Energy, 97, 506–516. https://doi.org/10.1016/J.ENERGY.2015.11.037

Najjar, M., Figueiredo, K., Hammad, A. W. A., & Haddad, A. (2019). Integrated optimization with building information modeling and life cycle assessment for generating energy efficient buildings. Applied Energy, 250, 1366–1382. https://doi.org/10.1016/j.apenergy.2019.05.101

Nebel, B., Alcorn, A., & Wittstock, B. (2011). Life cycle assessment: adopting and adapting overseas LCA data and methodologies for building materials in New Zealand. Ministry of Agriculture and Forestry.

Nebel, B., Kellenberger, D., Alcorn, A., & Garrett, P. (2009). Life Cycle Inventory–Review of data collection protocols '. Wellington: New Zealand Ministry of Economic Development, 90.

Neto, A. H., & Fiorelli, F. A. S. (2008). Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. Energy and Buildings, 40(12), 2169–2176. https://doi.org/10.1016/J.ENBUILD.2008.06.013

Neural Networks: Structure | Machine Learning Crash Course. (n.d.). Retrieved October 10, 2021, from https://developers.google.com/machine-learning/crash-course/introduction-to-neural-networks/anatomy

Ngo, N. T. (2019). Early predicting cooling loads for energy-efficient design in office buildings by machine learning. Energy and Buildings, 182, 264–273. https://doi.org/10.1016/J.ENBUILD.2018.10.004

Nguyen, A. T., Reiter, S., & Rigo, P. (2014). A review on simulation-based optimization methods applied to building performance analysis. Applied Energy, 113, 1043–1058. https://doi.org/10.1016/J.APENERGY.2013.08.061

Nizam, R. S., Zhang, C., & Tian, L. (2018). A BIM based tool for assessing embodied energy for buildings. Energy and Buildings, 170, 1–14. https://doi.org/10.1016/J.ENBUILD.2018.03.067

One Click LCA . (n.d.). LCA Database of Building products: local and global data for your LCA. (n.d.). Retrieved December 22, 2021, from https://www.oneclicklca.com/support/faq-and-guidance/documentation/database/

Ongsulee, P. (2018). Artificial intelligence, machine learning and deep learning. International Conference on ICT and Knowledge Engineering. https://doi.org/10.1109/ICTKE.2017.8259629

Optis, M., & Wild, P. (2010). Inadequate documentation in published life cycle energy reports on buildings. International Journal of Life Cycle Assessment, 15(7), 644–651. https://doi.org/10.1007/s11367-010-0203-4

Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J. v., Lakshminarayanan, B., & Snoek, J. (2019). Can You Trust Your Model's Uncertainty? Evaluating Predictive Uncertainty Under Dataset Shift. Advances in Neural Information Processing Systems, 32. https://arxiv.org/abs/1906.02530v2

P, W. B., & O, V. C. (n.d.). Swiss Centre for Life Cycle Inventories Overview and methodology (final) Acknowledgements v3.

Paleari, M., Lavagna, M., & Campioli, A. (2013). Life cycle assessment and zero energy residential buildings. In PLEA.

Paudel, S., Elmitri, M., Couturier, S., Nguyen, P. H., Kamphuis, R., Lacarrière, B., & le Corre, O. (2017). A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. Energy and Buildings, 138, 240–256. https://doi.org/10.1016/J.ENBUILD.2016.11.009

Pinsonnault, A., Lesage, P., Levasseur, A., & Samson, R. (n.d.). Temporal differentiation of background systems in LCA: relevance of adding temporal information in LCI databases. LCI METHODOLOGY AND DATABASES. https://doi.org/10.1007/s11367-014-0783-5

Pradeep Kumar, P., Venkatraj, V., & Dixit, M. K. (2022). Evaluating the temporal representativeness of embodied energy data: A case study of higher education buildings. Energy and Buildings, 254, 111596. https://doi.org/10.1016/J.ENBUILD.2021.111596

Praseeda, K. I., Reddy, B. V., and Mani, M. (2016). Embodied and operational energy of urban residential buildings in India. Energy and Buildings, 110, 211-219.

Pray, R. (2016). National Construction Estimator 2017 (Craftsmen) (65th ed.). Craftsman Book Co.

Pullen, S. (1999). Estimating the embodied energy of timber building products. Journal of the Institute of Wood Science, 15(3).

Qiao, Q., Yunusa-Kaltungo, A., & Edwards, R. E. (2021). Towards developing a systematic knowledge trend for building energy consumption prediction. Journal of Building Engineering, 35, 101967. https://doi.org/10.1016/J.JOBE.2020.101967

Rahmani Asl, M., Zarrinmehr, S., Bergin, M., & Yan, W. (2015). BPOpt: A framework for BIM-based performance optimization. Energy and Buildings, 108. https://doi.org/10.1016/j.enbuild.2015.09.011

Ramesh, T., Prakash, R., and Shukla, K. K. (2010). Life cycle energy analysis of buildings: An overview. Energy and buildings, 42(10), 1592-1600.

Rasmussen, F. N., Malmqvist, T., Moncaster, A., Wiberg, A. H., & Birgisdóttir, H. (2018). Analysing methodological choices in calculations of embodied energy and GHG emissions from buildings. Energy and Buildings, 158, 1487–1498. https://doi.org/10.1016/J.ENBUILD.2017.11.013

Rauf, A., & Crawford, R. H. (2015). Building service life and its effect on the life cycle embodied energy of buildings. Energy, 79, 140-148.

Reap, J., Roman, F., Duncan, S., & Bras, B. (2008). A survey of unresolved problems in life cycle assessment. Part 2: Impact assessment and interpretation. In International Journal of Life Cycle Assessment (Vol. 13, Issue 5, pp. 374–388). https://doi.org/10.1007/s11367-008-0009-9

Robinson, C., Dilkina, B., Hubbs, J., Zhang, W., Guhathakurta, S., Brown, M. A., & Pendyala, R. M. (2017). Machine learning approaches for estimating commercial building energy consumption. Applied Energy, 208, 889–904. https://doi.org/10.1016/j.apenergy.2017.09.060

Rodrigues, C., and Freire, F. (2017). Environmental impact trade-offs in building envelope retrofit strategies. The International Journal of Life Cycle Assessment, 22(4), 557-570.

Roudsari, M. S., & Pak, M. (2013). LADYBUG: A parametric environmental plugin for grasshopper to help designers create an environmentally-conscious design.

Roux, C., Schalbart, P., & Peuportier, B. (2016). Accounting for temporal variation of electricity production and consumption in the LCA of an energy-efficient house. Journal of Cleaner Production, 113, 532–540. https://doi.org/10.1016/j.jclepro.2015.11.052

Runge, J., & Zmeureanu, R. (2019). Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. Energies 2019, Vol. 12, Page 3254, 12(17), 3254. https://doi.org/10.3390/EN12173254

Samuel, A.L., (1959). Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development, 3(3).

Santos, L., Schleicher, S., & Caldas, L. (2017). Automation of CAD models to BEM models for performance based goal-oriented design methods. Building and Environment, 112, 144–158. https://doi.org/10.1016/J.BUILDENV.2016.10.015

Sartori, I., & Hestnes, A. G. (2007). Energy use in the life cycle of conventional and low-energy buildings: A review article. Energy and buildings, 39(3), 249-257.

Scheuer, C., Keoleian, G. A., & Reppe, P. (2003). Life cycle energy and environmental performance of a new university building: modeling challenges and design implications. Energy and buildings, 35(10), 1049-1064.

Sebestyen, A. and T. J. (2020). Machine Learning Methods in Energy Simulations for Architects and Designers - The implementation of supervised machine learning in the context of the computational design process. Werner, L and Koering, D (Eds.), Anthropologic: Architecture and Fabrication in the Cognitive Age - Proceedings of the 38th ECAADe Conference - Volume 1, TU Berlin, Berlin, Germany, 16-18 September 2020, Pp. 613-622.

Sekhar Roy, S., Roy, R., & Balas, V. E. (2018). Estimating heating load in buildings using multivariate adaptive regression splines, extreme learning machine, a hybrid model of MARS and ELM. Renewable and Sustainable Energy Reviews, 82, 4256–4268. https://doi.org/10.1016/J.RSER.2017.05.249

Seo, B., Yoon, Y. B., & Cho, S. S. S. (2019). ANN-Based Thermal Load Prediction Approach for Advanced Controls in Building Energy Systems. ARCC Conference Repository. https://arcc-journal.org/index.php/repository/article/view/657

Seo, S., Foliente, G., & Ren, Z. (2018). Energy and GHG reductions considering embodied impacts of retrofitting existing dwelling stock in Greater Melbourne. Journal of Cleaner Production, 170, 1288-1304.

Setiawan, A., Koprinska, I., & Agelidis, V. G. (2009). Very short-term electricity load demand forecasting using support vector regression. Proceedings of the International Joint Conference on Neural Networks, 2888–2894. https://doi.org/10.1109/IJCNN.2009.5179063

Seyedzadeh, S., Pour Rahimian, F., Rastogi, P., & Glesk, I. (2019). Tuning machine learning models for prediction of building energy loads. Sustainable Cities and Society, 47, 101484. https://doi.org/10.1016/J.SCS.2019.101484

Seyrfar, A., Ataei, H., Movahedi, A., & Derrible, S. (2021). Data-Driven Approach for Evaluating the Energy Efficiency in Multifamily Residential Buildings. Practice Periodical on Structural Design and Construction, 26(2). https://doi.org/10.1061/(asce)sc.1943-5576.0000555

Shadram, F., & Mukkavaara, J. (2018). An integrated BIM-based framework for the optimization of the trade-off between embodied and operational energy. Energy and Buildings, 158, 1189–1205. https://doi.org/10.1016/j.enbuild.2017.11.017

Shadram, F., Johansson, T. D., Lu, W., Schade, J., & Olofsson, T. (2016). An integrated BIM-based framework for minimizing embodied energy during building design. Energy and Buildings, 128, 592-604.

Shao, M., Wang, X., Bu, Z., Chen, X., & Wang, Y. (2020). Prediction of energy consumption in hotel buildings via support vector machines. Sustainable Cities and Society, 57, 102128. https://doi.org/10.1016/J.SCS.2020.102128

Sharif, S. A., & Hammad, A. (2019). Simulation-Based Multi-Objective Optimization of institutional building renovation considering energy consumption, Life-Cycle Cost and Life-Cycle Assessment. Journal of Building Engineering, 21, 429–445. https://doi.org/10.1016/J.JOBE.2018.11.006

Sharma, A., Shree, V., & Nautiyal, H. (2012). Life cycle environmental assessment of an educational building in Northern India: A case study. Sustainable Cities and Society, 4, 22-28.

Sholahudin, Alam, A. G., Baek, C. I., & Han, H. (2016). Prediction and Analysis of Building Energy Efficiency Using Artificial Neural Network and Design of Experiments. Applied Mechanics and Materials, 819, 541–545. https://doi.org/10.4028/WWW.SCIENTIFIC.NET/AMM.819.541

Shrivastava, S., and Chini, A. (2012). Using building information modeling to assess the initial embodied energy of a building. International Journal of Construction Management, 12(1), 51-63.

Simonen, K., Rodriguez, B. X., & Wolf, C. de. (2017). Benchmarking the Embodied Carbon of Buildings. Https://Doi.Org/10.1080/24751448.2017.1354623, 1(2), 208–218. https://doi.org/10.1080/24751448.2017.1354623

Singaravel, S., Suykens, J., & Geyer, P. (2018). Deep-learning neural-network architectures and methods: Using component-based models in building-design energy prediction. Advanced Engineering Informatics, 38, 81–90. https://doi.org/10.1016/J.AEI.2018.06.004

Solomon, D. M., Winter, R. L., Boulanger, A. G., Anderson, R. N., & Wu, L. L. (2011). Forecasting Energy Demand in Large Commercial Buildings Using Support Vector Machine Regression. https://doi.org/10.7916/D85D90X7

Somu, N., Raman M R, G., & Ramamritham, K. (2021). A deep learning framework for building energy consumption forecast. Renewable and Sustainable Energy Reviews, 137, 110591. https://doi.org/10.1016/j.rser.2020.110591

Soust-Verdaguer, B., Llatas, C., & García-Martínez, A. (2017). Critical review of bim-based LCA method to buildings. Energy and Buildings, 136, 110–120. https://doi.org/10.1016/J.ENBUILD.2016.12.009

Stephan, A., & Stephan, L. (2016). Life cycle energy and cost analysis of embodied, operational and user-transport energy reduction measures for residential buildings. Applied Energy, 161. https://doi.org/10.1016/j.apenergy.2015.10.023

Stephan, A., Crawford, R. H., & de Myttenaere, K. (2013). A comprehensive assessment of the life cycle energy demand of passive houses. Applied Energy, 112, 23-34.

Stephan, A., Jensen, C. A., & Crawford, R. H. (2017). Improving the Life Cycle Energy Performance of Apartment Units through Façade Design. Procedia Engineering, 196, 1003-1010.

Stuart Geiger, R., Yu, K., Yang, Y., Dai, M., Qiu, J., Tang, R., & Huang, J. (2020). Garbage in, garbage out? Do machine learning application papers in social computing report where human-labeled training data comes from? FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3351095.3372862

Su, S., Li, X., Zhu, Y., & Lin, B. (2017). Dynamic LCA framework for environmental impact assessment of buildings. Energy and Buildings, 149, 310–320. https://doi.org/10.1016/J.ENBUILD.2017.05.042

Sun, L., Kunkel, K. E. (Kenneth E. 1950-, Stevens, L. E., Buddenberg, A., Dobson, J. G., & Easterling, D. R. (2015). Regional surface climate conditions in CMIP3 and CMIP5 for the United States : differences, similarities, and implications for the U.S. National Climate Assessment. https://doi.org/10.7289/V5RB72KG

Sun, Y., Haghighat, F., & Fung, B. C. M. (2020). A review of the-state-of-the-art in data-driven approaches for building energy prediction. In Energy and Buildings (Vol. 221, p. 110022). Elsevier Ltd. https://doi.org/10.1016/j.enbuild.2020.110022

Taghizade, K., Heidari, A., & Noorzai, E. (2019). Environmental Impact Profiles for Glazing Systems: Strategies for Early Design Process. Journal of Architectural Engineering, 25(2). https://doi.org/10.1061/(asce)ae.1943-5568.0000343

Tally. (n.d.). Retrieved November 13, 2020, from https://www.choosetally.com/

Thiers, S., & Peuportier, B. (2012). Energy and environmental assessment of two high energy performance residential buildings. Building and Environment, 51, 276–284. https://doi.org/10.1016/J.BUILDENV.2011.11.018

Thilakarathna, P. S. M., Seo, S., Baduge, K. S. K., Lee, H., Mendis, P., & Foliente, G. (2020). Embodied carbon analysis and benchmarking emissions of high and ultra-high strength

concrete using machine learning algorithms. Journal of Cleaner Production, 262, 121281. https://doi.org/10.1016/J.JCLEPRO.2020.121281

Thomas, A., Menassa, C. C., & Kamat, V. R. (2016). System dynamics framework to study the effect of material performance on a building's lifecycle energy requirements. Journal of Computing in Civil Engineering, 30(6), 04016034.

Thormark, C. (2002). A low energy building in a life cycle—its embodied energy, energy need for operation and recycling potential. Building and environment, 37(4), 429-435.

Touloupaki, E., & Theodosiou, T. (2017). Energy Performance Optimization as a Generative Design Tool for Nearly Zero Energy Buildings. Procedia Engineering, 180, 1178–1185. https://doi.org/10.1016/J.PROENG.2017.04.278

Treloar, G., & Fay, M. R. (1998). The embodied energy of living. BEDP Environment Design Guide, (August), 1-8.

Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy and Buildings, 49, 560–567. https://doi.org/10.1016/J.ENBUILD.2012.03.003

Turhan, C., Kazanasmaz, T., Uygun, I. E., Ekmen, K. E., & Akkurt, G. G. (2014). Comparative study of a building energy performance software (KEP-IYTE-ESS) and ANN-based building heat load estimation. Energy and Buildings, 85, 115–125. https://doi.org/10.1016/J.ENBUILD.2014.09.026

U.S. Energy Information Agency. (2019). Annual Energy Outlook 2020 with projections to 2050. Annual Energy Outlook 2019 with Projections to 2050, 44(8).

U.S. Life Cycle Inventory Database | NREL. (n.d.). Retrieved December 22, 2021, from https://www.nrel.gov/lci/

USCB (nd). 2012 Economic Census, All sectors: Core Business Statistics Series: Comparative Statistics for the U.S. and the States (2007 NAICS Basis): 2012 and 2007. United States Census Bureau. Retrieved from https://data.census.gov/cedsci/table?q=23&tid=ECNCOMP2012.EC1200CCOMP1&n=23

USCB (nd). NAICS Definitions. United States Census Bureau. Retrieved from https://www.census.gov/cgi-bin/sssd/naics/naicsrch?code=238910&search=2017%20NAICS%20Search

USGBC, A. (2007). National Green Building Research Agenda. USGBC, Washington.

Utama, A., and Gheewala, S. H. (2009). Indonesian residential high-rise buildings: A life cycle energy assessment. Energy and Buildings, 41(11), 1263-1268.

Venkatraj, V., & Dixit, M. K. (2021). Life cycle embodied energy analysis of higher education buildings: A comparison between different LCI methodologies. Renewable and Sustainable Energy Reviews, 144. https://doi.org/10.1016/j.rser.2021.110957

Venkatraj, V., & Dixit, M. K. (2022). Challenges in implementing data-driven approaches for building life cycle energy assessment: A review. Renewable and Sustainable Energy Reviews, 160, 112327. https://doi.org/10.1016/J.RSER.2022.112327

Venkatraj, V., Dixit, M. K., Yan, W., & Lavy, S. (2020). Evaluating the impact of operating energy reduction measures on embodied energy. Energy and Buildings, 226. https://doi.org/10.1016/j.enbuild.2020.110340

Vukotic, L., Fenner, R. A., & Symons, K. (2010, September). Assessing embodied energy of building structural elements. In Proceedings of the Institution of Civil Engineers-Engineering Sustainability (Vol. 163, No. 3, pp. 147-158). Thomas Telford Ltd.

Wahid, F., & Kim, D. (2016). A Prediction Approach for Demand Analysis of Energy Consumption Using K-Nearest Neighbor in Residential Buildings. International Journal of Smart Home, 10(2), 97–108. https://doi.org/10.14257/ijsh.2016.10.2.10

Waldman, B., Huang, M., & Simonen, K. (2020). Embodied carbon in construction materials: a framework for quantifying data quality in EPDs. Buildings and Cities, 1(1). https://doi.org/10.5334/bc.31

Wang, C., Lu, S., Chen, H., Li, Z., & Lin, B. (2021). Effectiveness of one-click feedback of building energy efficiency in supporting early-stage architecture design: An experimental study. Building and Environment, 196, 107780. https://doi.org/10.1016/j.buildenv.2021.107780

Wang, W., Rivard, H., & Zmeureanu, R. (2005). An object-oriented framework for simulation-based green building design optimization with genetic algorithms. Advanced Engineering Informatics, 19(1). https://doi.org/10.1016/j.aei.2005.03.002

Wang, Y., & Li, Y. (2010). Applying LS-SVM to predict primary energy consumption. 2010 International Conference on E-Product E-Service and E-Entertainment, ICEEE2010. https://doi.org/10.1109/ICEEE.2010.5660473

Wang, Z., & Srinivasan, R. S. (2016). A review of artificial intelligence based building energy prediction with a focus on ensemble prediction models. Proceedings - Winter Simulation Conference, 2016-February, 3438–3448. https://doi.org/10.1109/WSC.2015.7408504

Wang, Z., Hong, T., & Piette, M. A. (2020). Building thermal load prediction through shallow machine learning and deep learning. Applied Energy, 263, 114683. https://doi.org/10.1016/J.APENERGY.2020.114683

Wang, Z., Liu, J., Zhang, Y., Yuan, H., Zhang, R., & Srinivasan, R. S. (2021). Practical issues in implementing machine-learning models for building energy efficiency: Moving beyond obstacles. In Renewable and Sustainable Energy Reviews (Vol. 143, p. 110929). Elsevier Ltd. https://doi.org/10.1016/j.rser.2021.110929

Wang, Z., Wang, Y., & Srinivasan, R. S. (2018). A novel ensemble learning approach to support building energy use prediction. Energy and Buildings, 159, 109–122. https://doi.org/10.1016/J.ENBUILD.2017.10.085

Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. Energy and Buildings, 171, 11–25. https://doi.org/10.1016/J.ENBUILD.2018.04.008

Wei, Y., Zhang, X., Shi, Y., Xia, L., Pan, S., Wu, J., Han, M., & Zhao, X. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. Renewable and Sustainable Energy Reviews, 82, 1027–1047. https://doi.org/10.1016/J.RSER.2017.09.108

Wong, S. L., Wan, K. K. W., & Lam, T. N. T. (2010). Artificial neural networks for energy analysis of office buildings with daylighting. Applied Energy, 87(2), 551–557. https://doi.org/10.1016/J.APENERGY.2009.06.028

Wu, X., Kumar, V., Ross Quinlan, J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P. S., Zhou, Z.-H., Steinbach, M., Hand, D. J., & Steinberg, D. (2007). Top 10 algorithms in data mining. Knowledge and Information Systems 2007 14:1, 14(1), 1–37. https://doi.org/10.1007/S10115-007-0114-2

Y. G., and Norton, B. (2002). Life-cycle operational and embodied energy for a generic single-storey office building in the UK. Energy, 27(1), 77-92.

Yan, C. W., & Yao, J. (2010). Application of ANN for the prediction of building energy consumption at different climate zones with HDD and CDD. Proceedings of the 2010 2nd International Conference on Future Computer and Communication, ICFCC 2010, 3. https://doi.org/10.1109/ICFCC.2010.5497626

Ye, Z., & Kim, M. K. (2018). Predicting electricity consumption in a building using an optimized back-propagation and Levenberg–Marquardt back-propagation neural network: Case study of a shopping mall in China. Sustainable Cities and Society, 42, 176–183. https://doi.org/10.1016/J.SCS.2018.05.050

Ye, Z., Cheng, K., Hsu, S.-C., Wei, H.-H., & Cheung, C. M. (2021). Identifying critical building-oriented features in city-block-level building energy consumption: A data-driven machine learning approach. Applied Energy, 301, 117453. https://doi.org/10.1016/J.APENERGY.2021.117453

Yezioro, A., Dong, B., & Leite, F. (2008). An applied artificial intelligence approach towards assessing building performance simulation tools. Energy and Buildings, 40(4), 612–620. https://doi.org/10.1016/J.ENBUILD.2007.04.014

Yildiz, B., Bilbao, J. I., & Sproul, A. B. (2017). A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renewable and Sustainable Energy Reviews, 73, 1104–1122. https://doi.org/10.1016/J.RSER.2017.02.023

Yousif, S., & Bolojan, D. (2021). Deep-performance: Incorporating deep learning for automating building performance simulation in generative systems. Projections - Proceedings of the 26th International Conference of the Association for Computer-Aided Architectural Design Research in Asia, CAADRIA 2021, 1.

Yu, W., Li, B., Jia, H., Zhang, M., & Wang, D. (2015). Application of multi-objective genetic algorithm to optimize energy efficiency and thermal comfort in building design. Energy and Buildings, 88. https://doi.org/10.1016/j.enbuild.2014.11.063

Zeng, A., Liu, S., & Yu, Y. (2019). Comparative study of data driven methods in building electricity use prediction. Energy and Buildings, 194, 289–300. https://doi.org/10.1016/J.ENBUILD.2019.04.029

Zeng, R., and Chini, A. (2017). A review of research on embodied energy of buildings using bibliometric analysis. Energy and Buildings.

Zhang, F., Deb, C., Lee, S. E., Yang, J., & Shah, K. W. (2016). Time series forecasting for building energy consumption using weighted Support Vector Regression with differential evolution optimization technique. Energy and Buildings, 126, 94–103. https://doi.org/10.1016/J.ENBUILD.2016.05.028

Zhang, L., Wen, J., Li, Y., Chen, J., Ye, Y., Fu, Y., & Livingood, W. (2021). A review of machine learning in building load prediction. Applied Energy, 285, 116452. https://doi.org/10.1016/J.APENERGY.2021.116452

Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16(6), 3586–3592. https://doi.org/10.1016/J.RSER.2012.02.049

Zhao, H. X., & Magoulès, F. (n.d.). Parallel Support Vector Machines Applied to the Prediction of Multiple Buildings Energy Consumption. In Journal of Algorithms & Computational Technology (Vol. 4, Issue 2).

Zhong, H., Wang, J., Jia, H., Mu, Y., & Lv, S. (2019a). Vector field-based support vector regression for building energy consumption prediction. Applied Energy, 242, 403–414. https://doi.org/10.1016/J.APENERGY.2019.03.078

Zou, Y., Xiang, K., Zhan, Q., & Li, Z. (2021). A simulation-based method to predict the life cycle energy performance of residential buildings in different climate zones of China. Building and Environment, 193, 107663. https://doi.org/10.1016/J.BUILDENV.2021.107663.

Zuo, J., Pullen, S., Rameezdeen, R., Bennetts, H., Wang, Y., Mao, G., ... and Duan, H. (2017). Green building evaluation from a life-cycle perspective in Australia: A critical review. Renewable and Sustainable Energy Reviews, 70, 358-368.