THESIS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Leveraging Machine Learning to Improve Early-stage Building Energy Optimization

XINYUE WANG

Department of Architecture and Civil Engineering

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2025

Leveraging Machine Learning to Improve Early-stage Building Energy Optimization
XINYUE. WANG
ISBN 978-91-8103-343-4

Doktorsavhandlingar vid Chalmers tekniska högskola
Ny serie nr 5800
ISSN 0346-718X

Department of Architecture and Civil Engineering
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone + 46 (0)31-772 1000

Cover:
An illustration displays how a machine learning prediction model can substitute the energy simulation engine in the building energy optimization workflow.
The icons are from https://thenounproject.com/, https://www.istockphoto.com/, and https://pngtree.com/, and generated with ChatGPT

Chalmers Reproservice
Gothenburg, Sweden

Leveraging Machine Learning to Improve Early-stage Building Energy Optimization
XINYUE WANG
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

As more than one-third of global greenhouse gas emissions are related to the operation of buildings, reducing building energy demand is a key area in the architecture, engineering, and construction (AEC) industry. One promising means is to conduct early-stage building energy optimization. Early-stage energy optimization offers substantial potential, as influential architectural design variables (ADVs) can be adjusted at low cost to achieve significant efficiency gains. However, existing optimization workflows depend heavily on energy simulations, which are time-consuming and computationally expensive. To address this, this thesis investigates the application of machine learning (ML) for accelerating early-stage building energy optimization, focusing on three key areas: developing ML prediction models, extending their generalizability through transfer learning (TL), and embedding them into practical optimization workflows.

A key contribution of this research lies in systematically identifying influential ADVs through both literature review and stakeholder surveys. Findings highlight building plan, window-to-wall ratio (WWR), and wall material as consistently important across sources, while practitioners additionally emphasize orientation, shading devices, storey number, storey height, roof type, and roof material. The thesis incorporates ADVs from both evidence-based and practice-based perspectives to ensure the development of robust and practically relevant ML models. Comparative ML experiments further provide recommendations for algorithm selection: Support Vector Machine (SVM) for small datasets, Multiple Linear Regression (MLR) for limited and low-diverse datasets, Artificial Neural Network (ANN) for larger and diverse datasets, and Random Forest (RF) when accuracy outweighs computational efficiency. Guidelines are also proposed for synthetic dataset generation, stressing the need for adequate size and diversity to achieve reliable predictions.

To evaluate generalizability, an ANN model trained on Gothenburg data is transferred to five cities with different climates through transfer learning (TL). TL substantially improves prediction accuracy in heating-dominant contexts (Stockholm, Seattle, Chicago), reducing the need for up to 1,600 training samples and saving over 180 hours of computation. Its effectiveness declines in cooling-dominant climates (Madrid, Miami) but remains beneficial when data availability is limited. While its effectiveness is highest in heating-dominant contexts with data scarcity, the results confirm TL's potential to reduce training requirements and computational time.

Finally, the ML model is integrated into a Grasshopper-based optimization workflow and exemplified with a case study. Results show that while ML-based optimization yields slightly higher energy demand than simulation-based methods, it drastically reduces computation time and provides comparable design outcomes.

Overall, this thesis advances methodological knowledge on selecting ADVs, algorithms, and datasets for ML-based building energy prediction, while also confirming the feasibility of cross-climate adaptation and workflow integration. The findings offer valuable guidance for researchers, software developers, and practitioners seeking to accelerate sustainable building design.

Keywords: Machine Learning, Building Energy, Early-stage Optimization, Stakeholder, Synthetic Dataset, Transfer Learning

List of Publications

This thesis is based on the following work:

- Wang, X., Teigland, R., & Hollberg, A. (2024). Identifying influential architectural design variables for early-stage building sustainability optimization. *Building and Environment*, 111295. https://doi.org/10.1016/j.buildenv.2024.111295
- Wang, X., Yu, Y., Teigland, R., & Hollberg, A. (2025). Size or diversity? Synthetic dataset recommendations for machine learning heating energy prediction models in early design stages for residential buildings. *Energy and AI*, 100557.https://doi.org/10.1016/j.egyai.2025.100557
- Wang, X., Yu, Y., Abbasabadi, N., Teigland, R., & Hollberg, A. (2025). Transfer Learning for Generalizing ANN-Based Building Energy Prediction Across Climate Zones. *Submitted.*

Author contribution:

Xinyue Wang – Writing original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, and Conceptualization.

Other work and publications not appended:

- Wang, X., Teigland, R., & Hollberg, A. (2022, September). A Pathway to Climate Neutral Buildings: Definitions, Policy and Stakeholder Understanding in Sweden and China. In IOP Conference Series: Earth and Environmental Science (Vol. 1078, No. 1, p. 012122). IOP Publishing.

- Wang, X., Harrison, J., Teigland, R., & Hollberg, A. (2024, May). Machine Learning (ML) as a Surrogate Model for Early-stage Heating Demand Optimization. In *Proceedings of the SimBuild Conference*.

- Egerlid, H., Wang, X., Thuvander, L., & Maiullari, D. (2025). Carbon efficiency of passive cooling measures in future climate scenarios: Renovating multi-family residential buildings in a Swedish context. *Energy and Buildings*, *334*, 115502.

Acknowledgement

This research project is funded by the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS) [Grant No. 2020-00934]. The research in this thesis was conducted between September 2021 and December 2025.

It seems like only yesterday that I stepped into this very office I am sitting in right now and began my PhD journey. These four years just slipped away in the blink of an eye. I met so many people that I am truly grateful for along the way. First and foremost, I would like to express my sincere gratitude to my main supervisor, Alexander Hollberg. Thank you for giving me the opportunity to pursue my PhD in such an open and inspiring environment, for continually encouraging me to explore the research directions that sparked my curiosity, and for supporting every decision I made along the way. I not only learned immense knowledge from you, but also learned how to grow into a researcher under your guidance. Your enthusiasm for research has truly inspired me, and your belief in me has given me the confidence to continue pursuing an academic career. I would also like to thank my co-supervisor, Robin Teigland, for the positive energy you always bring to our meetings, and for all the interesting and inspiring discussions we had. A big thank you to Yinan Yu, for all your support in the data science aspect of my research; it was really fun to collaborate with you. To Narjes Abassabadi for hosting me during my visit in University of Washington. To Ingela for helping me with all the paperwork. To my examiner, Holger Wallbaum, for always checking on me and for keeping me on track.

I would also like to express my gratitude to all participants in my survey study, for taking time from their busy schedules to help me in my research.

I feel extremely lucky to be part of the Sustainable Building Research Area and have so many wonderful and kind colleagues. I made amazing friends here and had great companion that made my journey truly enjoyable. I would like to express my sincere gratitude to all my colleagues. In particular, I would like to extend my special thanks to: Toivo for being the best deskmate, for our research discussion, and for your friendship; Sanjay for all your support in coding since I started my PhD; Anna for your firendship and for all our yappings that motivate me to go to work; Amrita for our shared time and sweet brunches; Fredrik, Ali, Shuang, Xin, and Quan for making our office a happy place; Josie for your inspiration and all the machine learning stuff you taught me; Jieming for helping me with my modeling; Dawn for your friendship from Sweden and Netherlands. And to Aida, David, Emiline, Luca, Sara, Yagmur, for all the fun events and for the laughter you bring to the hallway and the lunchroom.

I would also like to thank all my friends in Gothenburg, both inside and outside of the climbing gym. You made my life a thousand times better. Above all, I want to thank my mom, who, despite always worrying about me and missing me from afar, has always tried her best to trust me and given me the freedom to pursue whatever I truly wished to do. And my dad, who always tried to give me the best, I know he would be very proud if he could see this work.

Lastly, I would like to express gratitude and pay tribute to myself, for somehow managing to finish this, despite the countless setbacks and frustrations along the way, for still being able to enjoy this PhD, and for still preserving the ability to enjoy research itself.

Xinyue Wang

Gothenburg, December 2025

Abbreviation

ADVs	Architectural Design Variables
AEC	Architecture, Engineering, and Construction
GHG	Greenhouse Gas
ML	Machine Learning
ANN	Artificial Neural Network
SVM	Support Vector Machine
SVR	Support Vector Regression
RF	Random Forrest
MLR	Multiple Linear Regression
DT	Decision Tree
PCA	Principal Component Analysis
GH	Grasshopper
PC	Principal Components
MAE	Mean Absolute Error
RMSE	Root Mean Squared Error
WWR	Window-to-wall Ratio
IDF	Input Data File
RNN	Recurrent Neural Network
DNN	Deep Neural Network
GP	Gaussian Processes
LCA	Life Cycle Assessment
LCC	Life Cycle Cost
XGBoost	eXtreme Gradient Boosting
ELM	Extreme Learning Machine
CNN	Convolutional Neural Network
TL	Transfer Learning

Contents

Part I

Overview

Chapter 1

Introduction

1.1. Background

Buildings account for more than one-third of global greenhouse gas emissions, making them a major contributor to environmental issues [1][2][3]. Consequently, finding ways to reduce energy use in buildings has become important in working toward sustainability goals. One means to achieve this is to conduct building energy optimization in the early design stages, as 70% of decisions related to a building's sustainability are made during the early stage [4] and these decisions are responsible for 80% of the building's environmental impact throughout its life cycle [5]. Building energy simulation models have played an important role in early-stage building energy optimization. Being able to predict building energy at an early stage can support architects in developing more sustainable design proposals and, therefore, reduce building energy demand significantly.

Building energy optimization at an early stage often combines the methods of parametric design, energy simulation, and optimization algorithms simultaneously [6]. The goal of building energy optimization is to determine the optimal design by identifying the combination of different architectural design variables (ADVs) that results in a low energy demand. ADVs are the physical design elements that describe the building's physical and thermal features, such as building shape, orientation, and U-values of materials. To facilitate optimization in the early stage, various computational optimization tools have been developed to facilitate design choices and develop optimal solutions. Today, there are multiple approaches to developing these tools. One mainstream way is through physics-based modeling, which means developing tools based on physics-based energy simulation engines such as Energy Plus [7], IES [8], and Daysim [9]. These approaches typically demand detailed information and a thorough understanding of the building and its energy systems [10]. Furthermore, physics-based building energy prediction models can be very time-consuming and computationally heavy. During the optimization process, many design alternatives with different ADVs are generated. This requires running hundreds or even thousands of simulations to identify the optimum design. Even though one building energy simulation may take only a few minutes, a large number of simulations consume a lot of time, up to a few days. Therefore, these optimization tools are generally very time-consuming [11] and inefficient for the early stage.

More recently, to solve this problem, many researchers are turning to machine learning (ML) as a means to improve the speed and efficiency of these optimization tools[12]. ML is a collection of methods used to fit mathematical models from historical data and to make quick and accurate predictions [13]. Figure 1.1 presents the three steps of the current building energy optimization process: generation of numerous design alternatives with various ADV settings through parametric design; conduct of building energy simulation to identify each alternative's energy demand value; application of optimization algorithms to select one or multiple design alternatives with lower energy demand. Figure 1.1 shows how ML models can replace the existing building energy simulation engine in the current optimization process and improve efficiency and speed.

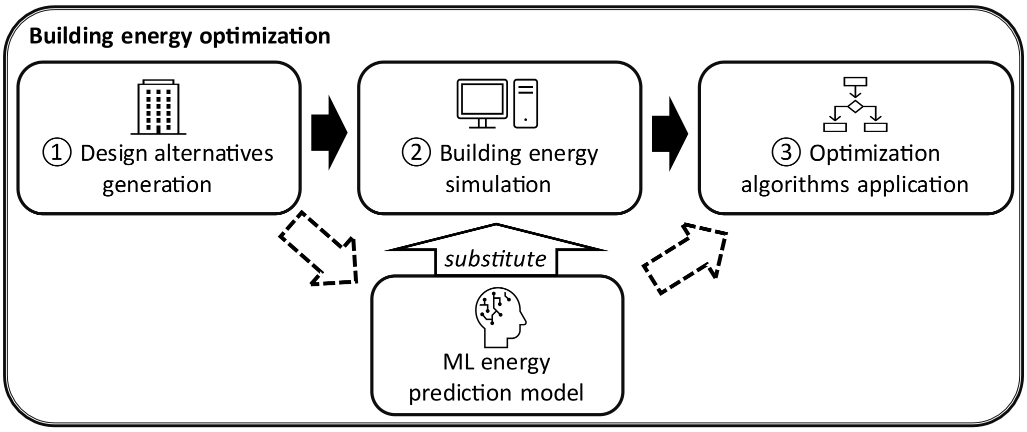


Figure 1.1. Workflow of building energy optimization and how ML models could be used.

1.2 Problem Statement

While showing promise, these ML prediction models are difficult to use in early-stage optimization. As shown in Figure 1.1, the key to an ML-based building energy optimization tool is to develop an ML energy prediction model to substitute the time-consuming simulation engine. Figure 1.2 shows what an ML energy prediction model comprises. To develop this model, the ADVs used as input data must be defined first. Second, the output energy demand for various building design alternatives is required. A dataset containing both the input ADVs and the respective energy demand values output is used to train the proposed ML model. Finally, a separate dataset is used to test the model. A high-quality training dataset and a suitable ML algorithm are significant in developing an ML model with good performance.

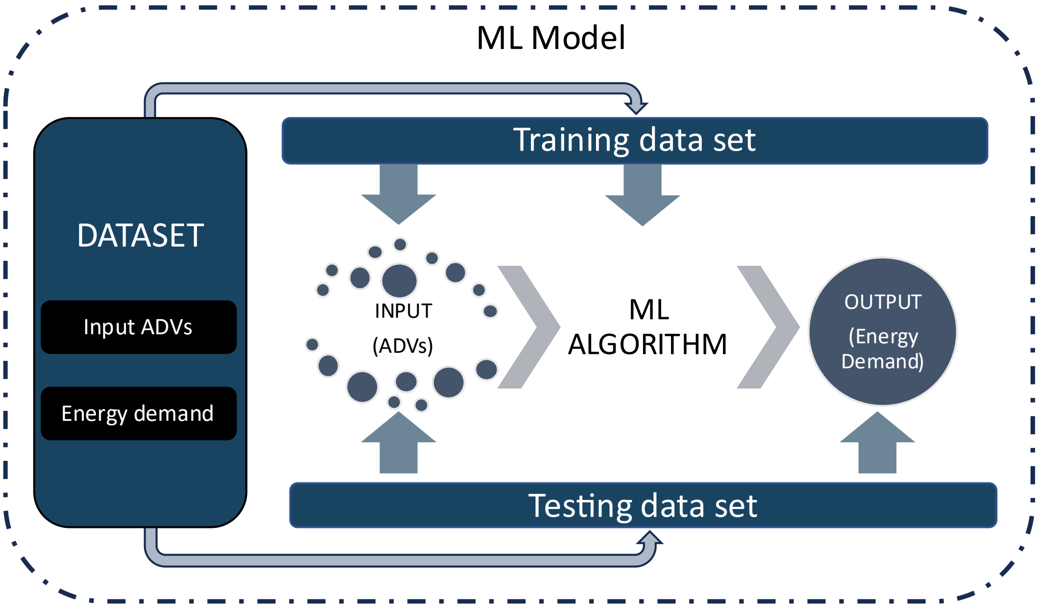


Figure 1.2. The composition of an ML model for building energy prediction.

Most existing ML energy prediction models rely on measured data, and the available datasets include data of ADVs that can typically only be retrieved at a later stage, e.g., the number of heated rooms, the fabric of the built envelope, and insulation conditions [14][15][16]. To provide an alternative, researchers have started to train ML models on synthetic datasets [17][18]. Synthetic data refers to data that is artificially manufactured rather than generated by real-world events. In this case, it means the building configurations that are generated by the software and the simulated energy demand. However, even with synthetic datasets, some studies used ADVs that are difficult to determine at the early stage, such as equipment density and lighting [19]. Others did use early-stage ADVs as input, but their selected ADVs are often not the ones used in optimization. For instance, building area is an ADV frequently used as an input in early-stage ML energy prediction; however, it is usually pre-defined before the architects start to design. Furthermore, including all ADVs in one optimization model not only exponentially increases the number of potential solutions but also the computational costs [20]. In general, there is a lack of proper datasets with the right ADVs for developing ML energy prediction models for early-stage optimization.

It is well-known that the size and diversity of training datasets are crucial for developing ML prediction models [21]. As such, most previous studies use very large datasets with data points exceeding 10000 when developing ML energy prediction models to ensure accuracy [22]. Therefore, the synthetic dataset cannot be as large as the researchers want. Under this context, the questions 'How much data is enough?' and 'How diverse should the data be?' are often raised. The size and the diversity of synthetic datasets in current ML-based early-stage building performance optimization vary [23][24][25]. At the same time, the applied ML algorithms also vary. However, no single ML algorithm has been proven to outperform other ML algorithms for all circumstances. Previous studies only investigated the best-performing algorithms in terms of accuracy under specific training datasets [22][26][27], or different training datasets in terms of size [28], or how to most efficiently increase data points to have better model performance [29]. So far, there have been no studies looking into the compatibility between different ML algorithms and datasets with different sizes and diversity at the same time in building energy predictions. Therefore, there is a lack of understanding of which algorithm works best, in terms of accuracy and computational efficiency, for which synthetic dataset.

Apart from the data scarcity, another major limitation of the ML building energy prediction models is the lack of generalizability, especially in the climatic context. Building energy performance is highly influenced by local climate conditions. The ML model trained on building samples in one specific location can only learn the pattern between building configuration and energy demand under the specific climate conditions. When such models are applied to buildings in a different location with distinct weather patterns, energy codes, or usage habits, their predictive performance tends to degrade significantly. This geographic dependency limits the scalability and practical application of ML-based energy models. To predict building energy under different locations or climatic conditions, it is necessary to regenerate the dataset and retrain the machine learning model, which can be time-consuming and computationally expensive. Therefore, overcoming this limitation is critical to enable broader adoption of data-driven methods in diverse climatic contexts. Transfer learning (TL) can be a solution to tackle the generalizability problem. TL techniques can apply the knowledge gained from relevant data in a previous ML task of data-rich scenarios to improve the performance of a newly given target ML task that lacks sufficient data [30]. However, there are certain gaps in applying TL in building energy prediction models. First, most previous research mainly applied TL from one or a few buildings to another building (Figure 1.3. A). Both the base dataset from the data-rich scenario and the insufficient dataset from the target data-poor scenario are collected from one or a few limited buildings. To develop a data-driven model for early-stage building energy prediction, it is typically necessary to have a dataset containing information including various ADVs from multiple building configurations (Figure 1.3. B), instead of having data from only one or a few buildings. Moreover, most applications of TL in the building domain have been limited to buildings located in the same city, making it unclear whether such methods can be effectively transferred across different climate zones.

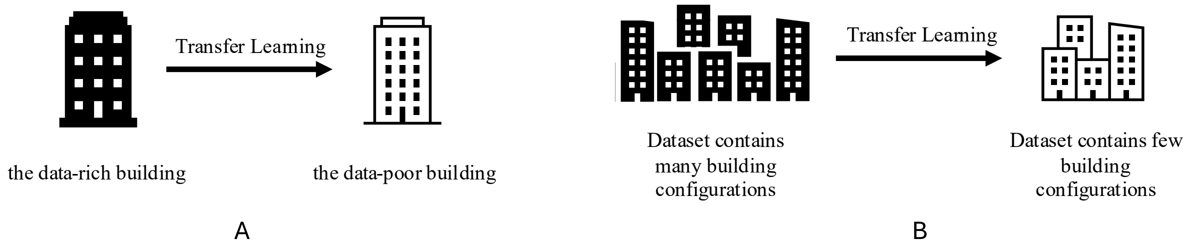


Figure 1.3. Applying TL in building energy prediction.

1.3 Aim and Research Questions

To address the above problems, the aim of this thesis is to investigate how ML can support early-stage building energy optimization. To achieve this aim, three research questions (RQ) are further explored:

RQ1. Which ADVs should be the input for an ML-based early-stage building energy optimization tool?

RQ2. How should the synthetic dataset be developed for training, and which ML algorithm should be used when developing an ML-based building energy prediction model?

RQ3. How can an existing ML model for early-stage optimisation be applied to other climates to avoid new data generation and the model tuning process?

1.4 Research Design

Four studies were conducted to answer the above research questions shown in Figure 1.4. It is worth noting that all case studies take places in Gothenburg, Sweden.

Study A aims to identify the most influential ADVs for early-stage building energy optimization and use them as input for ML building energy prediction models. Currently, sensitivity analysis based on computer simulations is the most commonly used means to identify which ADVs are the most influential in the early stage. However, although valid, this method primarily investigates individual cases only in their specific contexts, thereby restricting their generalization to other contexts. Therefore, Study A suggests a combination of the methods of literature review and stakeholder survey. Bringing in a stakeholder's perspective is beneficial as on the one hand, stakeholders are experts in the relevant fields, and they possess not only domain-specific knowledge and expertise but also a contextual understanding that can greatly enhance the development and effectiveness of optimization tools [40]. Their engagement can improve the optimization tools by providing actual practical experience. Some stakeholders are also the end users of the optimization tools, and extensive research in areas such as user-design and user-driven development [31][32][33] clearly shows the importance of integrating users in the development process for an effective product or service result. Further, previous research also indicates that users should be engaged during the early stage to improve a building's final performance [34]. Study A combines a literature review with survey data from 24 architects and sustainability consultants in the Nordics. By comparing and analyzing the results, the most influential ADVs in early-stage building energy optimization were found. These ADVs could be used as the input for the ML energy prediction model. A paper detailing Study A is published and attached in Part II.

To answer research question two, Study B investigates the best-performing ML algorithm in building energy prediction models as well as the characteristics of the corresponding synthetic dataset. The performance of ML algorithms are evaluated from the perspective of accuracy and computational efficiency. A literature review was first conducted on which ML algorithms are the most frequently used in building energy prediction models. A parametric model was developed to generate random building design alternatives based on different input ADVs that were identified in Study A within nine building shapes. Synthetic datasets with different characteristics, including the size of the dataset and the diversity of data points were generated in this step. ML experiments were later carried out to test the best algorithms depending on the different characteristics of synthetic datasets. Multiple combinations of ML algorithms and synthetic datasets were proposed as outcomes. A paper detailing Study B is published and attached in Part II.

To answer research question three, Study C takes an ML model developed from Study B as an example and applies it to five different cities using TL to investigate the effectiveness of transfer learning (TL) approaches in improving the accuracy of energy prediction models for residential buildings across different climate zones when having limited data availability for early-stage optimization. Parametric modeling is once again used to generate datasets for different climates. Multiple TL models will be developed based upon the pre-trained ML model, and fine-tuned using the training dataset from a target city. Subsequently, the testing dataset from the same test city is used to test and evaluate the fine-tuned TL model. Meanwhile, an ML model will also be developed from scratch using the insufficient data from the target city. The new model will be used as a comparison to evaluate whether applying TL can improve the model's performance. A submitted manuscript detailing Study C is attached in Part II.

Study D integrates the ML model developed in Study B into an early-stage building energy optimization workflow to exemplify the implementation of ML in early-stage optimization. Two optimization tasks, including a single-optimization task on building energy and a multi-objective optimization task on both building energy and embodied carbon emissions, are proposed. An optimization workflow based on a simulation engine is also developed as a benchmark. A case study is carried out to test the proposed optimization workflows.

Combining Study A and Study B, recommendations on how to develop an ML building prediction model to support early-stage building energy optimization, including the selection of input ADVs, the best-performing ML algorithm, and the generation of synthetic training datasets, are provided in this thesis. Study C provides TL models that could be applied to other climates with limited datasets. Study D provides an ML-based early-stage building energy optimization workflow for residential buildings in Gothenburg. These outcomes can support researchers and developers who want to integrate ML into the building energy optimization workflow in the early stage to accelerate the process, and demonstrate the potential of applying ML to significantly improve the efficiency of building energy optimization processes.

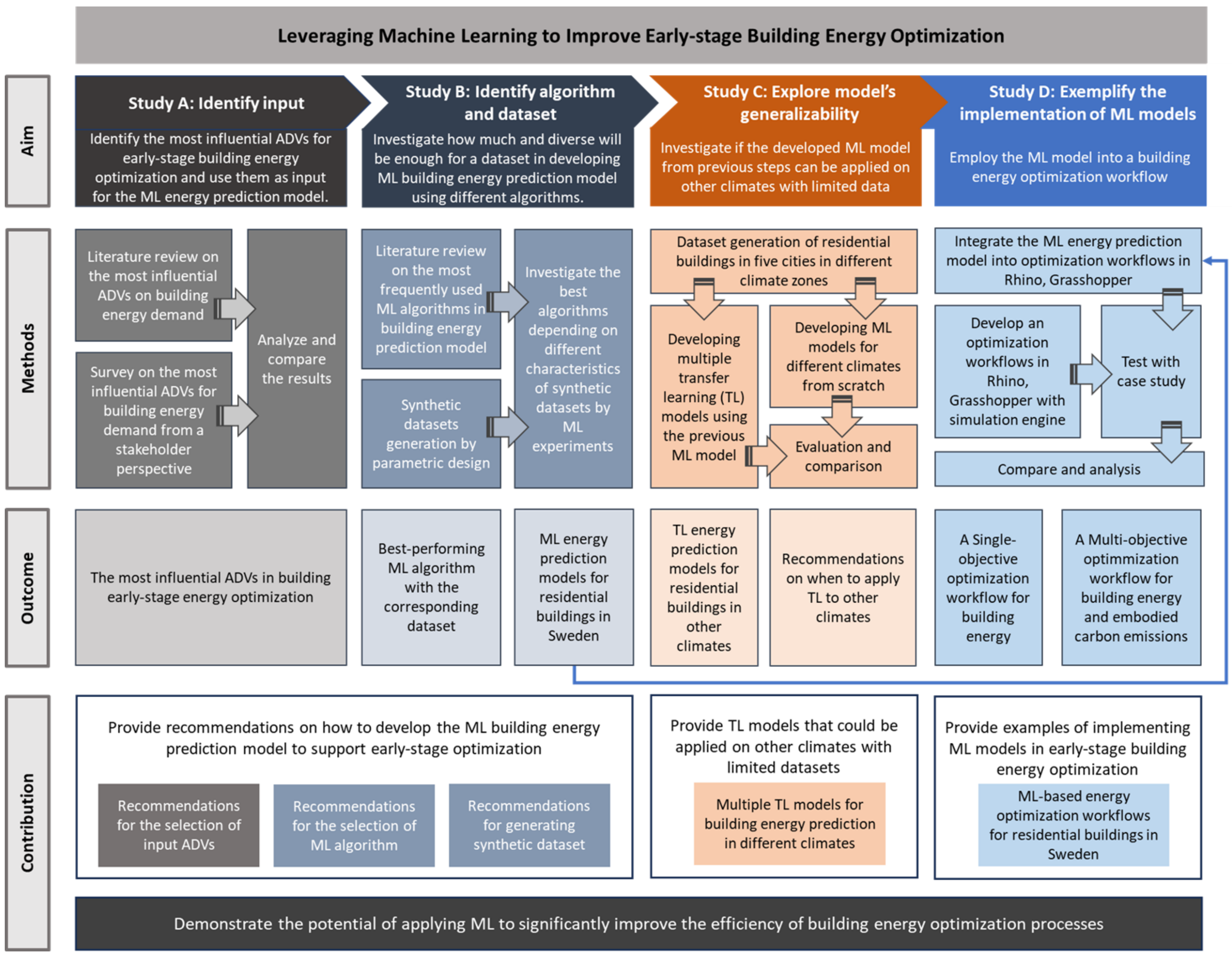


Figure 1.4. Research design.

1.5 Research Scope

1.5.1 Climate context

The ADVs for energy demand can vary depending on the climate context and local policies. Apart from Study C, where the main focus is to transfer the developed ML model to other climates, this thesis mainly focuses on the Nordic countries' context, as Nordic countries are among the global sustainability leaders [35][36]. Nordic countries include Sweden, Denmark, Finland, Norway, and Iceland.

1.5.2 Building type

This thesis focuses on residential buildings as the Nordic residential sector has one of the highest resource requirements in the Nordic countries due to the harsh climate. It is argued that sustainable housing is thus key to the AEC industry's sustainability [37]. It is also worth noting that this thesis only focuses on newly constructed buildings. This is because there is a greater flexibility in terms of choosing ADVs compared to renovation projects. The methods developed here can also be adaptable to retrofit contexts in the future, where informed decision-making remains critical despite greater constraints.

1.5.3 Early-stage

This thesis only focuses on optimization in the early stage of building design. There are many ways to define the early stage and what will be done in this stage. This thesis takes the definition of Stage 1-3 from the Royal Institute of British Architects [38] and 'Preliminary study' and 'Programming' from Swedish building design process. More descriptions can be seen at Section 2.1.

1.5.4 Stakeholders

There are many stakeholders involved in the early stage of building design, such as architects, consultants, engineers, clients, contractors, public authorities, etc. In this thesis, stakeholders only include architects and consultants. Architects consider both the building's aesthetic and functional aspects when creating the building plans, blueprints, and facades and are responsible for making the final decisions regarding the building's design. Consultants, especially sustainability consultants, are generally not directly designing a building but rather providing sustainability insights into building projects. For example, their role can involve assessing the environmental impact of a building's different design alternatives and developing strategies to improve a building's sustainability. Of note is that consultants are usually more familiar with the computational optimization process and energy simulation than architects.

1.5.5 ADVs and features

ADVs in this thesis mainly refer to the architectural design variables which are the physical design elements that describe the building's physical and thermal features. The process of identifying influential ADVs to use as input to feed into an ML model for the prediction of certain building performance objectives is referred to as feature selection in the ML context. The input ADVs in this study are the features in an ML model. To avoid confusion, this thesis only refers to them as input ADVs.

1.5.6 Training dataset size

When the training dataset is large in size, for instance, containing over 10000 data points [15][23], the accuracy of the developed ML model is high. However, as previously mentioned, when using synthetic datasets, it takes too much computational time to generate large datasets. Therefore, this thesis only considers small training datasets. The largest training datasets in the experiments in this study only contain 4800 data points, however, they are referred to as 'large datasets' in the result section.

1.6 Structure of the Thesis

This thesis contains two parts. The first part consists of background, methodology, findings, discussion, and conclusion. Chapter 1 provides the general background as well as the problem statement, aim, and research questions, research scope, and research design. Chapter 2 introduces the extended background of previous research. Chapter 3 presents the methods applied in this thesis. Chapter 4 includes the crucial findings of the conducted studies. Chapter 5 provides discussions, limitations, and outlook. Chapter 6 indicates the main contribution of the work. The second part of the thesis presents the appended papers and manuscripts detailing the conducted studies.

Chapter 2

Extended Background

2.1 Early Stage in the Building Design Process and ADVs

This thesis mainly focuses on the early stage of the building design process. The early stage is a general statement, and different researchers can interpret it differently depending on the geographical context as well as local regulations. This thesis adopts both the building design process proposed by RIBA (Royal Institute of British Architects) as it is the most frequently used process in the literature regarding early-stage optimization and the Swedish building design process as this thesis is conducted in Sweden and takes Swedish residential buildings as a case study. The design process and the selected stages can be seen in Figure 2.1. According to RIBA, the building design process consists of eight stages, Stage 1 ‘Preparation and Briefing’, and Stage 2 ‘Concept Design’ align with the definition of ‘early stage’ [38]. The Swedish design process is similarly organized according to the delivery of a sequence of drawing packages [39]. In the Swedish design process, ‘Preliminary study’ and ‘Programming’ can be considered as early stage. In general, the main tasks in the early stage are to develop the architectural concept for the project and determine an initial design. Figure 2.1 also shows that the greatest potential for design optimization is in the early stage [40]. As the design progresses and deepens, optimizing the architecture becomes increasingly difficult. Therefore, the early stage is the most efficient and proper time to conduct building sustainability optimization.

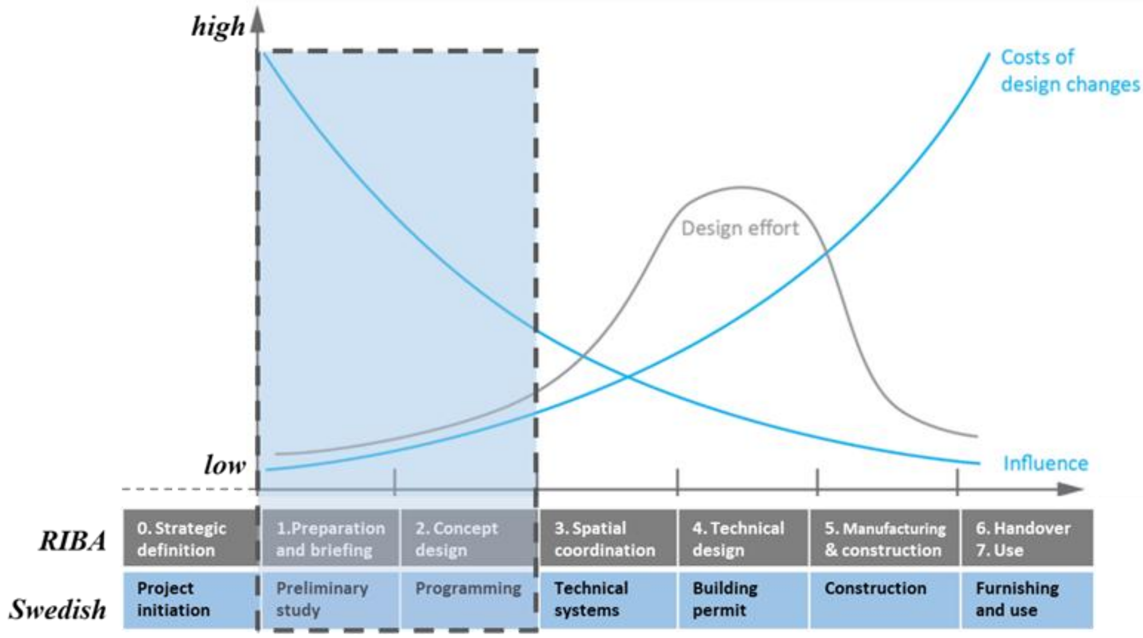


Figure 2.1. Design effort and cost of design changes across project Stages (adapted from 'Paulson curve' [41])

The main subject to optimize is the ADVs. As mentioned in the introduction section, ADVs are the physical design elements that describe the building's physical and thermal features, such as building shape, orientation, and materials. Common ADV categories include the composition of the opaque building envelope, such as wall thickness and material; the composition of the transparent building envelope, such as g-value and u-value of windows, shape, and form; the type of mechanical systems; and the operation of the mechanical systems [42]. The ADV decisions made are critical for a building's sustainability performance. For example, selecting the right wall material can lead to approximately a 17% energy cost reduction [43] while adjusting window scenarios can improve the useful daylight illuminance by approximately 20% [44]. As such, optimizing a building's ADVs in its early stage can significantly reduce the building's environmental impact. However, it is worth noting that not all influential ADVs can be decided in the early stage. ADVs related to concept design such as the shape of the building or the main material are usually defined in the early stage, while ADVs related to technical design such as the mechanical systems or interior materials might be decided much later in the design process. This leads to a main common drawback in conducting optimization in the early stage: the accuracy might be compromised due to insufficient information on ADVs [45].

2.2 Stakeholders' Engagement in Early Stage

Researchers agree that high-performance buildings are best achieved when a project's relevant stakeholders are involved early in the design process. To capture stakeholders' insights, Study A developed surveys to ask related stakeholders' opinions on the influential ADVs to include in the building energy optimization. Literature related to different stakeholders and their main roles have been reviewed. The major category of stakeholders, their job descriptions, and whether they are active in the early stage and listed in Table 2.1.

Table 2.1. Stakeholder categories

Stakeholder category	alternative names in the literature	Main responsibilities	If active in early stage
Architects	Architect[46][47][48] Design professionals [49] Design and drafting personnel [50]	clarify the requirements of the building project clarify project scope and condition develop the building plan decide all related ADVs for the building project sometimes assist with the building's sustainability assessment	√
Sustainability consultants	Sustainability consultants [48] LCA experts [49] Environmental impact consultants [46] Consultants [51] Estimators [50]	assess the project's sustainability performance and improve it. consult the design per all requirements and standards for different sustainability objectives provide technical advice if necessary coordinate between architects and engineers to satisfy certification conditions and environmental compliance	√
Engineers	Structure engineers [47] Building service engineers [46]	in charge of the technical design of the building structure confirm primary engineering configurations and specifications align the building project with safety requirements.	×
Clients	Clients [48][51] Clients and project managers [47] Building owners [46] Financial organization [51] Home buyers [50]	provide financial support participate in deciding ADVs as well as assess the aesthetics of the design and the sustainability performance constantly communicate with architects and sustainability consultants	√
Contractors	Contractors [51]	ensure environmental management occurs and the project is delivered within budget and time.	×
Public authorities	Building permission authorities [47] Government [49][51] Local authorities [46]	law and policy-making and funding provision.	√
End users	End users [46]	use the building once the project is fully finished and put into use.	×

As shown above, the AEC industry involves many different types of stakeholders, but not all of them have significant influence in the early stage. It is worth noting that although end users normally do not get involved in the early stage, recently, user-design and user-driven development [31][32][33] clearly shows the importance of integrating users in the early development process for an effective product or service result. Therefore, it is encouraged that end users should be engaged during the early stage to improve a building's final performance

[34]. The criteria for the chosen stakeholders in this study are that they need to be active in the early stage, and have to either be involved in ADVs' selection or the energy performance assessment. Architects and Sustainability consultants are the stakeholders who have the most interest and influence in the early stage. Architects have the most power towards deciding the ADVs while sustainability consultants are most familiar with building energy performance. Engineers are also involved in deciding ADVs but in a supporting role compared to architects. However, they are not active in the early stage. Contractors are also only active in later stages and therefore excluded from this research. Clients, public authorities, and end users are also the influential stakeholders who are involved in the early stage and related to either the ADVs' selection or the energy performance assessment. Clients will be engaged in the decision of ADVs and energy performance assessments during their interaction with architects and sustainability consultants by providing feedback. Public authority is one of the most important decision-makers concerning several specific ADVs, such as shape and orientation. However, they are excluded in this thesis for two reasons. One reason is that although they are important, their influence and interest in the early stage are much less compared with architects and consultants. The other reason is that it was found only very limited stakeholders in the above three categories are willing to participate in the research. The results would not be valid when there are only a few participants involved. In summary, this study only considers architects and sustainability consultants as related stakeholders.

2.3 Building Energy Optimization Tools in Early Stage

Building energy optimization tools are software applications or platforms designed to assist architects, consultants, or other related stakeholders in designing more sustainable buildings with lower energy demand by proposing one or multiple design alternatives with the best energy performance. Since the turn of the millennia, publications about building optimization have roughly increased tenfold due to the advances and developments in computer science [52]. As mentioned in Figure 1.1 in Section 1, a building energy optimization tool always consists of three parts: design alternative generation through parametric modeling, building energy simulation, and optimization algorithm application.

Parametric modeling is always conducted with computer-aided tools such as Grasshopper in Rhino, or Dynamo in Revit. The basic idea of parametric modeling is to use parametric equations to describe the ADVs of building design alternatives. Compared with the traditional design method where architects use CAD or even draw their designs on paper, parametric design can allow architects to change the ADVs quickly and immediately see how the new design looks. Moreover, the main advantage of the parametric approach is that it can generate numerous different ADVs quickly and easily. Once a parametric model has been developed, generating further design alternatives is nearly effortless [53]. Therefore, parametric modeling can provide numerous design alternatives for the optimization algorithm to choose from. The ADVs that the tool wants to optimize are also defined in the parametric modeling step. For instance, if the optimization tool aims to achieve a low energy demand by optimizing the shape of the building, the dimensions of the building have to be included in the parameterized ADVs; when the aim is to find the most proper building materials that can lead to better energy performance, the materials need to be included in the parametric modeling while the dimensions of the building do not need to. In previous studies, the optimized ADVs can vary depending on the focus of the research, such as building shape [54][55], the layout of the building design [56][57], the setting for the HVAC system [58], the materials [59], window scenarios [60][61], etc.

Energy simulation is an important step in building energy optimization. Most studies use detailed building energy simulation tools or custom-developed tools [62]: including TRNSYS [63], DOE-2 [64], EnergyPlus [65], and ASHRAE toolkit [66] for building load calculations. However, the use of these tools requires detailed information about the design and needs special expertise to avoid garbage-in and garbage-out, therefore, is very time-consuming [62]. To avoid the time-consuming detailed building energy simulations, many studies introduce ML surrogate models to predict building energy demand. These ML prediction models are mathematical models derived from numerous actual measurements or detailed building energy simulations. These models can then be used as a replacement for the detailed building energy simulation during the optimization process for less computational time with an acceptable margin of error. Further information about applying ML in developing energy prediction models for optimization is discussed in section 2.4.

Optimization means finding the best solution(s) among different feasible alternatives, where feasible solutions mean those that satisfy all the constraints [67]. It is summarized that there are mainly three optimization methods applied to building energy optimization tools: exhaustive methods, calculus methods, and stochastic methods [68]. Exhaustive methods aim to list all possible building alternatives with the combination of ADVs and find the best solution. This method can guarantee finding the most optimal solution, however, might take way too much computational time. Calculus methods use mathematical expressions or gradients to find the value of a variable that yields the optimal value of the objective function. Stochastic methods include randomness in the optimization process. This method could reduce computational time, but might only find the relative optimal solution rather than the absolute optimal solution.

In general, when developing an energy optimization tool, how to define the parameterized ADVs, which simulation and optimization methods to use highly depends on the nature of the problem. However, the energy simulation part is most time-consuming in most cases. Therefore, many studies have been focusing on finding a less computationally expensive alternative for building energy simulation methods while maintaining the accuracy of the assessments in an acceptable range to improve the efficiency of building energy optimization tools.

2.4 ML Algorithms in Developing Building Energy Prediction Models

As mentioned in Section 2.3, as energy simulation can be very time-consuming in building energy optimization, studies are increasingly developing ML-based energy prediction models to accelerate the optimization process due to their low time consumption and high performance [69]. ML is a collection of methods used to fit mathematical models from historical data and to make predictions [29]. Based on the historical data, with suitable models and algorithms, machine learning methods could “learn” the non-linear relationship between the independent variables and target variables [27]. There are two types of algorithms, regression and classification. Regression algorithms are used to predict continuous values while classification algorithms are used to predict discrete values. The algorithms used in developing the building energy prediction model for early-stage optimization are the regression models. To investigate the most proper algorithms for this study, a literature review of the existing review papers of the most frequently used ML algorithms for developing building energy prediction models is conducted. The literature review is conducted in Web of Science and Scopus. The combination of the keywords in the search is shown in Figure 2.2.

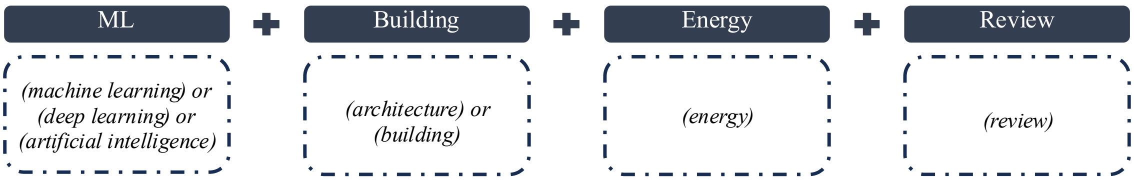


Figure 2.2. Literature search keywords for the most frequently used ML algorithms for developing building energy prediction models.

In total 35 review papers were found and reviewed in this study. Table A in Appendix A summarizes the algorithms mentioned in the 35 review papers. Figure 2.3 presents the top ten most frequently mentioned ML algorithms and their occurrence.

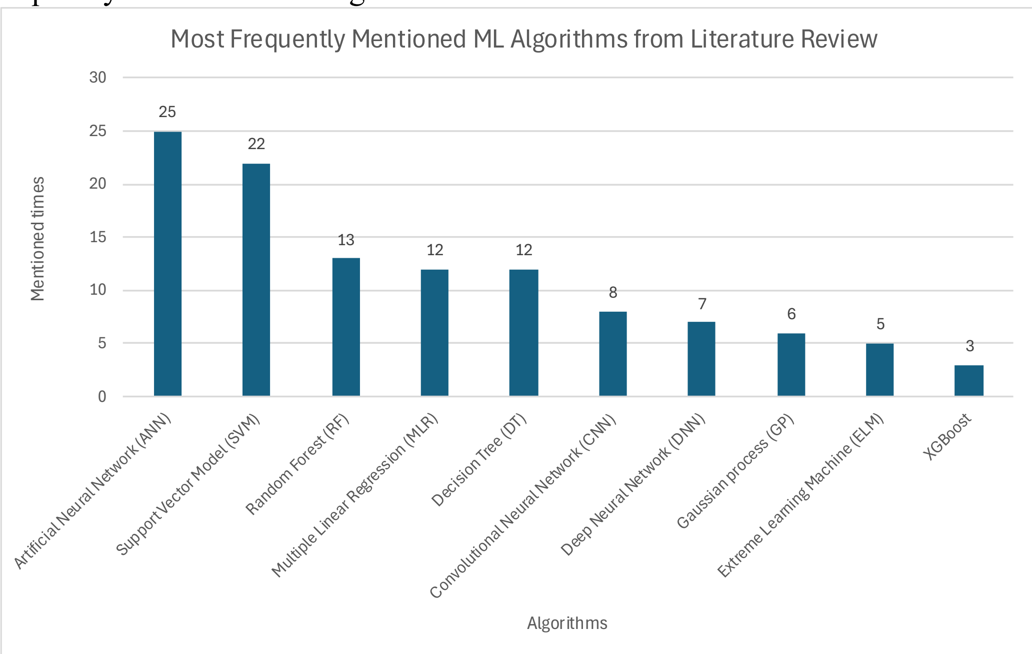


Figure 2.3. The most frequently mentioned ML algorithms from the literature review

Figure 2.3 shows that ANN, SVM, RF, MLR, and DT are the most frequently used ML algorithms in developing building energy prediction models.

Artificial Neural Network (ANN) ANN is the most frequently used algorithm in developing a building energy prediction model. ANN is an ML algorithm with a structural design based on the human brain [70]. The main computational aspects of ANN are their ability to learn from given instances, and then determine patterns and consistencies in data through self-organization [71]. ANN can efficiently emulate the complex relationships of biological networks to answer complex prediction problems [72]. In a neural network, information is broken down into many small pieces, each of which is contained within a single neuron. A neural network can be seen as a sort of “box” that can give an answer to a question or provide an output in the presence of one or more inputs [17]. Many studies have used ANN to develop building energy prediction models. Elbeltagi at el. developed an ANN model through parametric modeling to predict residential building energy consumption using a database of 12000 data points and achieve a mean absolute percentage error (MAPE) of 5.36% [70]. Li et al. developed an ANN-based fast-building energy consumption prediction method for complex architectural forms at the early design stage with accuracy within ±10% [73]. Aruta et al. used ANN to develop building energy load prediction models in different climates [17]. In general, ANN can achieve a high accuracy performance when using it for building energy prediction models.

Support Vector Model (SVM) SVM is centered on the kernel, a technique primarily designed for resolving regression problems developed by Vapnik in the early 1990s [74]. SVM is considered an effective model due to its exceptional characteristic of dealing with samples of extremely high dimensionality [75]. SVM does not especially rely on large amounts of data for training [27]. Fu et al. proposed using SVM to predict the load at a building's system level based on weather predictions and hourly electricity load input and achieve a root mean square error (RMSE) of 15.2% and mean bias error (MBE) of 7.7% [76]. Dong et al. applied SVM to predict building energy consumption in tropical regions with coefficients of variance (CV) less than 3% and percentage error within 4% [77]. Liu et al. developed a few ML building energy prediction models and stated that SVM performed best [27].

Random Forrest (RF) RF is one of the most applied ML models in predicting building energy and related performance: Tian et al. developed an RF model to predict the energy performance of office buildings based on building form with a root mean squared error (RMSE) of 2 kWh/m^2 [78]; Fang et al. used RF to predict construction stage carbon emissions of buildings in the early design phase [79], Olu-Ajayi et al. developed an RF-based building energy prediction strategy with an RMSE of 1.69 kWh/m^2 [28]; and Singh et al. combined RF with building information modeling in supporting the early stage design, reached a mean-absolute-percentage-error (MAPE) of 2.02% [80]. The RF Regressor combines several Decision Tree Regressors to form an ensemble model. Each decision tree splits a feature into 'branches', which ultimately end at a 'leaf' node, where a final value is decided. Since the RF Regressor combines several independent Decision Tree Regressors, it is known for being robust to overfitting [81]. Like all tree-based ML models, it provides feature importance as an output, which can be used to interpret the model. The speed of the RF Regressor largely depends on the 'number of estimators' hyperparameters; this instructs the model on how many independent Decision Tree Regressors should be used in the ensemble model. A high value (exceeding 600 estimators) can sometimes decrease the number of errors but at the cost of greatly slowing down the model's prediction time [81].

Multiple Linear Regression (MLR) An MLR model could also be considered an ML model when it uses the principles of supervised learning, where an algorithm learns to predict a target variable from a set of input features. In previous studies, MLR has been used in estimating commercial building energy [82], analyzing the relationship between ADVs and energy [83], and integrating building performance simulation in an agent-based model using trained regression surrogate models to simulate a building's energy use, with a MAPE of 4% [84]. An MLR model is also often chosen as a baseline to gauge the minimum number of errors that would be reasonable for the given use case. The Linear regression's version of feature importance is the estimated coefficients assigned to each feature after a model has been fitted to the data. Additionally, the MLR is likely to outperform all ML models in terms of speed since it is the simplest regression model.

Decision Tree (DT) DT uses a tree-like flowchart to partition data into groups [75]. DT is an adaptable process that could advance with an enlarged amount of training data [85]. A DT model manifests itself as a graph consisting of a root node and a couple of branch nodes. The model starts from the root node in which the input data are split into different groups based on predefined criteria. These split data are then disseminated to sub-nodes as branches emanating from the root node. The data on sub-nodes will undergo either further or no splits. The former are the internal nodes where the subsequent data split is conducted to form new subgroups as son-branches emanated graphically at the next level. Meanwhile, the latter are leaf nodes, treating their corresponding data group at the current level as their final outputs. DT is also used in previous studies for predicting building energy: Tso et al. use DT to predict building electricity energy consumption [86]. Yu et al. apply DT in building energy demand modeling [87].

2.5 TL in Building Energy Prediction

The definition of TL is: Given a source domain D_S and learning task T_S, a target domain D_T and learning task T_T, TL aims to help improve the learning of the target predictive function $f_T(\cdot)$ in D_T using the knowledge in D_S and T_S, where $D_S \neq D_T$, or $T_S \neq T_T$. In this thesis, the source domain D_S refers to the datasets in the base scenario, and the source task T_S refers to the building energy prediction for the base scenario. The target domain refers to the datasets in the target cities, and the target task T_T refers to the building energy prediction for the target cities.

TL has been applied to many fields, including healthcare [88], computer vision, and language processing. Recently, researchers have conducted exploratory work on applying TL in building energy prediction. TL can be applied across multiple scales, ranging from the prediction of appliance-level consumption via non-intrusive load monitoring (NILM), to specific systems such as heating, ventilation, and air conditioning (HVAC), as well as wastewater treatment processes. Furthermore, TL has been utilized at district and whole-building levels, with temporal resolutions varying from hourly to monthly. Among these applications, whole-building energy prediction has emerged as the most extensively studied topic in the context of smart buildings, largely due to the widespread availability of relevant data. Fang et al. [89] used transfer learning to enhance energy prediction in buildings with few labelled data, employing an LSTM as feature extraction and further fine-tuning a regression layer for domain adaption, studying the effects of different time horizons, architectures, and buildings. Fan et al. [90] compared several parameter-based architectures to enhance building forecasting prediction, analysing how data availability and duration period affect performance. Lastly, Cai et al. [91] exploited TL to increase the accuracy of incentive-based Demand Response (DR), characterized by stochastic and sporadic events, using data from similar customers. Li et al. developed an Artificial Neural Network (ANN) based energy prediction model for information-poor buildings using TL; Jung et al. used TL to enhance building monthly electric load prediction in different districts, with a DNN-based forecasting model on the source data, fine-tuned on the target data [92]; Ozer et al. employed TL to perform short-term load prediction, and proposed an effective method on how to find the single building most useful to perform weight-initialization TL [93]; Qian et al. exploited information of a building with a detailed sensor system to perform an energy prediction on another building with few available data using TL [94]. Previous research has proved that implementing TL can reduce the amount of training data required for the development of the target model, save time for tuning and training models, and improve prediction performance [10]. TL also has the potential to enhance energy model adoption and broaden the applicability of existing building energy prediction models.

However, as stated in Section 1.2, most previous research mainly applied TL from one or a few buildings to another building instead of from a dataset containing information including various architectural design variables from multiple building configurations to another one, which is what is required for early-stage building energy optimization.

Chapter 3

Method

This chapter presents the six-step method applied in this thesis to explore applying ML in early-stage building energy optimization. The overall workflow for the methods is presented in Figure 3.1. Study A investigates identifying input ADVs for the ML prediction model by combining a literature review and stakeholder survey. Study B investigates how to generate a synthetic dataset and select the most suitable ML algorithm. The synthetic dataset is generated by using parametric modeling and visualized by principal component analysis (PCA). In step 3, a literature review is conducted first to select the most frequently used ML algorithms. Multiple ML models with different algorithms are trained on different datasets. Hyperparameter tuning is applied to find the best-performing model. The developed ML models are evaluated by accuracy and computational efficiency at the end to provide recommendations for selecting the most suitable algorithm. Study C explores the ML model's generalizability by applying the ML models developed in step 3 to predict the energy of the same building type in different climates using TL. In step 4, target cities are defined and smaller synthetic datasets on respective climates are generated through parametric modeling, as in step 2. Multiple TL models were developed in step 5 based on the ML model trained in step 3, aiming to predict building energy consumption across different climatic conditions.

Simultaneously, baseline ML models were trained from scratch to serve as benchmarks for evaluating whether TL can enhance prediction accuracy and computational efficiency. In the final step, the developed ML model is embedded in an optimization workflow to exemplify howML can help accelerate the optimization process at early stage by using a case study.

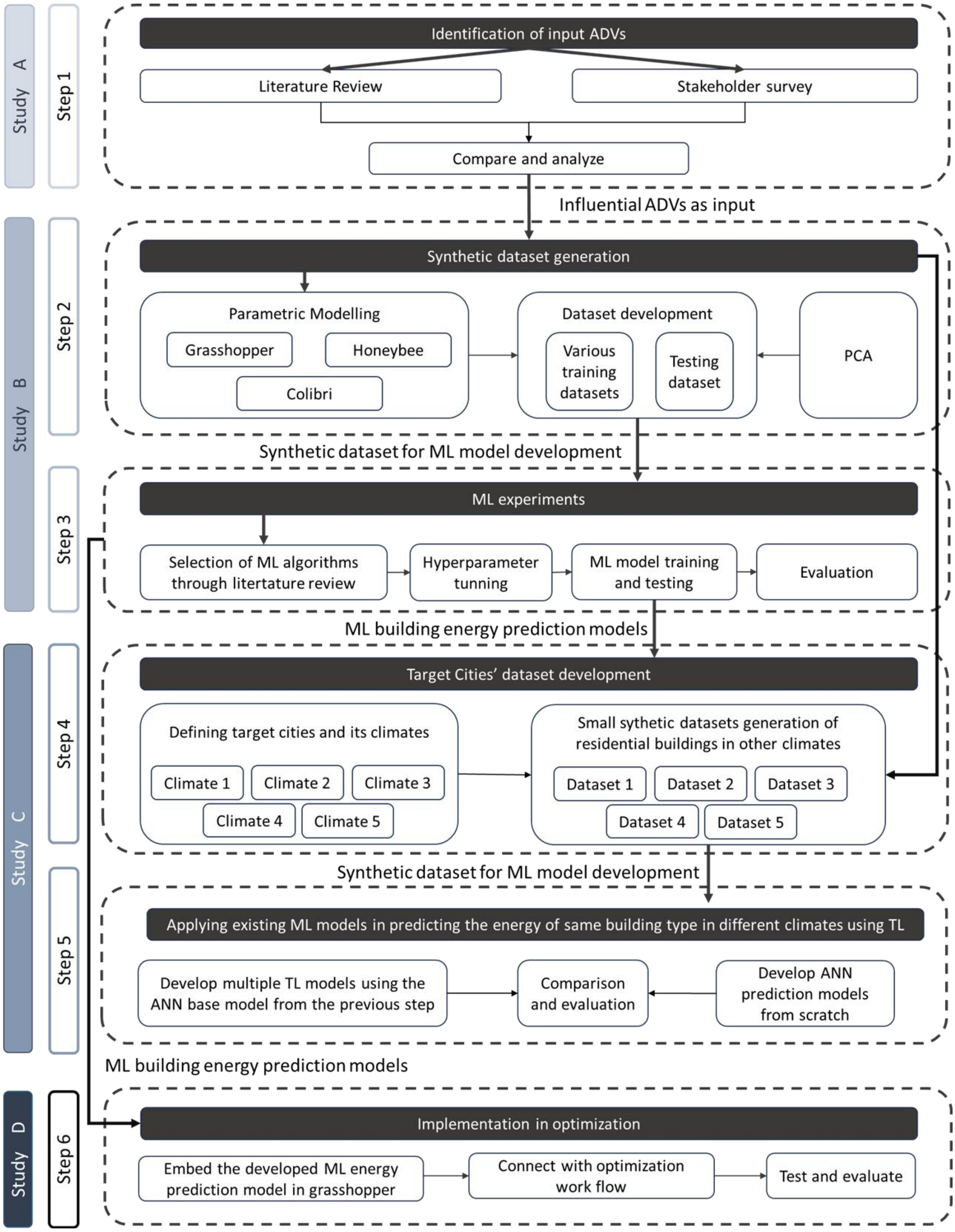


Figure 3.1. Workflow of the methods applied in this thesis.

3.1 Study A: Identification of Input ADVs

The most influential ADVs are investigated by combining a literature review and a stakeholder survey.

3.1.1 Literature review

A literature review is a critical analysis and synthesis of existing literature, research, and scholarly articles relevant to a particular topic or research question [95]. A literature review serves many different purposes and entails a wide variety of activities including the identification of the research gap by examining the inconsistencies in current publications, contextualization of current knowledge, and providing the broader academic discourse [96]. In this thesis, it is mainly used to understand what is going on in the related research field and to identify the trends.

The assumption for the literature review in finding the most influential ADVs for building energy is the ones that are most investigated. If one particular ADV occurs frequently in related literature, it is implied that the ADV holds relatively greater significance. For instance, the 'window-to-wall ratio' is the most frequently encountered ADV when searching for literature focused on early-stage building energy optimization, thus indicating its paramount importance in this context.

The literature review in this thesis was conducted by utilizing two academic search engines: (1) Web of Science, renowned for its comprehensive database of scholarly works and sophisticated search features [29]; and (2) Scopus, esteemed in the realm of architectural research. In Web of Science, the topic search (TS) function was employed, while in Scopus, the Title-Abstract-Keywords search function was applied. The combination of the keywords in the search is shown in Figure 3.2. A two-round article selection was adopted to filter the preliminary results. Duplicate articles and articles from other disciplines that were out of context are removed first. For example, many studies found were from computer science due to the search term "architecture", which can refer to software architecture, among others. The articles that only dealt with certain technical parts of a building, e.g., HVAC system, energy storage system, as well as those that did not focus on newly constructed buildings, e.g., renovations, or historical buildings, were also removed.

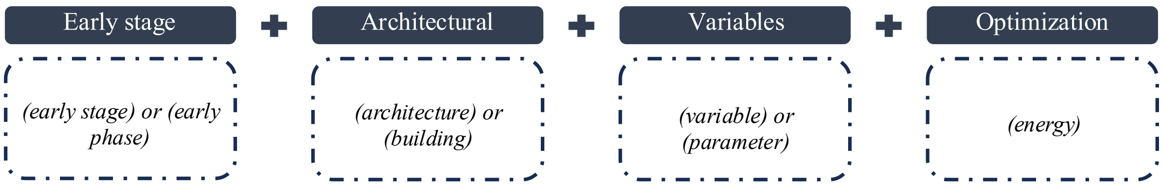


Figure 3.2. Literature search keywords for the most influential ADVs for building energy optimization

3.1.2 Stakeholder survey

The purpose of stakeholder interaction is to ask stakeholders which ADVs they consider important or influential for building energy optimization. There are many approaches for retrieving information from stakeholders, such as focus groups [97], interviews, and surveys. However, both focus group and interview methods require extensive time from participants. Upon contacting several potential participants, it became evident that they were hesitant to commit extensive time helping academic research. Consequently, the survey methodology was adopted at the end. Surveys consist of a series of questions, crafted by a researcher and centered on a particular topic. They are self-administered, with questions presented in a fixed and standard order to all participants [98]. In contrast to other methods, it can be distributed online, reaching a broader and more international audience, and garnering more responses efficiently.

A straightforward rating system was adopted in the survey. Participants were asked to rate the influence of each ADV on a scale from 1 to 5, where 1 means the ADV has nearly no influence, and 5 means the ADV has extremely significant influence. Other than the rating system, the survey also included questions related to the respondent's professional background, work location, years of experience, and job description. An open-ended question asking respondents if there were any other variables that they thought were important were also included. The survey can be seen in the appended paper. It in total included 15 questions and took around 10 minutes to complete. The surveys were distributed between August 2022 to November 2022. They were distributed on LinkedIn and in person in three architecture firms in Sweden. A follow-up interview of 5 to 15 minutes will be initiated if the respondent's survey answers are not clear enough or if the respondent is willing to further explain their answers, in which a free-flowing discussion will be initiated.

24 completed survey responses were collected: 12 from architects and 12 from consultants. The respondents were primarily from Sweden and Norway, and relevant experience varied between 2 and 35 years. All responding architects had experience in residential building design, and the responding consultants all had experience in improving building sustainability. Six follow-up interviews including four with architects and two with consultants were conducted at the end.

3.1.3 Comparison and analysis

To rate the influence that early-stage ADVs had in building energy in the literature review, an indicator that equals the occurrence of an ADV for one objective divided by the total number of papers found looking at building energy was created. For instance, the ADV building plan is mentioned 19 times in the 31 papers looking at building energy, so the rating for the influence of building plan is 19 divided by 31, which is 0.61. The higher the rating value is, the more influential the ADV is. The ADVs appearing at least in half the papers are considered influential, which means the more influential ADVs from the literature should have a rating higher than 0.5. For the survey analysis, the average ratings of all participants were calculated. The ADVs with an average rating higher than 3 are considered as influential as the scale is from 1 to 5.

3.2 Study B: Identification of Algorithms and Dataset for ML Energy Prediction Model.

3.2.1 Synthetic dataset generation

3.2.1.1 Parametric modeling

As previously indicated, parametric modeling enables the efficient generation of datasets with diverse building design alternatives. Since the influential ADVs have already been identified, the first step in the parametric modeling process is to parameterize these ADVs. Given that the definition and representation of ADVs can be broad, different studies may adopt different parameterization strategies. In this thesis, the parameters corresponding to the identified influential ADVs—including their symbols, units, descriptions, value ranges, and step sizes—are presented in Table 3.1. It should be noted, however, that not all parameterized variables are directly used as inputs to the ML prediction model. Several ADVs are represented by derived parameters that are computed within the Grasshopper script rather than explicitly parameterized. These calculated input variables are summarized in Table 3.2.

Table 3.1. Parameters in the GH model

ADV	Parameter	Symbol	Description	Unit	Range	Step	If selected as input
Building plan	Staircase edge	SE	The distance between the first staircase to the edge of the building.	Meter	1-10	1	×
	Staircase width	SW	The width of the staircase.	Meter	3-8	1	×
	Staircase distance	SD	The distance between two staircases.	Meter	30-50	1	×
	Apartment width	AW	The width of each apartment.	Meter	2-8	1	×
WWR	Window room ratio	WRR	The ratio between window width and room width.	×	0.5-0.9	0.1	×
	Window height	WH	Distance between upper and lower edge of window.				×
Building orientation	angle	AG	The degree between the main orientation of the building and the north.	degree	0-360	1	√
Storey number	Storey number	SN	Number of floors.	×	2-7	1	√
Storey height	Storey height	SH	Height of each floor.	Meter	2.5-3.5	0.1	√
Shading device	Shading type	ST	Type of shading: 0 means horizontal shading, 1 means vertical shading, 2 means combination shading	×	0-2	1	√
	Shading length	SL	Width of overhang	Meter	0.1-1	0.1	√
Roof material	Roof U value	RU	U-value of building roof	W/m²·K	0.1-0.6	0.01	√
Wall material	Wall U value	WU	U-value of the exterior building wall	W/m²·K	0.1-0.2	0.01	√

Table 3.2. Transformed parameters in the GH model

ADV	Parameter	Symbol	Description
Building plan	Building area	BA	The total area of each floor in the building
	Building perimeter	BP	The perimeter of each floor in the building
	Circulation area	CA	The total area of the circulation area on each floor, including corridor and staircases
	Circulation perimeter	CP	The perimeter of the circulation area on each floor, including corridor and staircases
WWR	Window-to-wall ratio on north	WWR_N	Fraction of exterior wall above grade covered by fenestration on the north facade
	Window-to-wall ratio on south	WWR_S	Fraction of exterior wall above grade covered by fenestration on the south facade
	Window-to-wall ratio on east	WWR_E	Fraction of exterior wall above grade covered by fenestration on the east facade
	Window-to-wall ratio on west	WWR_W	Fraction of exterior wall above grade covered by fenestration on the west facade

To investigate the diversity of the training dataset's impact on different ML model performances, a GH script that can generate nine shapes of building alternatives was developed (Figure 3.3). The nine building shapes were chosen and adapted from [99]. For each building shape, multiple building configurations with different ADVs were generated.

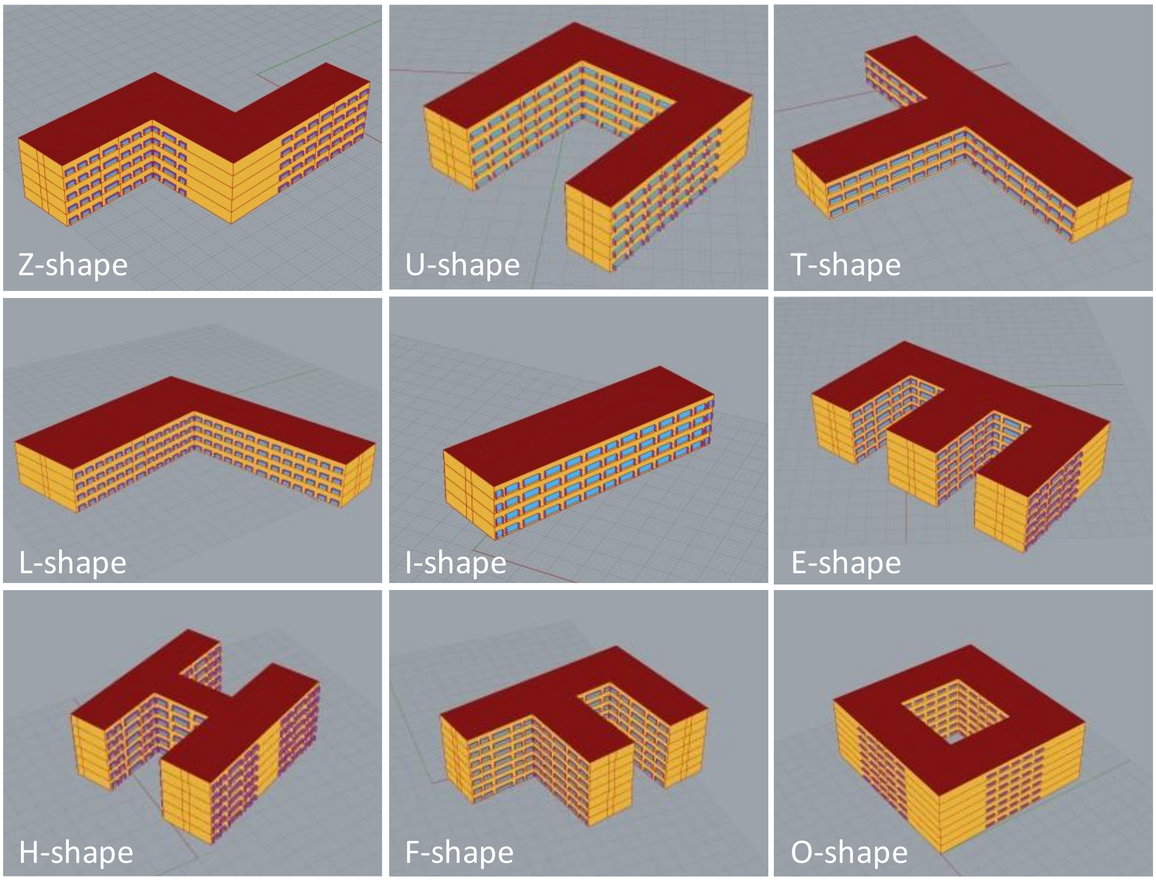


Figure 3.3. Nine building shapes generated in parametric modeling.

Figure 3.4 presents the part of the parameters used in this study representing building plan for various building shapes. Table 3.3 presents the range and unit of the parameters. It is worth noting that these parameters are not the ones fed into the ML model as input values. The parameters representing building plan that serve as input for the ML model are building area, building perimeter, circulation area, and circulation perimeter. The calculation method and explanation for these parameters are presented in Table 3.3.

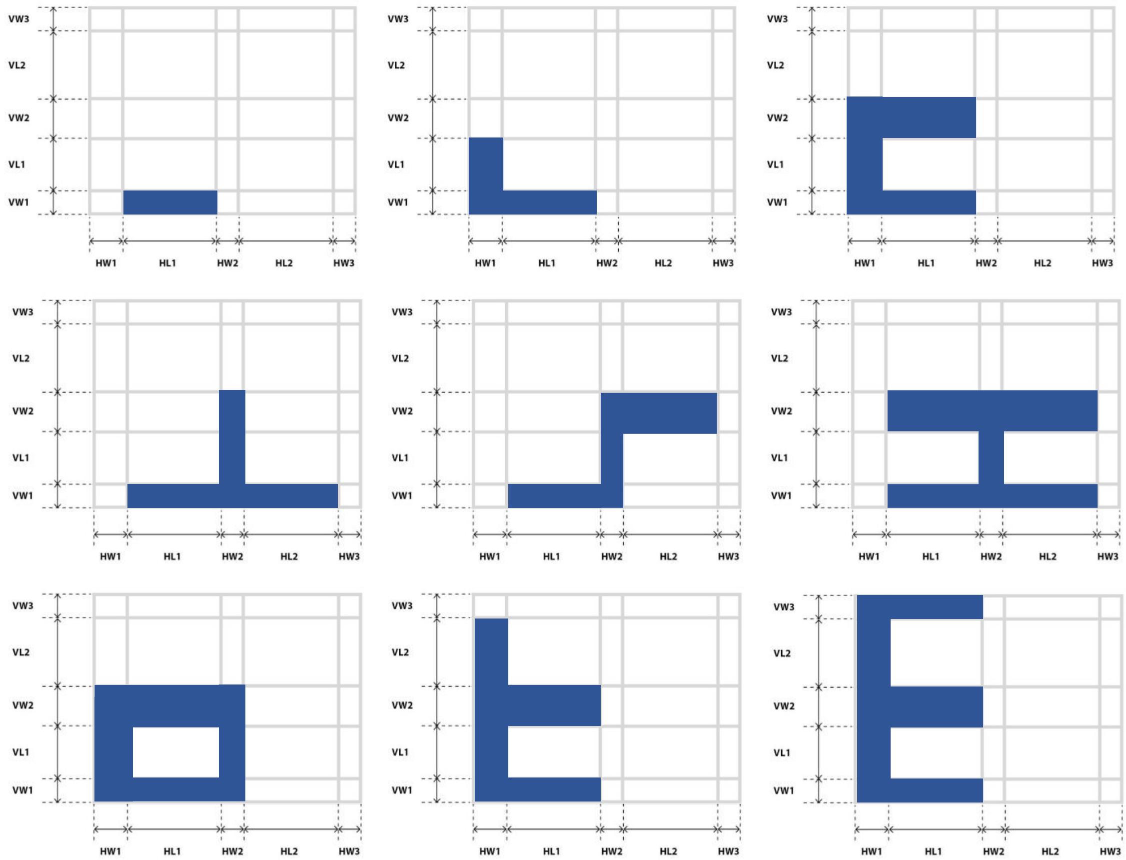
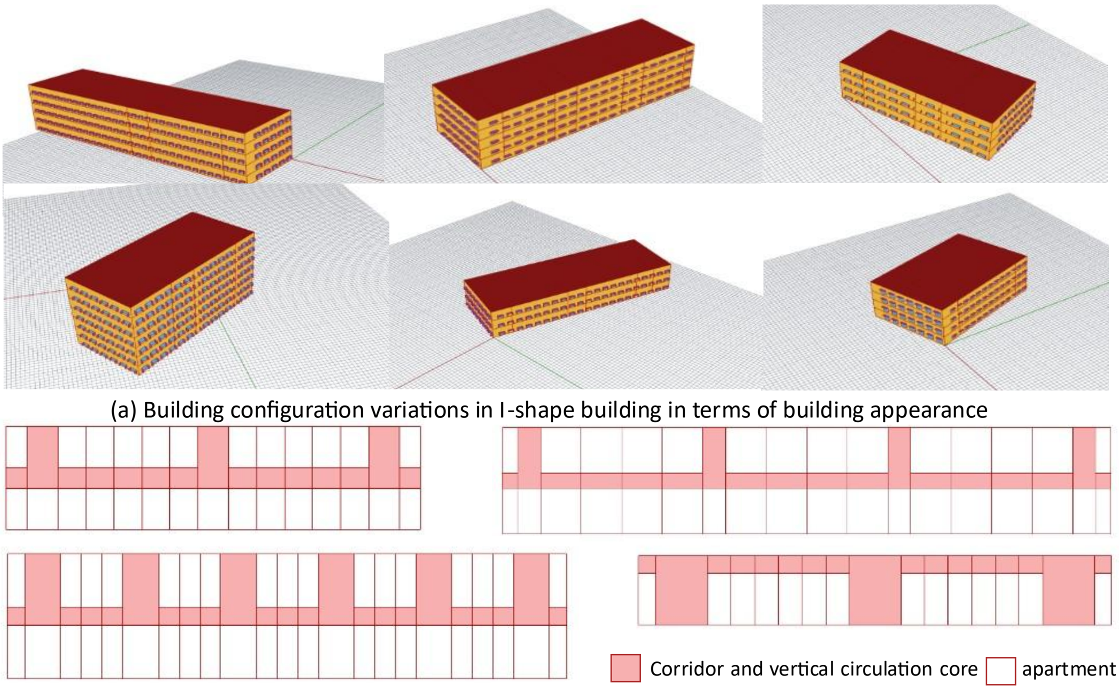


Figure 3.4. Parameters used in representing building plan for various building shapes.

Table 3.3. Parameters related to building plan for different building shapes and their range (Unit: meter, Step:1)

	I-shape	**L-shape**	**U-shape**	**T-shape**	**Z-shape**	**H-shape**	**O-shape**	**F-shape**	**E-shape**
HW1	×	5-25	5-15	×	×	×	10-20	10-20	10-20
HW2	×	×	×	5-20	5-20	5-20	10-20	×	×
HW3	×	×	×	×	×	×	×	×	×
HL1	20-100	15-70	15-40	15-40	15-40	10-30	15-30	15-30	15-30
HL2	×	×	×	15-40	15-40	10-30	×	×	×
VW1	5-30	5-25	5-20	×	5-20	5-20	10-20	×	10-20
VW2	×	×	5-20	5-20	5-20	5-20	10-20	10-20	10-20
VW3	×	×	×	×	×	×	×	10-20	10-20
VL1	×	15-70	10-30	15-40	5-15	8-25	15-30	15-30	10-15
VL2	×	×	×	×	×	×		10-15	10-15

Figure 3.5 takes the I-shape building as an example to show how the building configuration, including building appearance and building plan layout, can vary within one building shape.



(a) Building configuration variations in I-shape building in terms of building appearance

(b) Building configuration variations in I-shape building in terms of building layout

Figure 3.5. Building configuration variations in I-shape building, including building appearance and layout.

Honeybee was then used to generate energy model for each building design alternative and create the input data file (IDF). Colibri was used to iterate through all ADV combinations and compile the resulting data into a CSV of input values (ADVs) and output values (IDF files) per iteration. All generated IDF files were later imported into a Python script that uses EnergyPlus to simulate the energy result for various building alternatives in parallel. The Energy Plus settings were based on a typical residential building scenario in Gothenburg, Sweden [100]. The model used a standard heating system for common Swedish residential buildings instead of an individual heating system. There are two types of thermal zones on each floor, the living area thermal zone and the corridor zone, which includes the vertical circulation core and corridors. The living area zone's heating set point is 21°C and the unoccupied zone's set point is 10°C. Since the location is in Sweden, there is no requirement for activating a cooling system. The unit of energy demand is kWh/m2. More settings for the energy simulation can be seen in Appendix B.

3.2.1.2 Synthetic dataset development and principal component analysis (PCA)

The final complete dataset contains the values of different ADVs as input and takes the annual energy demand per square meter from Energy Plus as output. Figure 3.6 presents the workflow of formulating the testing dataset and different training datasets.

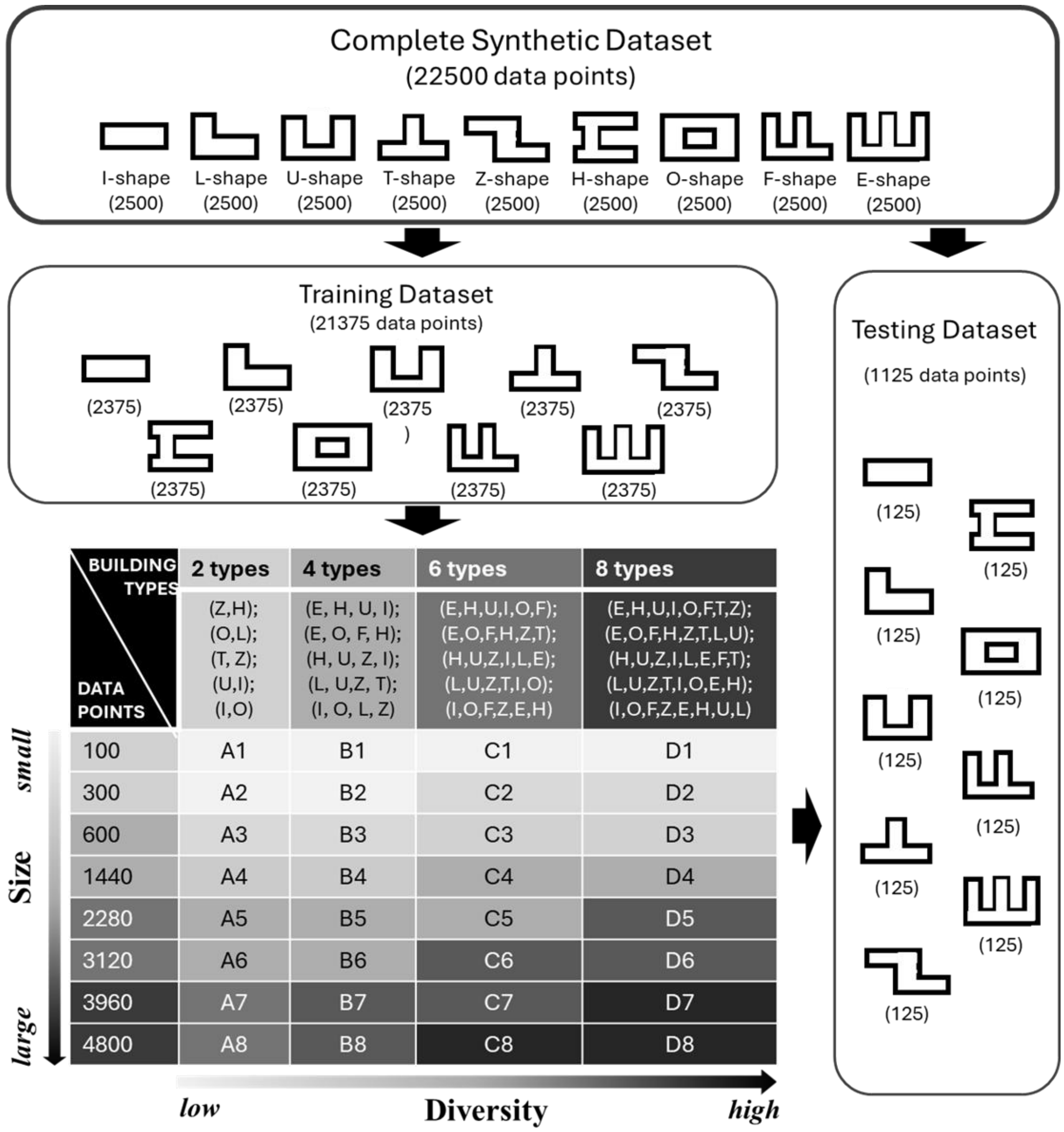


Figure 3.6. Workflow of synthetic dataset development.

The complete synthetic dataset contains a total of 22500 data points of different building alternatives within nine types. The testing dataset contains 125 data points selected from each building type randomly, in total 1125 data points. To investigate the influence of training dataset size and diversity on the ML model's performance, multiple training datasets were formulated from the rest of the data points in the developed dataset. 32 training dataset categories were developed, from A1 to D8. The letter indicates the diversity of datasets while the number indicates the size of the datasets. The A1 dataset category has only 2 building types and in total contains 100 data points while the D8 dataset category has 8 building types and contains 4800 data points. In each dataset category, the number of data points of each building type is the total data points divided by the included building types. For instance, in the dataset category C3, the total data points are 600, and 6 building types are included, the data points in each building type are 600 divided by 6, which is 100. Five datasets with different selections of building types were formulated and applied in each category. The final performance for each training dataset category was calculated by the average performance for the five datasets.

PCA was used to reduce the data features and visualize the diversity in datasets. PCA is a statistical approach for identifying principal features based on the total variance [101]. It is a general-purpose dimension reduction and data analysis tool, which is mainly used in the machine learning field [102]. It generates linear combinations of original features that are capable of projecting original data on a reduced dimensional space [103]. PCA works by transforming the original features into a new set of orthogonal features called principal components (PC). The PCs are orthogonal to each other and capture the maximum amount of variance in the data [104]. By retaining only the two PCs, the dataset can be easily visualized while preserving most of the important information since the PCs are linear combinations of uncorrelated attributes that best describe the variance among data. Figure 3.7 presents the visualization of the diversity of the complete synthetic dataset (a) and the testing dataset (b) by using PCA to show the distribution of the data points. The further the distance between two data points, the greater the diversity gap between them.

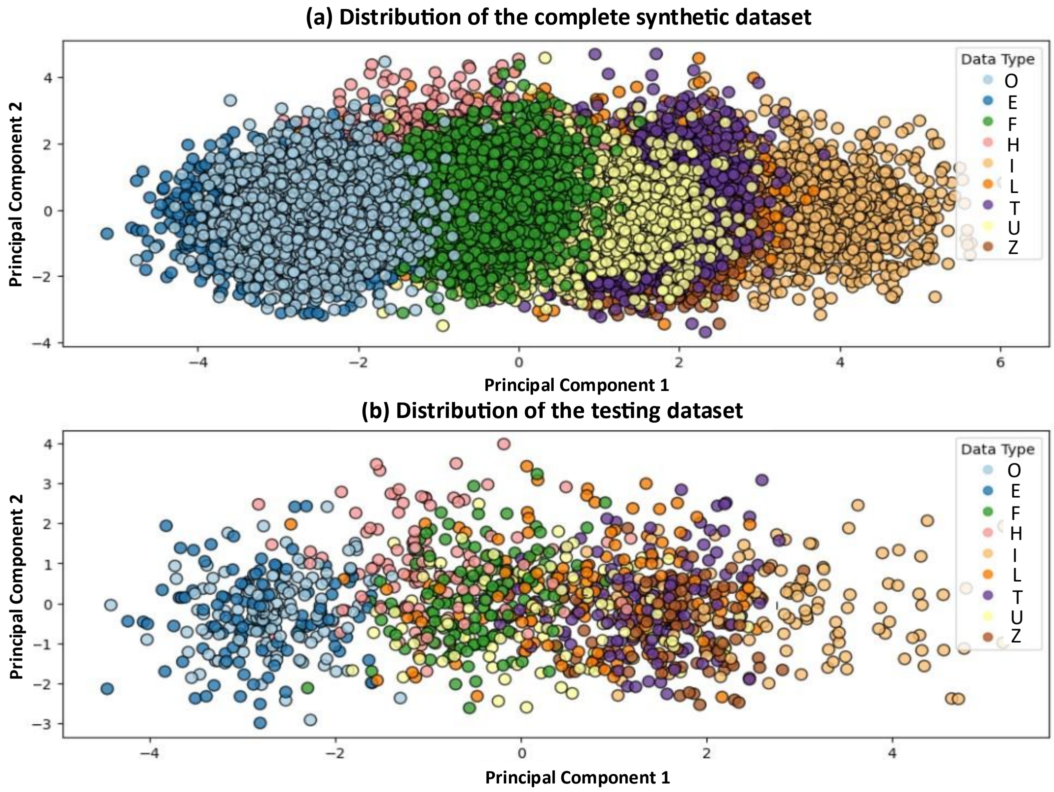


Figure 3.7. Distribution of the complete synthetic dataset (a) and the testing dataset (b) using PCA.

Figure 3.8 shows examples of the distribution of the testing dataset and different training datasets. It can be seen how different training datasets vary regarding size and diversity, and how much they cover the testing dataset.

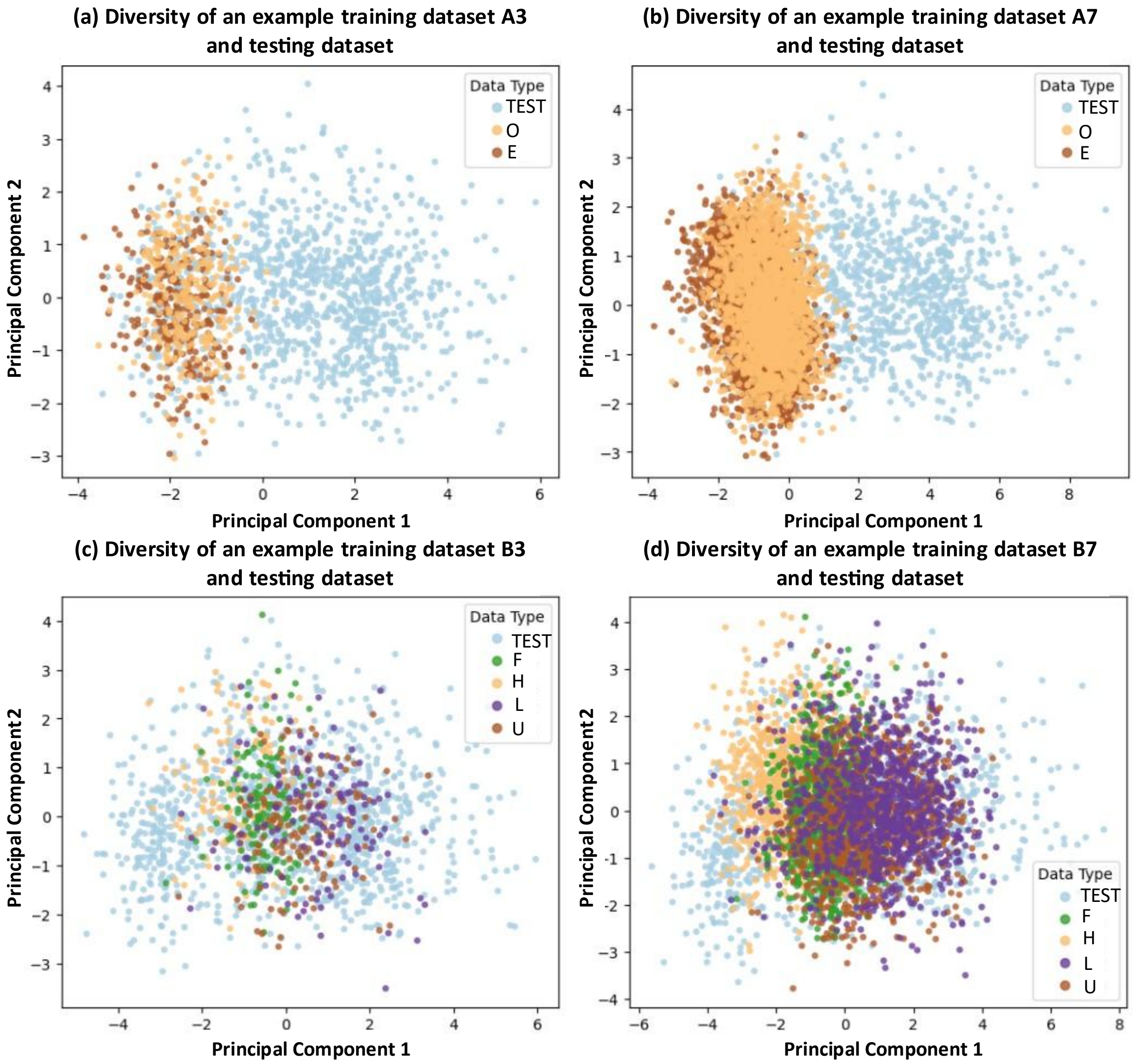


Figure 3.8. Distribution of examples of A3 (a), A7 (b), B3 (c), and B7 (d) training dataset categories and testing dataset.

3.2.2 ML model development for building energy prediction.

Various training datasets developed in the previous step were used to train different ML models with multiple algorithms. The performance of developed models is evaluated in terms of accuracy and computational efficiency. Through the literature review, five algorithms were selected in the ML model development, including Artificial Neural Network (ANN), Support Vector Machine (SVM), Decision Tree (DT), Random Forrest (RF), and Multiple Linear Regression (MLR). The reasons for choosing these five algorithms and their respective introductions are in Chapter 2. The models were developed in the Python 3.9 environment using Scikit-learn.

(1) Hyperparameter tuning

To generalize diverse data structures, most ML models require specific constraints or learning rates or weights [105]. These values set within the model are known as hyperparameters. Default hyperparameter settings cannot guarantee optimal performance and additional attention should be directed to this critical step. To reach robust performance results with these ML models, their respective hyperparameters must be optimized. The process of choosing a group of optimal hyperparameters for an ML algorithm is known as hyperparameter tuning. Hyperparameter tuning is performed for ANN, SVM, DT, and RF in this thesis. For MLR, no tuning is required as MLR has no hyperparameters and assumes a logit relationship between response and predictors.

(2) Cross-validation

Cross-validation is a resampling-based technique for the estimation of a model's predictive performance. The basic idea is to split the given dataset into multiple segments using a user-defined number of partitions. The dataset is split into k folds first, and k-1 sets are used as training datasets while the remaining one is used as the validation dataset. This process is repeated k times by changing the validation dataset [106]. Cross-validation provides a more reliable estimate of model performance compared to a single train-validation split. Applying cross-validation in hyperparameter tuning can also help mitigate overfitting by providing a more robust estimate of model performance across different subsets of the data [107]. In this thesis, the training dataset is split into ten folds to perform cross-validation in hyperparameter tuning.

(3) RandomizedSearchCV

RandomizedSearchCV is a popular technique for hyperparameter tuning in Scikit-learn. In contrast to GridSearchCV, which does an exhaustive search over specified parameter values for an estimator, not all parameter values are tried out in RandomizedSearchCV, but rather a fixed number of parameter settings is sampled from the specified distributions. RandomizedSearchCV is more time-efficient than GridSearchCV. However, the most optimal hyperparameters retrieved from RandomizedSearchCV might not be as optimal as the ones retrieved from GridSearchCV. The number of parameter settings that are tried is given by users. In this thesis, it is set that 40 random combinations for all parameters are tried, and the best-performing ones are considered the most optimal parameters.

3.2.3 Evaluation methods

The developed ML models are evaluated from two perspectives: accuracy and computational efficiency.

(1) Accuracy

When working with ML models, it's crucial to evaluate performance quantitatively using evaluation metrics. These metrics offer objective and measurable insights into a model's predictive accuracy, especially when ML models are always treated as 'black box' as they can be used without requiring people to understand what is going on within the model. For regression tasks, the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) are the most common and well-studied metrics.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \widehat{y_i}| \qquad (1)$$

Where n= the number of samples; i= the ith sample; yi=the true value of the ith sample's target value; $\widehat{y_1}$= the predicted value of the ith sample's target value.

The MAE (1) subtracts the predicted value from the true value for each sample, removes the direction of the error, and takes the mean across the entire test set. In this way, the MAE retains the scale of the target variable, which allows for an intuitive interpretation of the output. For example, an MAE of 2.35 for this study means that the model makes an average error of 2.35 kWh/m2 when predicting the normalized district heating demand.

$$RMSE = \sqrt{\sum_{i=1}^{n}\frac{(y_i - \widehat{y_i})^2}{n}} \qquad (2)$$

Where n= the number of samples; i= the i_{th} sample; y_i=the true value of the i_{th} sample's target value; $\widehat{y_1}$= the predicted value of the i_{th} sample's target value.

Similar to the MAE, the RMSE (2) subtracts the predicted value from the true value; however, it squares the error before taking the mean across the test set. The root square is added at the end to adjust the result to the same scale as the target variable. This results in a greater weight placed on larger errors, which can help detect outliers and substantial deviations from the ground truth.

(2) Computational efficiency

Computational efficiency is an essential evaluation metric as the main purpose of using ML to develop a building energy prediction model is to accelerate the speed of early-stage optimization. Computational efficiency is evaluated from two aspects: the time used for hyperparameter tuning during the ML model development and the time needed for a model to predict the entire testing dataset.

3.3 Study C: Exploring ML models' Generalizability.

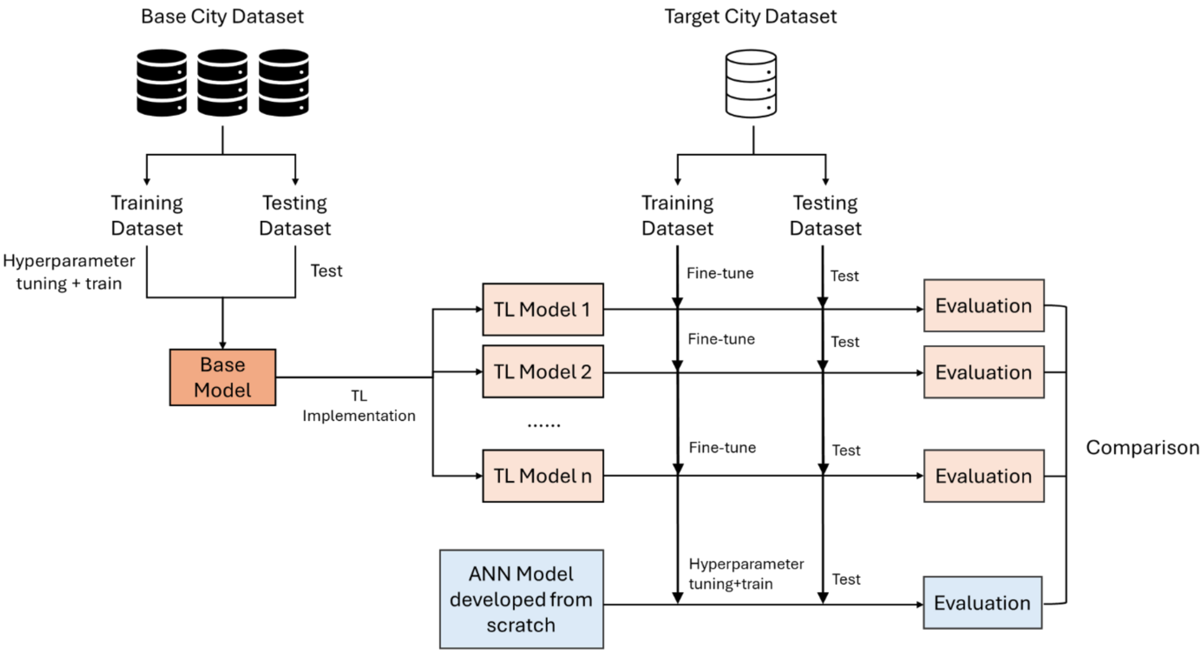


Figure 3.9. The main workflow of applying and evaluating TL in Study C.

The primary research methodology and workflow of Study C are illustrated in Figure 3.9. A base ML model trained from the previous step will be integrated here to ensure robust initial performance. Then, multiple TL models will be developed based on the pre-trained model and fine-tuned using the training dataset from a target city. Subsequently, the testing dataset from the same test city is used to test and evaluate the fine-tuned TL model. Meanwhile, an ANN model will also be developed from scratch using the insufficient data from the target city. The new ANN model will be used as a comparison to evaluate whether applying TL can improve the model's performance.

3.3.1 Development of base models

The base model was developed to predict heating energy demand for residential buildings in Gothenburg, Sweden, following the logic in Section 3.2. The training dataset is part of the synthetic dataset developed in Section 3.2.1, including building shape Z, U, T, L, I, and H. The input features for the base ML model are the same as listed in Section 3.2.1.1. The target prediction output of the base models is the annual heating energy demand per square meter (kWh/m^2). Energy Plus was used to simulate the annual energy demand for each building configuration. The base model was developed using ANN, as mentioned in Section 2.4. The model structure, training dataset size, and model performance, including root mean square error (RMSE), mean absolute error (MAE), R-squared (R^2), and mean absolute percentage error (MAPE), are listed in Table 3.4. The explanation for model performance metrics can be further seen in Sections 3.2.3 and 3.3.3.

Table 3.4. Characteristics of the Base Model

Training Dataset Size	Structure	RMSE (kWh/m^2)	MAE (kWh/m^2)	R^2	MAPE
12000	14-20-28-14-1	1.97	1.23	0.97	1.51%

3.3.2 Target cities' dataset development

To investigate the generalizability of TL for building energy prediction across diverse climatic conditions, five representative cities are selected based on the Köppen climate classification system [108], which categorizes global climates according to temperature and precipitation patterns. The Köppen system is one of the most widely used methods for categorizing global climates, combining temperature and precipitation patterns to define major climate zones such as tropical (A), arid (B), temperate (C), continental (D), and polar (E). Each major category is further subdivided based on seasonal characteristics. Due to its simplicity, spatial coverage, and climate-relevant grouping, the Köppen classification is widely used in building energy and environmental studies to distinguish climatic influences on energy demand [109], [110]. Following this system, we selected five target cities across different climate zones to represent a broad climatic spectrum. We excluded zone E as there is no major city with a high population in zone E. The selected target cities are listed as follow:

- **Miami, USA (Am)** – Tropical monsoon climate with high temperatures and no heating demand required, dominated by cooling loads.
- **Madrid, Spain (Bsk)** – Hot-summer Mediterranean climate with dry summers and mild winters, creating a mixed demand for both heating and cooling.
- **Seattle, USA (Csb)** – Warm-summer Mediterranean climate with wet winters and mild summers, resulting in moderate year-round energy use.
- **Chicago, USA (Dfa)** – Cold winters and warm to hot, often humid summers, exhibiting high heating demands for much of the year, with moderate but non-negligible cooling requirements during the summer season.
- **Stockholm, Sweden (Dfb)** – Warm-summer humid continental climate, characterized by significant heating needs, between the maritime temperate and continental subarctic climates.

The source domain city used for the pre-trained ANN model is:

- **Gothenburg, Sweden (Cfb)** – Temperate oceanic climate featuring mild winters and cool summers. The dominant energy demand type for Gothenburg is heating energy.

Figure 3.10 presents the global distribution of climate zones based on the Köppen classification, along with the climate types of the selected target cities and their relative similarity to the source city.

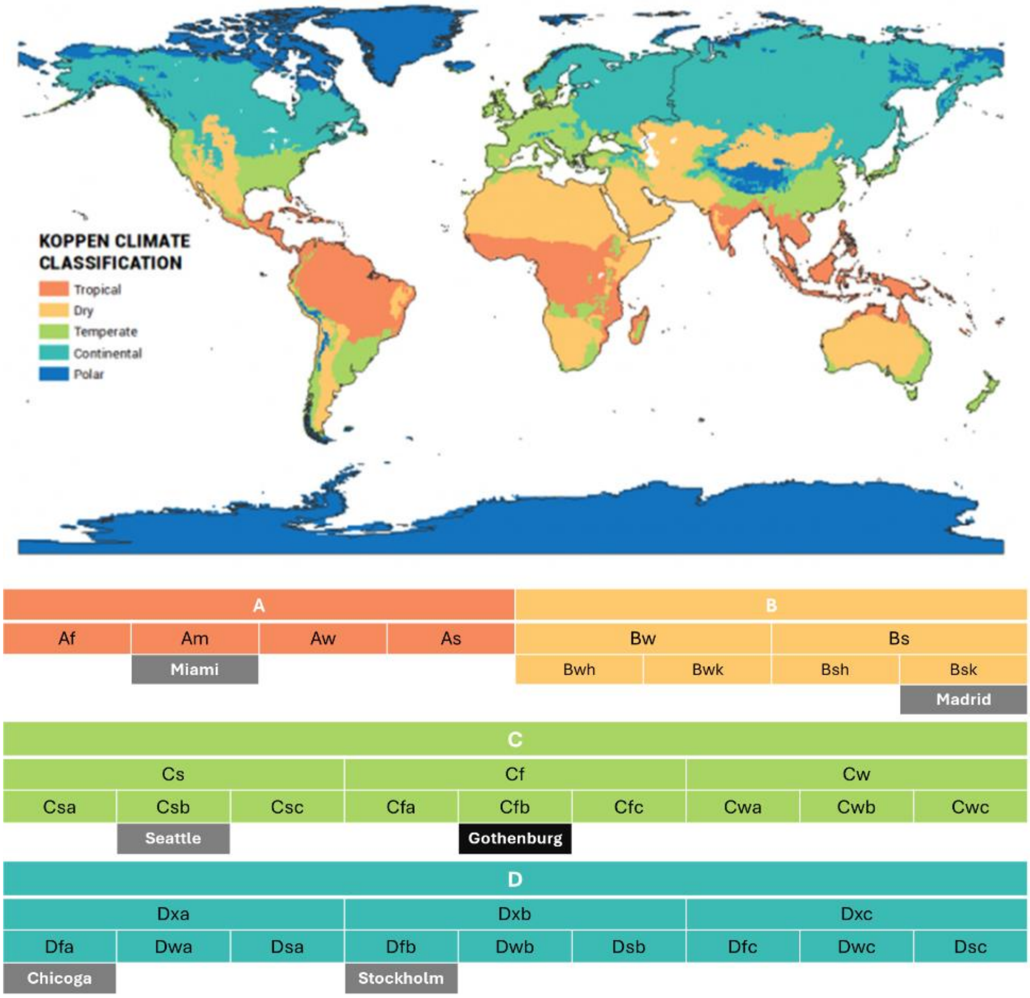


Figure 3.10. Global climate zones based on Köppen climate classification (adapted from [108]) and selected target cities.

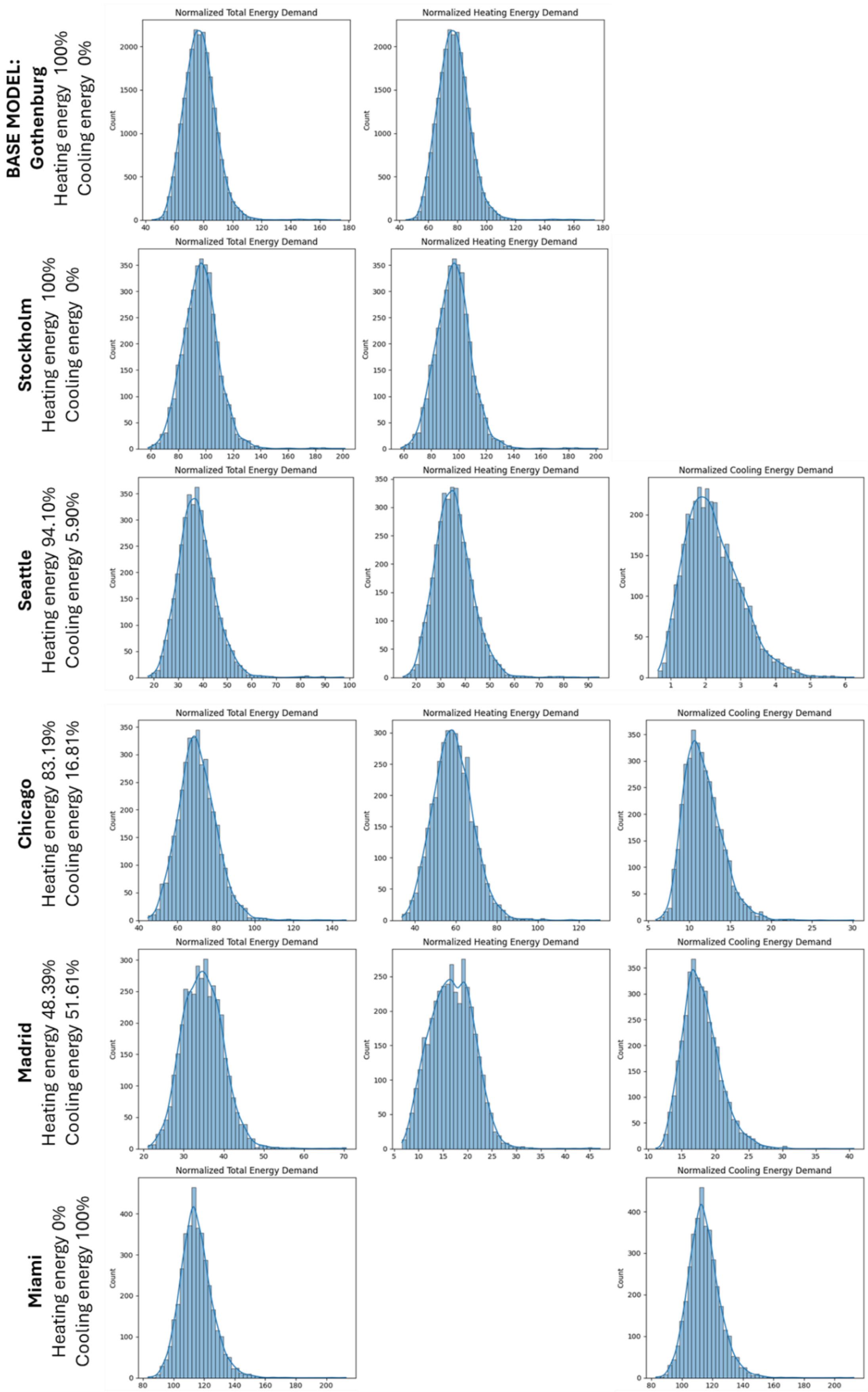


Figure 3.11. The distribution of the energy demand of the six selected cities.

For the target cities, the synthetic datasets were generated using the same parametric modeling approach employed in the database development for the base scenario, as shown in section 2.1. The building shapes are identical to those in the base model scenario. Appendix C lists Energy simulation settings for each target city's buildings. The developed datasets were further formulated into various sizes for the experiment settings, from 200 to 2000 data points. Figure 3.11 presents the distribution of energy demand, including heating energy and cooling energy, of all data points in the six cities.

3.3.3 Transfer learning implementation

There are multiple approaches to implementing TL. For a neural network model, the easiest and most typical way to apply TL is to copy the base network model's first n layers to the first n layers of the target model [111]. The copied layers will be frozen and do not need to be trained with new data. The remaining layers of the target network will then be randomly initialized and trained using the target train dataset in the target task [10]. In this study, the features of the new models and the base models are the same as the process tends to work if the features are general, meaning suitable to both base and target tasks, instead of specific to the base task [112]. Therefore, the first selected TL method is to freeze the first n layers and retrain the rest with new data. Based on this approach, we developed three TL models by freezing different numbers of layers from the original ANN model: the first layer (A-1), the first two layers (A-2), and the first three layers (A-3). The implementation is illustrated in Figure 3.12.

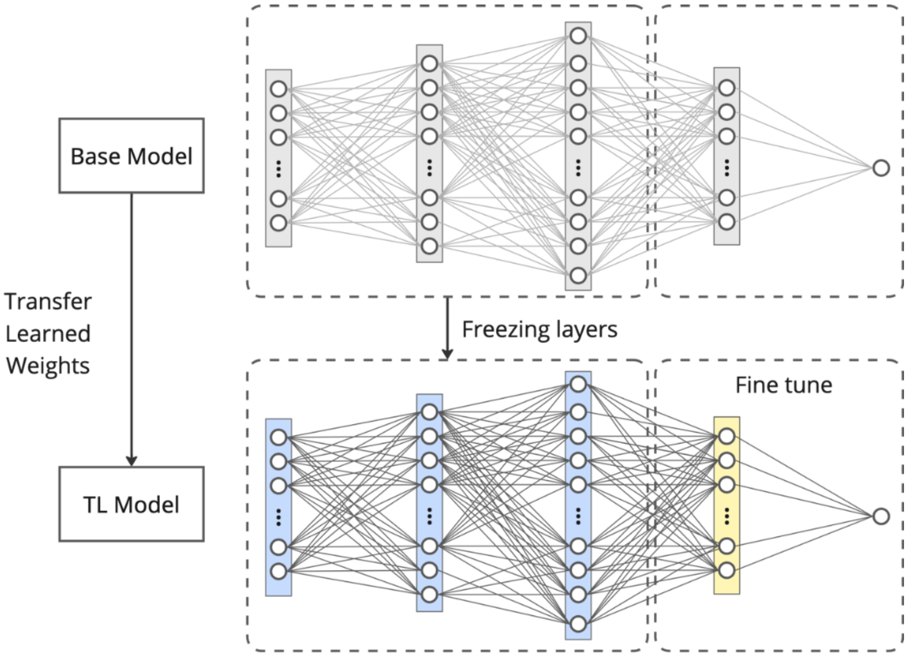


Figure 3.12. TL implementation method for model A-1, A-2, A-3.

One efficient way to mitigate the issue of the discrepancy in data distributions between source and target cities caused by differences in climate conditions and energy usage patterns is to employ a domain-adaptive transfer learning framework using a Gradient Reversal Layer (GRL) embedded in the neural network architecture [113]. Therefore, the second selected TL method integrates a GRL into the existing ANN architecture. The GRL passes features forward unchanged, but during backpropagation, it reverses gradients associated with domain-specific signals, acting as an implicit regularizer. This discourages overfitting to city-specific patterns and promotes the extraction of domain-invariant features, improving the model's generalization ability across cities [114]. Without requiring an explicit domain classifier or labeled data from the target city, this approach enables effective knowledge transfer under different climatic conditions. Following the GRL, we added several additional fully connected layers to form the task-specific regression head, which refines domain-invariant features into accurate energy predictions. These layers are fully trainable and preserve the model's capacity to learn expressive, task-optimized representations while benefiting from adversarial domain adaptation. Based on this approach, we developed three TL models by freezing different numbers of layers from the original ANN model: the first layer (B-1), the first two layers (B-2), and the first three layers (B-3). For each model, we added three more layers after the GRL: the first layer contains 20 neurons, the second 16 neurons, and the third 8 neurons. The implementation is illustrated in Figure 3.13.

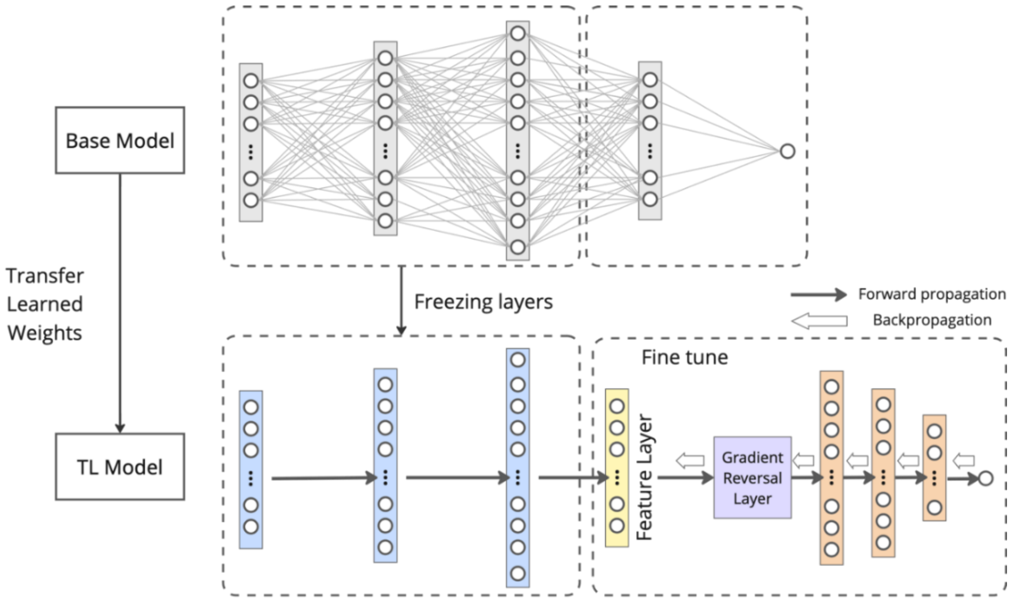


Figure 3.13. TL implementation method for model B-1, B-2, B-3.

It is worth noting that if the target dataset is large or the number of features is small, overfitting becomes less of a concern. Although the target dataset is not considered large in this study, there are only 14 features, which is quite small, so overfitting is not considered a problem. The TL models were developed in the Python 3.9 environment using Scikit-learn. For each TL model development, five experiments were conducted, and the results are the average value of all experiments. Table 3.5 summarizes the details of the experimental setting, including the TL models, how many layers they freeze, and the training dataset sizes.

Table 3.5. TL experiment setting.

Method	**A. Freezing Layers**			**B. GRL+New Layers**			**C. Retrain Model**
Name	A-1	A-2	A-3	B-1	B-2	B-3	C
Frozen Layer Number	1	2	3	1	2	3	×
Base Model Training Data	12000						×
Target City Training Data	200, 400, 600, 800, 1000, 1200, 1400, 1600, 1800, 2000						

3.3.4 Evaluation method

The primary logic of the evaluation is to apply TL to the base models on the new tasks' datasets to assess whether this approach improves the prediction accuracy compared to developing a new ML model from scratch using the dataset of the new task. The three evaluation metrics selected in this study are root mean squared error (RMSE), determination coefficient (R^2), and mean absolute percentage error (MAPE). Apart from RMSE, the description could be seen at Section 3.2.2, the description of the rest two metrics can be seen below. Time efficiency is also included as an evaluation metric as reducing model developing time is one of the main reasons to apply TL to begin with.

(1) R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{3}$$

Where n= the number of samples; i= the i_{th} sample; yi=the true value of the i_{th} sample's target value; $\hat{y}_i$= the predicted value of the i_{th} sample's target value; $\bar{y}$ = the mean of the actual values.

The R^2 indicates the proportion of the variance in the dependent variable that is explained by the independent variables in the model. A higher R^2 value means that the model explains more of the variability in the data. When R^2 equals 1, it indicates all predicted values match the actual values; when R^2 equals 0, it indicates all model explains none of the variance in the data, and predictions are no better than using the mean of the actual values; when R^2 is negative, it indicates the model fits the data worse than simply predicting the mean of the data.

(2) MAPE

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \times 100 \tag{4}$$

Where n= the number of samples; i= the i_{th} sample; y_i=the true value of the i_{th} sample's target value; $\hat{y}_i$= the predicted value of the i_{th} sample's target value.

MAPE calculates the average of the absolute percentage differences between predicted and actual values and is always expressed as a percentage. A lower MAPE indicates a more accurate model. For example, an MAPE of 3% means the model's predictions are off by an average of 3% from the actual values.

3.4 Study D: Exemplify the Implementation of ML Models.

The ML heating energy prediction model developed in Study B is tested in an optimization workflow using a case study to compare its performance with that of an optimization workflow utilizing a simulation engine. Two tasks are designed in this study to exemplify how the ML models can accelerate the optimization. Task 1 is a single-objective optimization task, which optimizes building energy demand only. Task 2 is a multi-objective optimization task that optimizes both building energy demand and embodied carbon emissions. Embodied carbon emissions is this study refer to greenhouse gas emissions related to material production in life cycle module A1-A3 according to EN15978 [115]. The embodied carbon factors are taken from the Swedish database [116].

Both tasks use the same case study. It should be noted that evaluating the best optimization algorithms is not the primary focus of this study; rather, the aim is to exemplify applying ML models in optimization workflow and demonstrate the potential of using ML in supporting early-stage energy design decisions.

3.4.1 Case study description

The selected case study is a residential building project designed by Kaminsky Arkitektur. The project is located in Björlanda, Gothenburg, Sweden. The design concept features a centralized circulation area and four apartments, each with a corner location. The length of the initial design is 32.3 meters, and the width is 15.8 meters. There are in total four floors, and each floor has four apartments, two of which are 87 square meters and two are 70 square meters. There is an equipment room that is 30 square meters. The initial building plan can be seen in Figure 3.14.

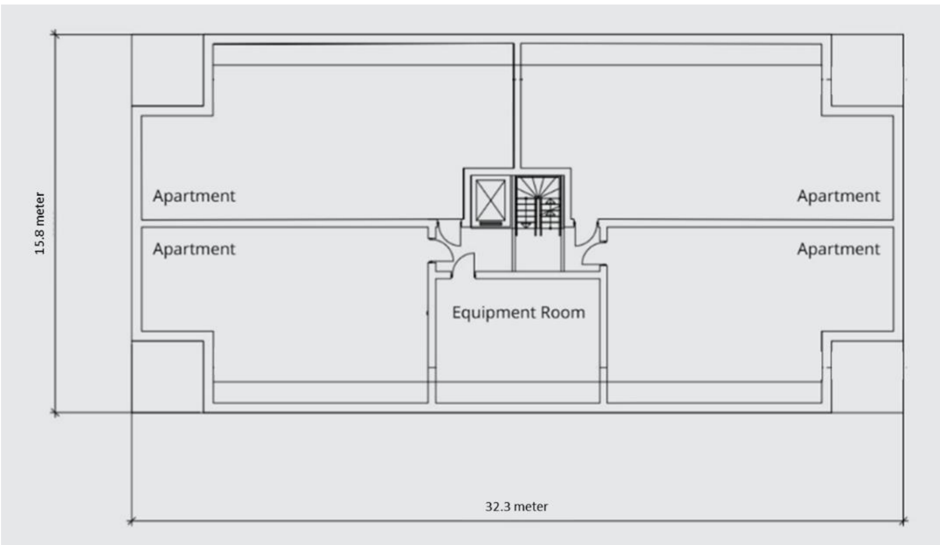


Figure 3.14. Initial plan of the case study project.

The heating setpoint for the living area (apartments) is 21℃, the heating setpoint for unoccupied areas (staircases and equipment room) is 10 ℃. The rest of the energy setting is listed in Appendix B. The initial energy demand for this project is 99.53 kWh/m^2.

3.4.2 Parameter setting

The aim of the early-stage optimization is to optimize the design to achieve a better sustainability performance, but at the same time, keep the architect's design concept. As mentioned in Section 1, the first step of the optimization is to use parametric modelling to generate numerous building design configurations. In this case, the building project is parameterized in Grasshopper, Rhino. Table 3.6 presents the ADVs, the parameters that control them, both in the initial design and the range in the optimization.

Table 3.6 Parameter setting.

ADV	Parameter	Description	Unit	Value in the initial design	If fixed in optimization	Range	Step
Building Plan	Building Area	The total area of each floor.	m^2	475	√		
	Building width	The width of the building.	Meter	15.8	×	8-18	0.1
WWR	Window height	The height of each window.	Meter	0.6	√		
	Windon length/ wall length	The ratio of window length to wall length.		0.5	×	0.2-0.8	0.1
Building orientation	angle	The angle between the main orientation of the building and the north.	degree	24.5	√		
Storey	Storey number	Number of floors.		4	√		
	Storey height	Height of each floor.	Meter	3	×	3-3.5	0.1
Shading device	Shading type	Type of shading: 0 means horizontal shading, 1 means vertical shading, 2 means combination shading		0	×	0-2	1
	Shading length	Width of overhang	Meter	0.2	×	0.2-0.8	0.1
Envelope Material	Wall U value	Wall U value	W/m²·K	0.30	×	0.10-0.60	0.01
	Roof U value	Roof U value	W/m²·K	0.15	×	0.10-0.20	0.01
	Floor U value	Floor U value	W/m²·K	0.15	×	0.10-0.20	0.01
	Window U value	Window U value	W/m²·K	1.5	×	1.0-2.0	0.1

For Task 1, the parameters listed in Table 3.3 are sufficient, since the objective is to simulate and predict building operational energy, which only requires the U-values of the envelope materials rather than their specific compositions. In contrast, Task 2 involves calculating embodied carbon emissions, which necessitates more detailed material information. The parameter settings of the ADVs, including Building Plan, WWR, Building Orientation, Number of Storeys, and Shading Device, are consistent with those presented in Table 3.3. The only exception is the parameter setting of the Envelope Material, which differs and is detailed in Table 3.7.

Table 3.7 Envelope material parameter setting for Task 2

ADV	Parameter	range	Step	Unit	Value in the initial design	If fixed in optimization	description
Wall	Index	0-2	1	×	0	×	Different wall composition
	Insulation thickness	100-500	10	mm	300	×	The thickness of the insulation layer in the wall composition
Roof	Index	0-3	1	×	0	×	Different roof composition
	Insulation thickness	100-500	10	mm	300	100-500	The thickness of the insulation layer in the roof composition
Window	Window U value	1.1	×	×	×	√	×

Table 3.8 lists all the wall and roof archetypes used in this study. Each table lists the construction type, its material layers arranged from interior to exterior. These archetypes are retrieved and adapted from Boverket [116], serve as the candidate options for the optimization process in Task 2, ensuring consistent and standardized representation of building envelope configurations.

Table 3.8 Building Envelope Archetypes with Material Layers.

Construction Type	Index	Material Layers
Wall	0	Gypsum board → Mineral wool → Sheathing → Vapor control layer → Timber cladding
	1	Gypsum board → CLT 120 mm → Mineral wool → Timber cladding
	2	Concrete 150 mm → Mineral wool → Concrete 70 mm
Roof	0	Gypsum board → Mineral wool → Sheathing → Bitumen waterproofing
	1	Concrete 200 mm → XPS → Bitumen waterproofing
	2	Metal → Mineral wool → Bitumen waterproofing
	3	Concrete 200 mm → Bitumen waterproofing → XPS insulation → 50 mm screed

The optimization tasks keep the building design concept while optimizing the building shape and envelope. Figure 3.15 presents the three design configurations in the parametric model with different parameter settings as examples.

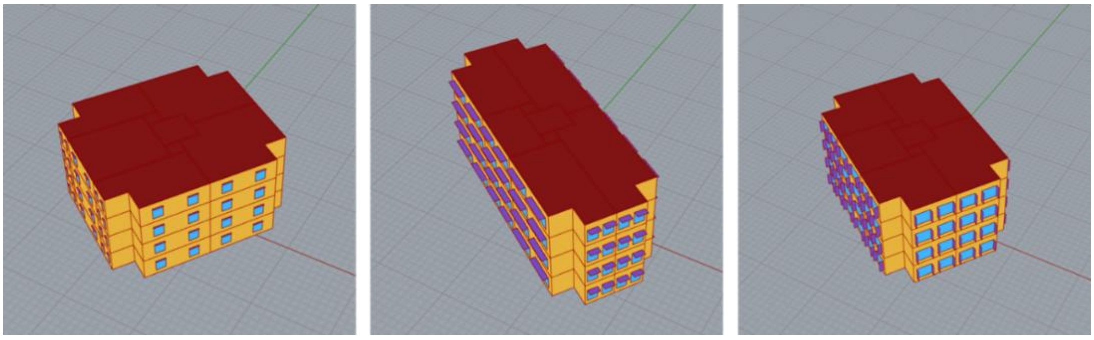


Figure 3.15. Three example configurations in parametric modelling.

3.4.3 Optimization workflow development.

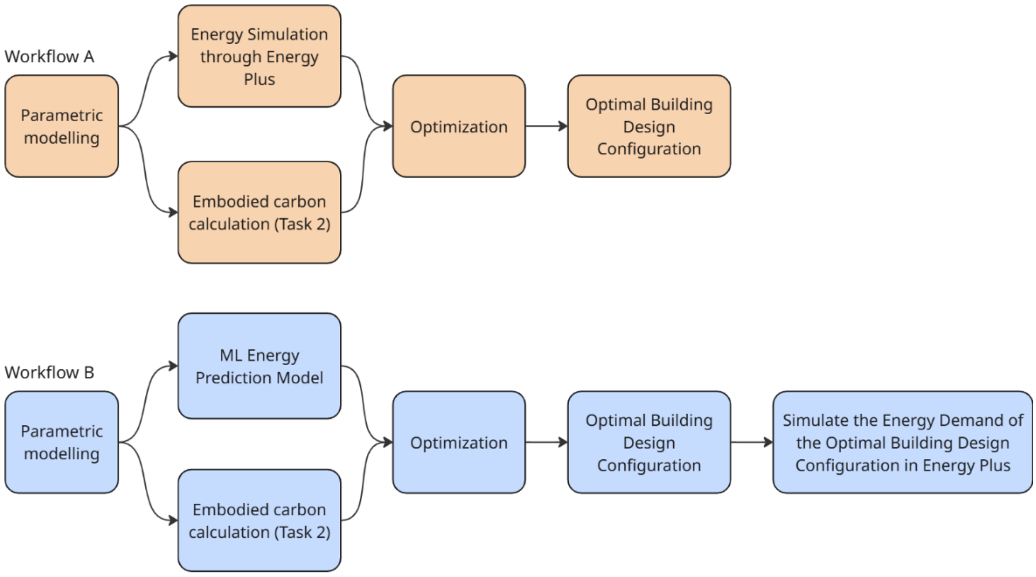


Figure 3.16. The composition of an ML model for building energy prediction.

Figure 3.16 presents the two workflows applied in this study. The two workflows are developed based on the concept presented in Figure 1.1. Workflow A represents the current most common optimization workflow that uses the simulation engine. In contrast, workflow B represents the proposed optimization workflow in this thesis that uses an ML model for more time-efficient predictions. Both workflows are developed in Grasshopper, Rhino. The two workflows are nearly identical; the only difference lies in the handling of the final evaluation. Since the prediction values generated by the ML model do not perfectly match those from the actual simulation engine, Workflow B includes an additional step at the end—re-simulating the most optimal building design configuration obtained from the optimization process using the original simulation engine.

The simulation in workflow A is achieved by using Honeybee, which uses the Energy Plus simulation engine. The ML prediction model is an ANN model that was developed in Study B and applied in Study C. The details of the ANN model can be seen in Table 3.1 in Section 3.3.1. The ML model is integrated into the parametric design workflow using a custom Flask-based inference server. First, a local API server was built using Flask, TensorFlow, and supporting Python libraries to host the trained ANN model. This server receives input parameters in JSON format and returns prediction results via HTTP POST requests. To connect the server with the Grasshopper environment, a GHPython scripting component was developed, allowing the user to send the input ADVs directly to the server. The server processes the input and returns predicted performance metrics, which are then visualized within Grasshopper. This setup enables real-time feedback during the optimization process.

The embodied carbon emission calculation for Task 2 are based on material emission factors and metadata retrieved from the Boverket material database [116]. A Python 3 script is embedded in Grasshopper to conduct the calculation by multiplying material quantities derived from layer thickness and element area by the respective emission factors, and then summing across all layers to produce the total embodied carbon for a given envelope configuration. This workflow ensures that each archetype evaluated in the optimization is associated with a traceable, standardized embodied carbon value.

Wallacei was employed as the optimization tool within the Grasshopper environment. Wallacei is a multi-objective evolutionary optimization engine based on the Non-dominated Sorting Genetic Algorithm II (NSGA-II), a widely used multi-objective evolutionary algorithm. The algorithm follows the principles of natural selection and evolutionary biology, where a population of solutions evolves over successive generations toward better performance in terms of the defined objectives. A key feature of NSGA-II is the preservation of a diverse set of Pareto-optimal solutions. The algorithm uses a crowding distance mechanism to maintain diversity within the population, preventing premature convergence toward a single solution. This ensures that the outcome is not one single "best" solution, but rather a Pareto front that illustrates the trade-offs between competing objectives. Moreover, compared with other optimization plugins, Wallacei provides comprehensive recording of the entire optimization process, including parameter values, objective scores, and population evolution across all generations. Wallacei extends this functionality by integrating visualization tools and data analytics, allowing users to explore and compare the evolutionary process, examine the distribution of solutions across the objective space, and select configurations that best align with project-specific priorities.

In the optimization setting, each iteration has 100 population, and the optimization process is configured to terminate if the best solution remains unchanged for five consecutive iterations. The rest of the settings for the optimization in this study are listed in Table 3.7. It should be noted that the specific optimization settings are not the primary focus of this study; rather, the aim is to validate the overall approach and demonstrate the potential of using an ML model in supporting early-stage energy design decisions, rather than identifying the absolute best configurations in this context.

Table 3.7. Optimization settings

Setting Parameter	Value
Elitism	0.5
Mult. Probability	0.2
Mutation Rate	0.75
Crossover Rate	0.8
Population Size	100
Max. Generation	30

3.4.4 Evaluation metrics.

Two key metrics are commonly used to evaluate optimization algorithms: computational time and solution quality [117]. In this study, the first metric is the total computing time required to complete the optimization. The second metric is task-dependent: For Task 1, solution quality is measured by the best objective value obtained. For Task 2, solution quality is assessed using the Hypervolume as an indicator. The Hypervolume measures the size of the objective space dominated by the Pareto optimal solutions. It inherently reflects both the convergence quality of the solutions and their diversity, since a more widely spread Pareto front results in a larger hypervolume [118]. The higher the hypervolume value, the better the quality of the optimized Pareto optimal solutions.

3.5 The Use of AI.

The four-year period during which this thesis was developed coincided with a rapid acceleration in AI technologies, particularly in large language models. These developments created new opportunities to enhance research workflows while maintaining methodological rigor. In this thesis, AI tools such as ChatGPT, GitHub Copilot, Google Gemini, and DeepSeek were used in a limited and transparent manner to support specific aspects of the research process.

The use of AI is primarily focused on two areas. First, AI-assisted language tools were employed to refine the clarity, coherence, and academic tone of English writing. This includes improving sentence structure, strengthening argumentation, and ensuring stylistic consistency across chapters. All conceptual content, analytical reasoning, and scientific conclusions remain the author's own work. Second, AI-based coding assistants were used to support programming tasks, particularly for Python scripting, debugging, and generating alternative implementations for comparison. These tools were especially useful for improving code efficiency, identifying syntax errors, and exploring different approaches within the parametric modeling and machine learning workflow. However, all algorithms, data-processing pipelines, and model-development decisions were designed, tested, and validated by the author.

Overall, the integration of AI tools reflects contemporary research practice and serves as a means to enhance productivity, accuracy, and clarity. Their use was carefully managed to ensure that the intellectual contributions, technical decisions, and scientific insights presented in this thesis are the result of independent scholarly work.

Chapter 4

Findings

4.1 Study A: Influential ADVs for Early-stage Building Energy Optimization

Study A reveals the most influential ADVs for building energy optimization from both a literature and a stakeholder point of view. The selected ADVs, their definitions, and the influence of various ADVs in building energy demand from the literature review using the metrics mentioned in Chapter 3 as the influential factor can be seen in Table 4.1. It is presented that the window-to-wall ratio (WWR) on the four different facades has the biggest impact on building energy demand, reaching an influential factor of 0.81. This means that more than 80% of the academic papers looking into building energy optimization include WWR as input ADV. As mentioned in Chapter 3, an ADV is considered influential in literature when the rating is over 0.5, which indicates more than half of the previous studies include it. Under this criterion, building plan and building orientation are also influential ADVs for building energy, as they have 0.61 and 0.58 influential factors respectively.

Table 4.1. Selected ADVs, their definitions, and their influence on building energy demand from the literature

ADV	Definition	Influential factor on building energy demand
Window-to-wall ratio on north/south/west/east (WWR N, WWR_S, WWR_E, WWR_W)	Fraction of the exterior wall above grade that is covered by fenestration on the north /south/west/east façade, respectively	0.81/0.81/0.81/0.81
Building plan	Vertical projection onto a horizontal plane cutting through the building, showing the size and arrangement of spaces	0.61
Wall material	Material used for external walls	0.58
Building orientation	Relationship of a building and the positioning of its windows, rooflines, and other features to the building site	0.35
Roof material	Material used for the roof	0.32
Shading device	An integrated component of a window or I protecting the interior space from direct sun, overheating, and glare	0.32
Storey height	Height of each floor	0.26
Storey number	Number of floors	0.16

Figure 4.1 provides the average rating that the surveyed architects and sustainability consultants provided for the influence each ADV has on building energy demand. The ADVs with an average rating higher than 3 are considered influential as the scale was from 1 to 5. The color of the legend implies the influence a certain ADV has on building energy demand. Red indicates high influence while blue indicates low influence. While the two stakeholder groups gave similar ratings across the ADVs, sustainability consultants tended to give higher ratings than the architects. Both architects and consultants consider WWR_N, WWR_S, shading device, and wall material influential ADVs, while consultants also consider WWR_E, WWR_W, storey number, roof type, and roof material influential.

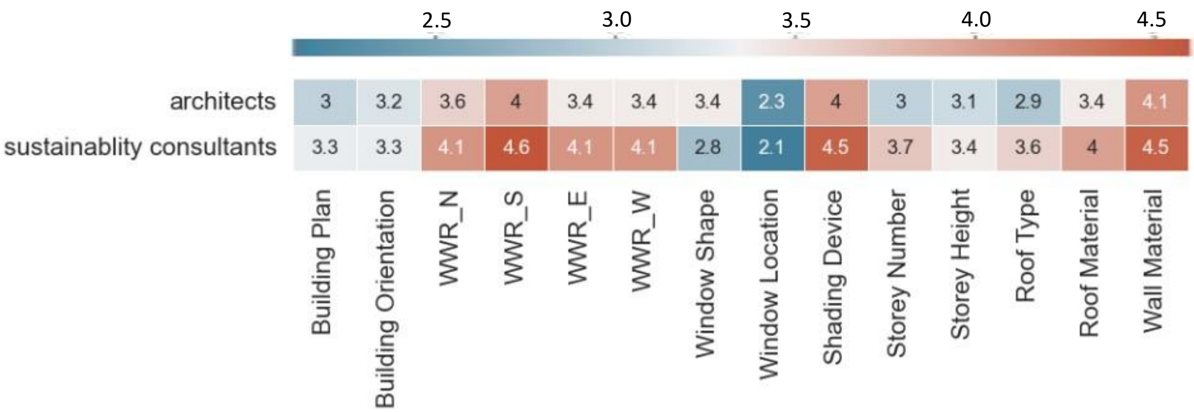


Figure 4.1. Rating of ADV influence on building energy demand from stakeholders.

Figure 4.2 compares the influence of ADVs from the literature and stakeholder perspective. The ratings from stakeholders are the mean value of all 24 participants, including 12 architects and 12 sustainability consultants. It can be seen that WWR, building plan, and wall material are considered influential by both literature and stakeholders, while the stakeholders' opinion on the influence of WWR_S is stronger than in the literature yet the influence of building plan is weaker than In the literature. Building orientation, shading device, storey number, storey height, roof type, and roof material are influential according to stakeholders but not in the literature.

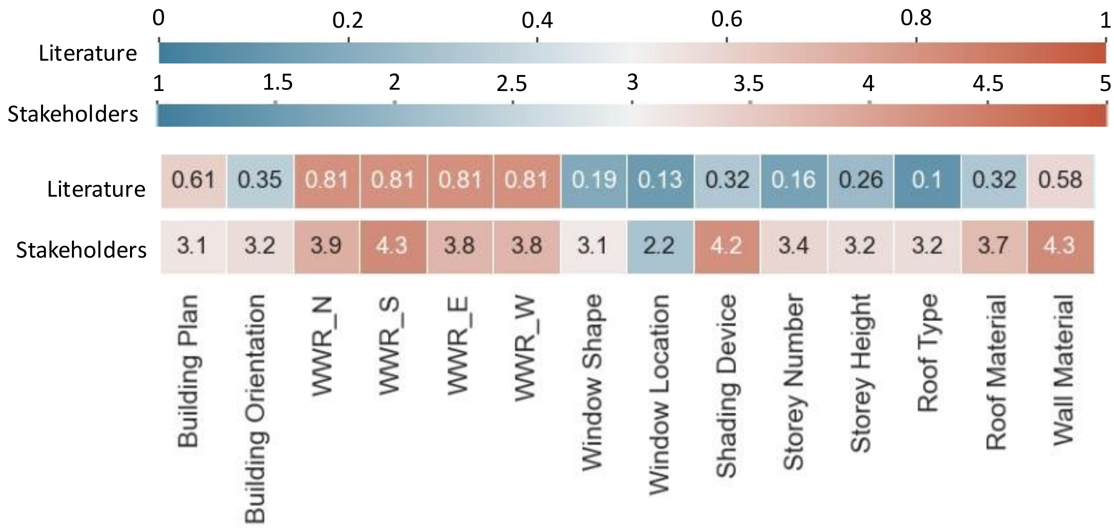


Figure 4.2. Comparison of ADV influence on building energy between the literature and stakeholders.

In summary, the ADVs that should be the input for the ML building energy prediction model are building plan, building orientation, WWR on four facades, shading device, storey number, storey height, roof type, roof material, and wall material. However, it is worth noting that although roof type is considered influential, it is not included in the further ML model development. This is because most mid-rise residential buildings in Sweden have cold roofs, which means the insulation is installed at the ceiling level, not on the roof itself. In this case, roof type will not influence the building's energy performance and therefore are excluded.

4.2 Study B: Synthetic Dataset Recommendations for ML Energy Prediction Models in Early-stage Optimization.

4.2.1 ML models' accuracy with various training datasets

The performance of the ML models in this study includes accuracy and computational efficiency. The size of the training dataset in this study refers to how many data points the dataset contains, and the diversity of the training dataset refers to how many building types the dataset contains. Figure 4.3 visualizes the training datasets in terms of size and diversity. Since the testing dataset contains nine building types, the training dataset with the lowest diversity covers around 22% of the testing dataset while the training dataset with the highest diversity covers around 89% of the testing dataset.

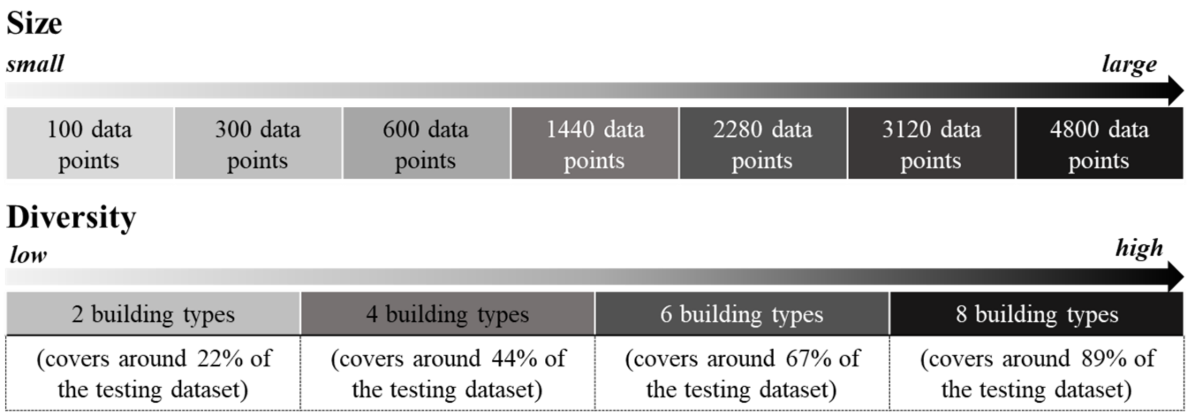


Figure 4.3. Visualization of the characteristics of the training datasets.

4.2.1.1 ML models' accuracy when increasing training dataset size

The five selected algorithms are applied to all 32 training dataset categories to investigate how the ML model's accuracy performance changes when increasing the size of the training dataset. In each training dataset category, five training datasets were developed with randomly selected building types. The value of RMSE and MAE presented in Figure 4.4 is the average value of all five sets of experiments.

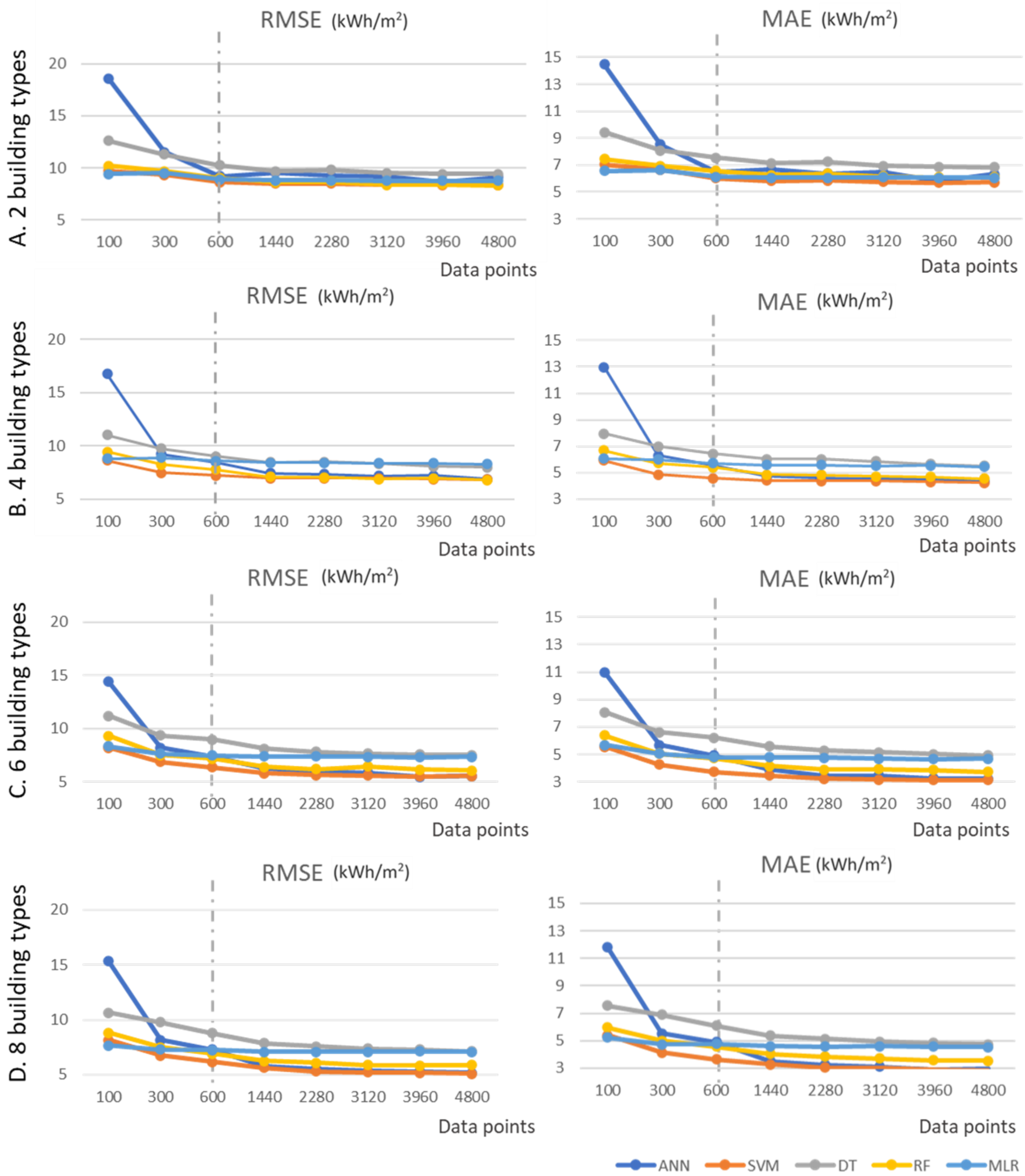


Figure 4.4. RMSE and MAE for five ML models when increasing dataset size for dataset with different diversity.

When the training dataset size is smaller than 300, increasing the dataset size can improve the ML model's accuracy significantly, especially when the algorithm is ANN. When increasing the dataset size from 300 to 1440, there is still an evident improvement in model performance. However, the accuracy of all ML models does not improve by more than 5% by increasing the dataset size after the dataset size reaches 1440 in general. Therefore, when generating synthetic datasets for an ML energy prediction model in the early stage, it might not be necessary to have more than 1440 data points in the training dataset considering that the generation of data points can be very time-consuming.

It Is also worth noting that the Impact on ML models' accuracy when Increasing data points also depends on the diversity of training datasets under certain circumstances. Before the dataset size reaches 600, increasing training data points can make more improvement for the ML model that is trained on lower-diversity datasets compared with the models trained on higher-diversity datasets. However, after the dataset size reaches 600, the extent of improvement for models trained on datasets with different levels of diversity does not show much difference when increasing data points.

Among the five algorithms, ANN is the most sensitive in terms of the training dataset size. When using ANN as the ML algorithm, it is more productive to improve the model's accuracy by increasing the number of data points in the training dataset, especially from 100 data points to 600 data points. Compared with other algorithms, MLR is the least sensitive algorithm in the size of the training dataset. Increasing the training dataset size does not make much impact on MLR's accuracy.

4.2.1.2 ML models' accuracy when increasing training dataset size

Dataset categories A3-D3 (600 data points with different diversity), A5-D5 (2280 data points with different diversity), and A8-D8 (4800 data points with different diversity) were taken as examples to investigate the impact of increasing dataset diversity on the ML models' accuracy performance under small, medium, and large dataset sizes.

Figure 4.5 shows that in general, increasing the training dataset's diversity can improve ML model accuracy significantly, especially when the dataset's diversity goes from covering only around 22% of the data in the testing dataset (2-type-building training set) to covering around 67% of the data in the testing dataset (6-type-building training set). However, the accuracy of all ML models does not improve by more than 10% by increasing the dataset size after the dataset contains more than six building types, covering around 76% of the building types in the testing dataset. It is also presented in Figure 4.5 that enhancing data diversity can achieve more improvement when the training dataset size is smaller. In general, the results achieved give the impression that increasing data diversity is more beneficial for reaching higher accuracy than increasing dataset size after the training dataset reaches 600 data points in this case.

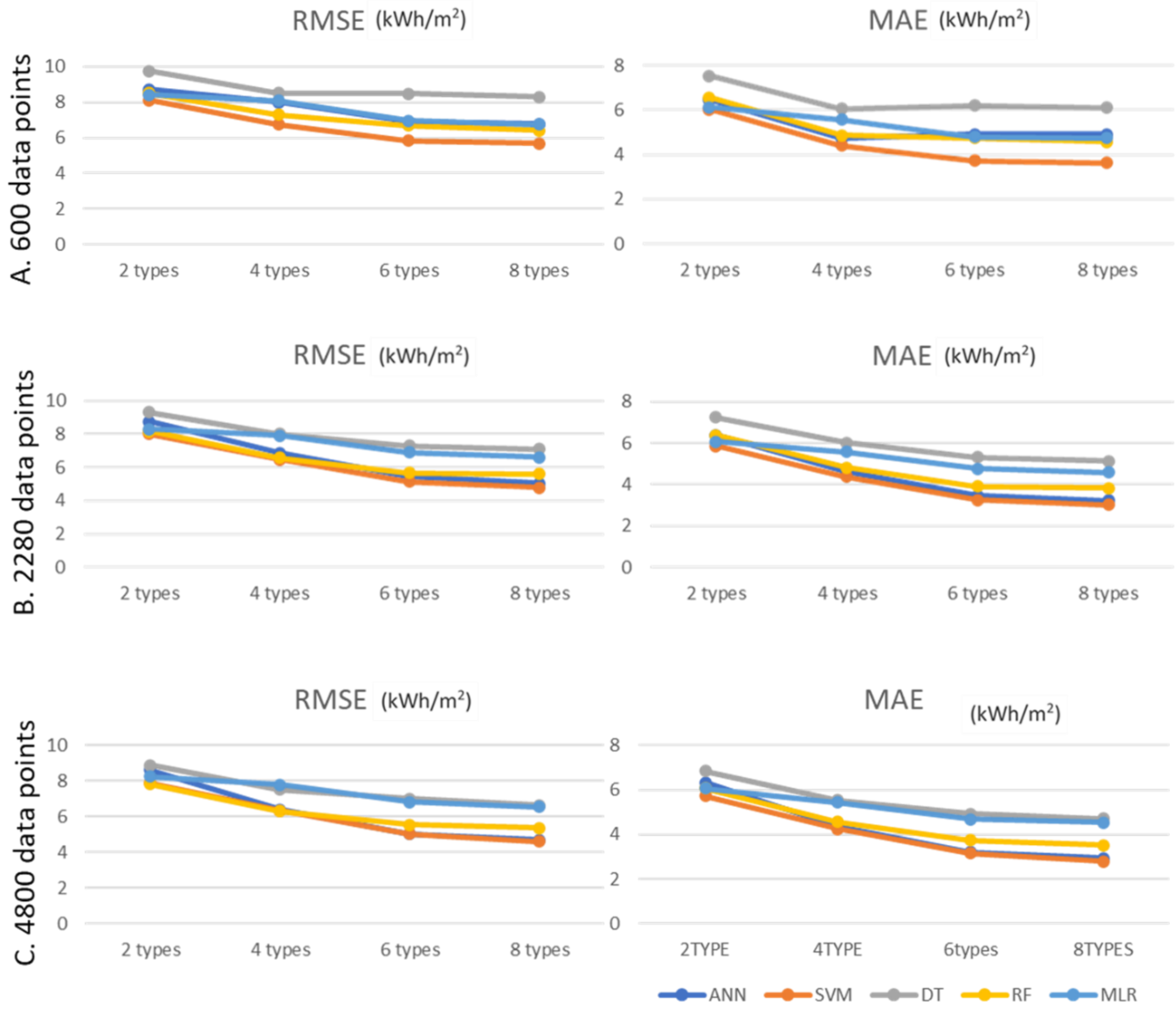


Figure 4.5. RMSE and MAE for five ML models when increasing dataset diversity under various sizes.

Among the five ML algorithms, SVM and RF are more sensitive to data diversity, which means that they improve more when increasing data diversity compared with other algorithms. DT is the least sensitive algorithm for training data diversity; its accuracy does not improve much when increasing dataset diversity. In the most extreme case, when increasing the training data diversity from 6 building types to 8 building types for 600-point datasets, the DT model's accuracy even worsens. This is further discussed in Chapter 5.

4.2.2 ML models' computational efficiency

Figure 4.6 presents the training time for five ML models when increasing dataset diversity for various sizes. It can be seen that MLR requires less than 0.005 seconds for the training process, which is the least amount of required time among the five algorithms. DT also has good performance in computational efficiency, the training time for DT does not exceed 40 seconds even for the biggest and the most diverse training datasets. The longest required training time for ANN and SVM does not exceed 110 seconds and 600 seconds respectively. RF has the worst performance in computational efficiency, the required training time can go up to more than 5000 seconds when the dataset is large. Even when the training dataset is small and has low diversity, the training time for RF is always the longest among all five selected algorithms.

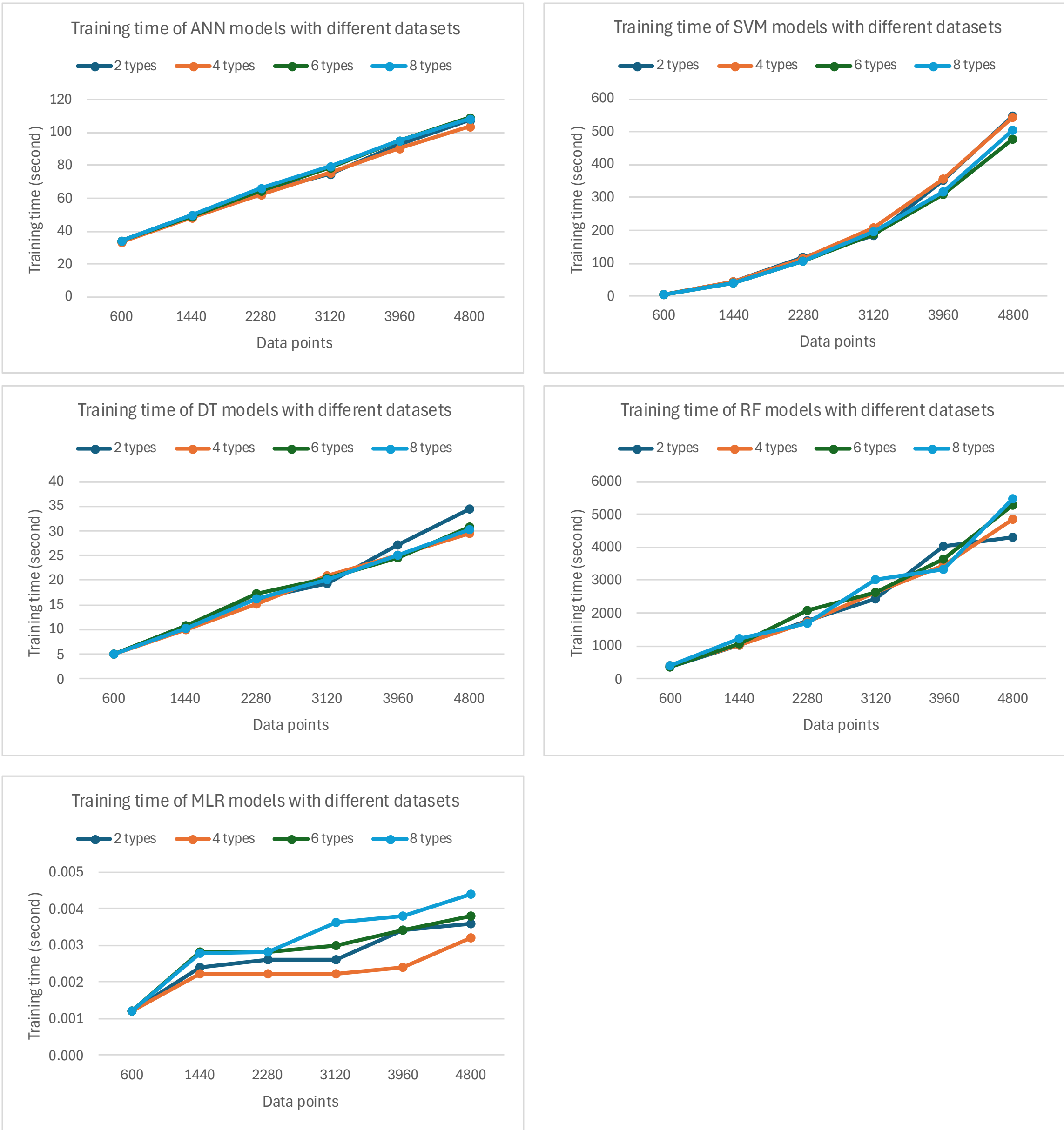


Figure 4.6. Training time for five ML models when increasing dataset size for various diversity.

It is also presented that the model computational efficiency is only influenced by the training dataset size, not diversity. Except that MLR's computational efficiency performance is always consistent regardless of the characteristics of the training dataset, the required training time for the rest ML models increases when the training dataset's size increases. However, the level of sensitivity for different ML models is different. The correlation between the growth rate of required training time and the growth rate of dataset size for SVM is exponential, while it is linear for the other algorithms. Specifically, Table 4.2 presents how much training time increases for all ML algorithms except for MLR for every 1000 more data points.

Table 4.2. The increase in training time for different ML algorithms

ML algorithm	Increased training time for every 1000 more data points (seconds)
ANN	Around 17
DT	Around 7
RF	Around 1100
SVM	From around 45 to around 230 depending on the size of the training datasets

Figure 4.7 shows that for all developed ML models, the predicting time for the entire testing dataset does not exceed 0.3 seconds, which means that the predicting time for one building configuration takes less than 0.0003 seconds. This is a significant improvement compared to using a building energy simulation engine such as Energy Plus, which takes around three minutes for one simulation on the same computer.

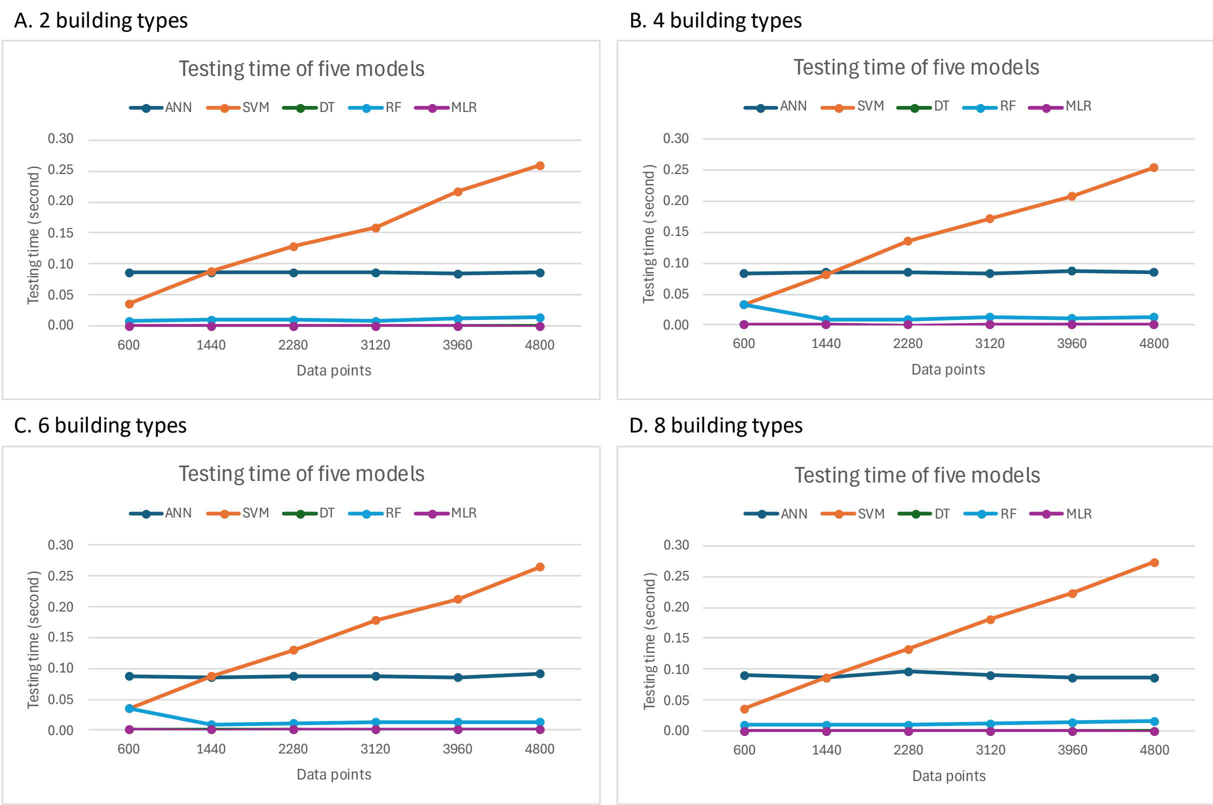


Figure 4.7. Testing time of five models.

4.2.3 Recommendations for developing ML energy prediction model for early-stage optimization

4.2.3.1 Recommendations for selecting algorithms for various training datasets

Figure 4.4 and Figure 4.5 show that different ML algorithms perform differently for different training datasets, and the best-performing algorithm varies. Figure 4.8 presents the best-performing ML algorithms for different training datasets in terms of accuracy. The X-represents the diversity of the training dataset while the Y-axis represents the size of the training dataset. The best-performing algorithms for each dataset category are listed. In general, SVM is the ML algorithm that performs best for almost all training datasets. MLR performs best compared to other algorithms when the training dataset is very small or has low diversity. RF performs best for medium-size training datasets with low-to-medium diversity. ANN performs best when the training dataset has high diversity and large size. DT has the worst performance in terms of accuracy in all training dataset categories and therefore is not a suitable algorithm for predicting building energy demand.

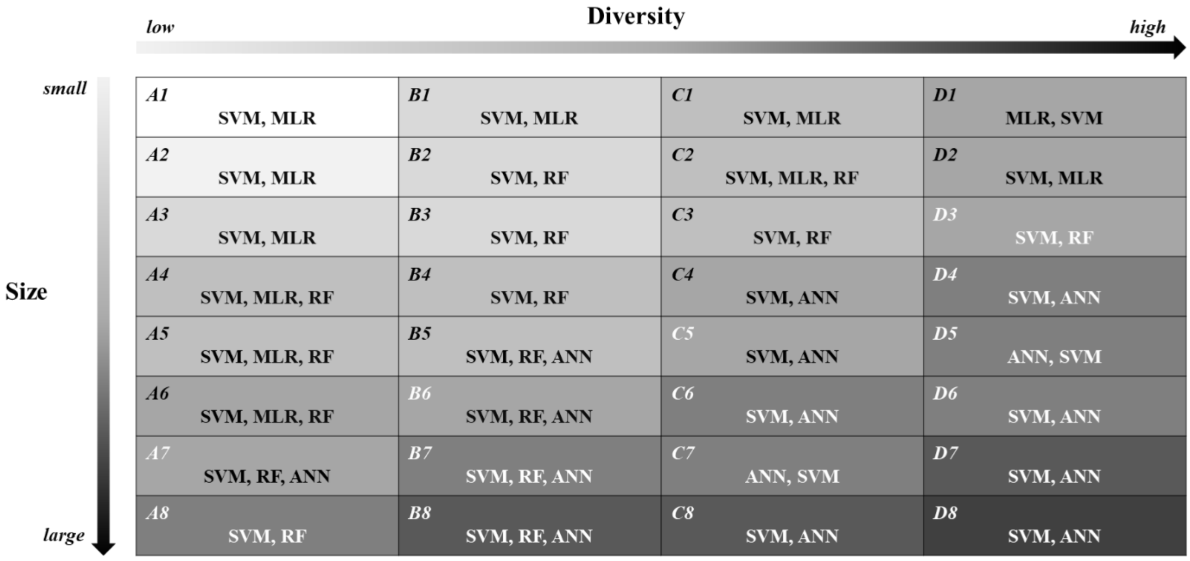


Figure 4.8. Best-performing ML algorithms for various training datasets in accuracy.

When selecting the most appropriate ML algorithm for developing energy prediction models, it is also important to take computational efficiency into consideration. Although RF performs well for certain datasets, the hyperparameter tuning for RF can require much more computational time compared to other algorithms. Taking the training dataset category A8 as an example, the time for training the RF model is almost eight times more than the SVM model. On the other hand, for smaller training datasets, MLR is also a good option. MLR models have the second-best accuracy performance for small datasets and require nearly no training time as they do not have the hyperparameter tuning process.

In summary, it is recommended to use SVM as the algorithm for building energy prediction models as it is the best-performing one for most training dataset categories, and the required training time is also in the middle position among all selected algorithms. When the training dataset has low diversity and/or small size, it is recommended to consider MLR. MLR has a relatively high accuracy performance and costs nearly no time in the training process. Moreover, MLR is one of the basic ML algorithms and is very easy to pick up. When the training dataset has high diversity and/or large size, ANN is also recommended.

4.2.3.2 Recommendations for generating synthetic datasets for different algorithms

As mentioned in Chapter 1, generating a synthetic dataset can be very time-consuming. Therefore, when developing ML building energy prediction models, it is essential to understand how many data points are enough and how diverse are enough. Figure 4.9 presents the accuracy performance of the five models using different training datasets.

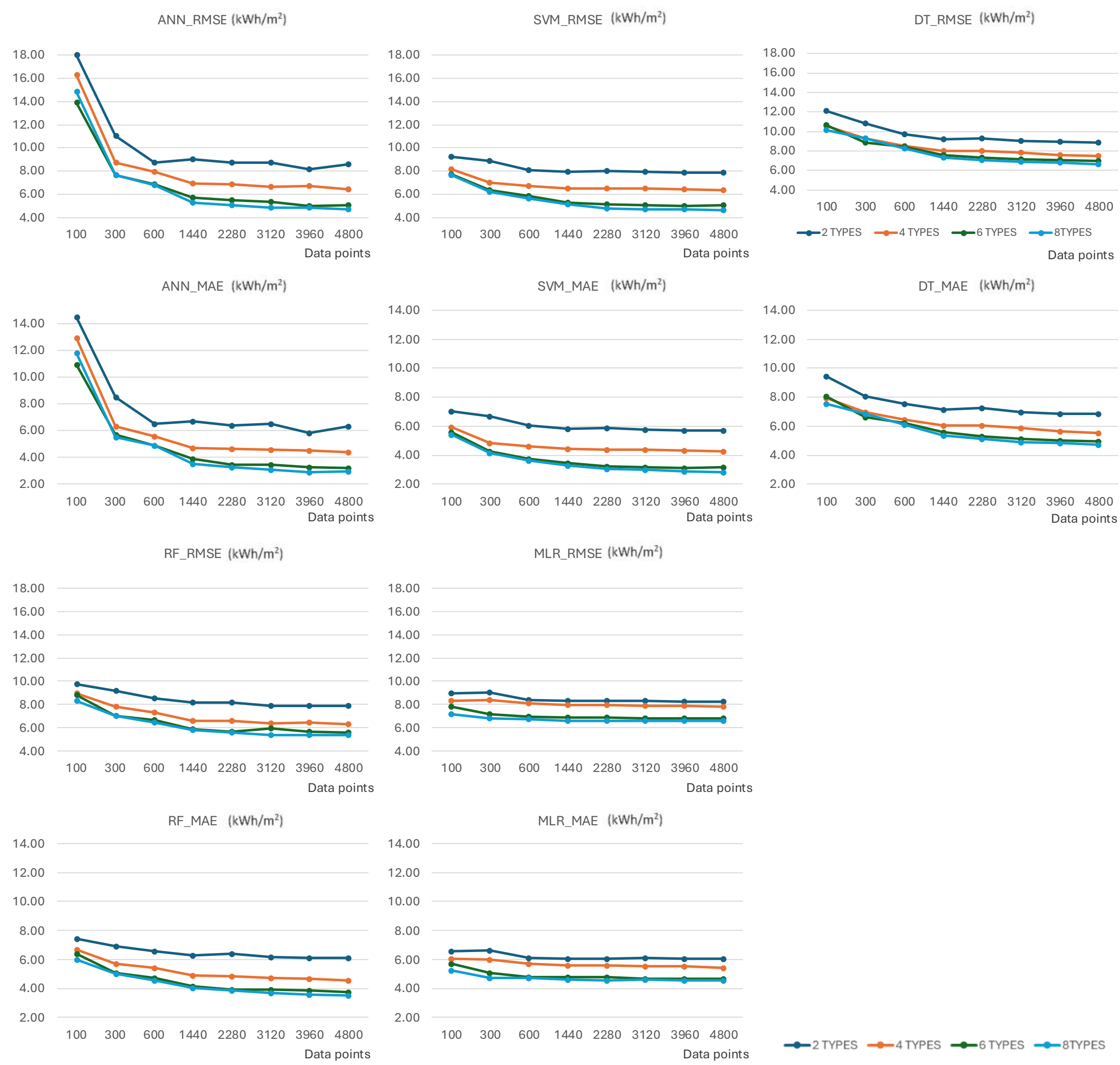


Figure 4.9. RMSE and MAE for five models with different training datasets.

When developing an ML prediction model for early-stage optimization, it is considered that an RMSE that is lower than 7% (around 5.5 in this study) and an MAE that is lower than 5% (around 4 in this study) are reasonable performance for early-stage predictions based on previous studies [70] [73] [78] [80]. Under this circumstance, SVM, ANN, and RF are the proper algorithms to select as they have the best accuracy performance among all the experiments. Table 4.3 presents the least requirement of a training dataset for each ML algorithm to have reasonable performance.

Table 4.3. The least requirement of a training dataset for each ML algorithm to have reasonable performance

ML algorithm	Size and diversity of the corresponding training dataset	
SVM	Size: 1440 data points Diversity: 6 building types (cover around 67% of the diversity in the testing dataset)	
ANN	Size: 2280 data points Diversity: 6 building types (covers around 67% of the diversity in the testing dataset)	Size: 1440 data points Diversity: 8 building types (covers around 89% of the diversity in the testing dataset)
RF	Size: 3120 data points Diversity: 8 building types (covers around 89% of the diversity in the testing dataset)	

In general, a synthetic dataset with more than 1440 data points and a diversity that covers around 67% of the testing dataset with SVM is the least requirement for a high-performance ML model. A training dataset containing more than 3120 data points and a diversity that covers around 89% of the testing dataset with either of the three mentioned algorithms can guarantee to develop a high-performance model.

However, when willing to slightly compromise the accuracy of the ML model for higher efficiency and less time in generating the training dataset, multiple ML algorithms can be considered. For instance, if the set goal is to achieve an RMSE that is lower than 10% (around 7.8 in this study) and an MAE that is lower than 7% (around 5.5 in this study), small training datasets can also be enough for developing the ML models. A training dataset with 300 data points and a diversity that covers around 67% of the diversity in the testing dataset can be enough for developing an MLR, RF, or SVM model with the above accuracy performance; a training dataset with 600 data points and a diversity that covers around 67% of the diversity in the testing dataset can be enough for developing an ANN model with the above accuracy performance. Although it is normally very time-consuming to conduct hyperparameter tunning for the RF model, the tuning time is only around 200 seconds when the dataset contains only 300 data points, making the process more affordable. In general, the dataset size can be smaller when comprising the accuracy for higher time efficiency, but the dataset diversity always needs to be up to at least covering around 67% of the diversity in the testing dataset.

4.2.4 Synthesis of Study B

Study B reveals the performance of five selected algorithms regarding training datasets with different sizes and diversity and makes recommendations. In terms of selecting the ML algorithm, SVM performs well in terms of accuracy for all training datasets and has an acceptable performance in computational efficiency. MLR is recommended for training datasets with smaller sizes and lower diversity while ANN is recommended for training datasets with larger sizes and higher diversity. In terms of generating synthetic training datasets, to achieve reasonable accuracy, the dataset needs to have more than 1440 data points and a diversity that covers around 67% of the diversity in the testing dataset. Overall, this section provides recommendations for developing an ML energy prediction model for building early-stage optimization.

4.3 Study C: Applying ANN Early-stage Building Energy Prediction Models to Alternative Climates Using TL

4.3.1 Improvement of applying various TL models for different scenarios.

Figure 4.10 presents the RMSE of various TL models and retrained ANN models in predicting heating energy demand in the selected target cities. The horizontal axis represents the size of the target city training dataset, while the vertical axis indicates the model's RMSE. As several selected metrics exhibit similar trends, only the RMSE results are presented in this section, while MAE, MAPE, and R^2 can be found in Appendix D.

As shown in Figure 4.10, the results indicate that multiple TL models outperform models trained from scratch in predicting heating energy demand across the four target cities. However, the number of effective TL models and the degree of improvement vary. In both Stockholm and Chicago, all TL models outperform the baseline retrained model. In Seattle, all TL models except for A-1 show better performance. For Madrid, only the TL models developed by method 2 yield improved results.

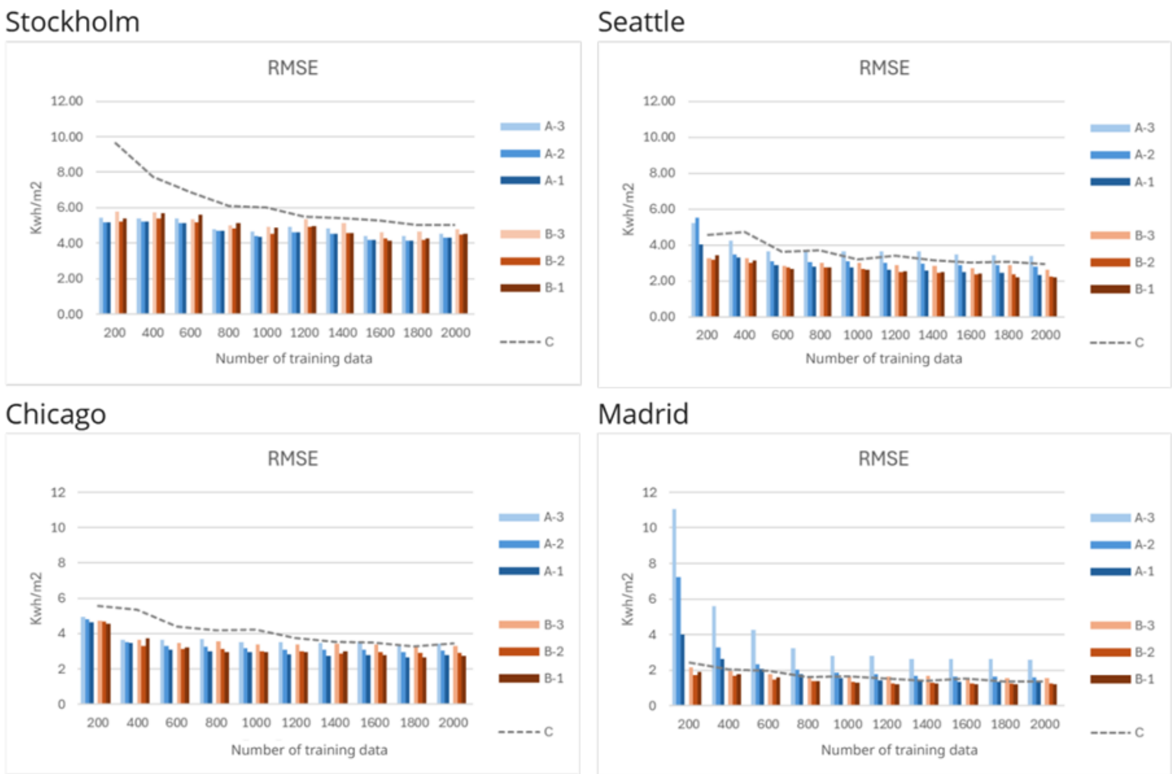


Figure 4.10. RMSE of TL models and retrained ANN models under different training dataset sizes in predicting heating energy demand for four cities.

Figure 4.11 presents the RMSE of various TL models and retrained ANN models in predicting total energy demand in the selected target cities. The results suggest that the performance of TL models in predicting total energy demand across different target cities generally follows a similar trend to that observed in heating energy demand prediction in Figure 4.10. Since Stockholm does not require cooling energy, it is excluded from Figure 4.11. The patterns observed in Seattle and Chicago in Figure 10 are consistent with those for heating energy in

Figure 4.10, where the majority of TL models improve the prediction accuracy of total energy demand. In contrast, TL performance in Madrid and Miami is less effective. For Madrid, only models B-2 and B-1 show improvements compared with retraining, while for Miami, only model B-1 is more effective than retraining, and even then, the improvement is limited.

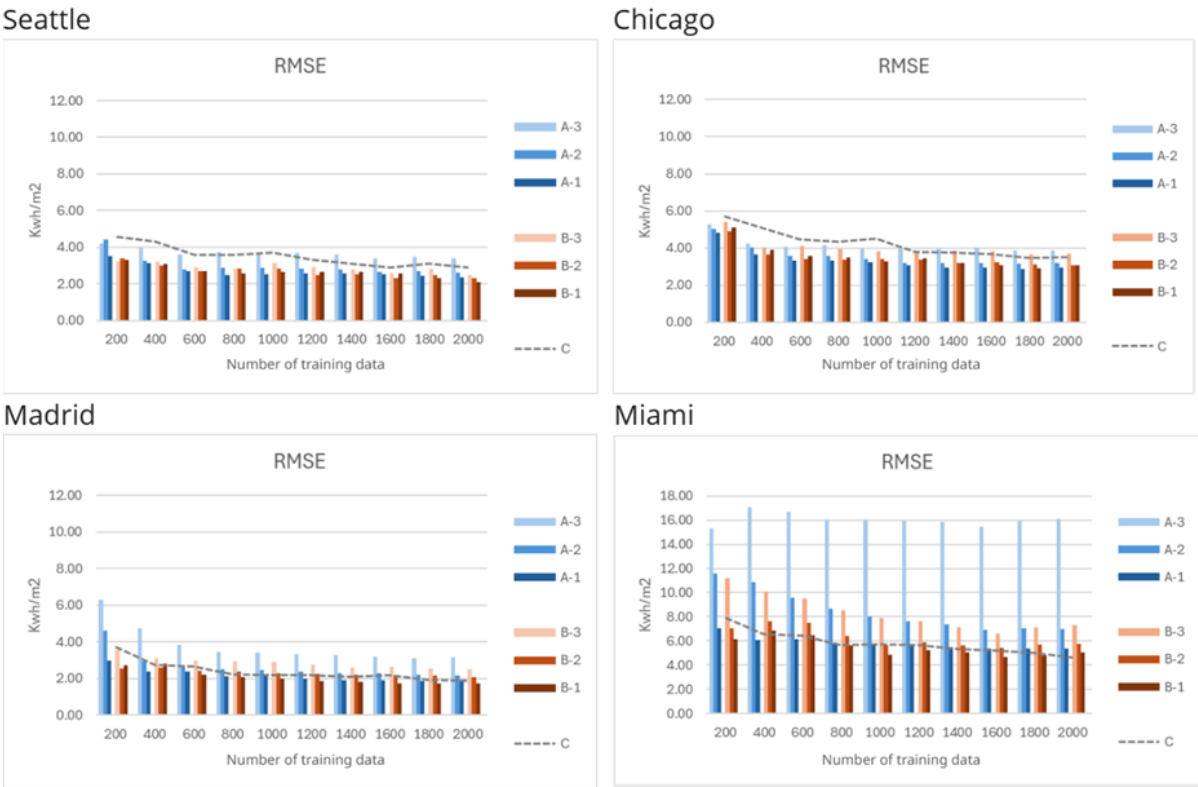


Figure 4.11. Accuracy of TL models and retrained ANN models under different training dataset sizes in predicting total energy demand for four cities.

Overall, it can be observed that when applying TL models to predict building energy demand in target cities that differ from the source domain city, the performance of Methods A and B depends on the similarity between the two cities. When the target and source cities share similar climatic conditions, the differences between Methods A and B are not prominent. In such cases, for heating demand prediction, all TL configurations tend to yield improvements regardless of the number of frozen layers, and the impact of the number of frozen layers is relatively small. For total energy demand, freezing one or two layers performs better, while freezing three layers may not lead to further improvement. Conversely, when the target and source cities differ significantly in climate and energy type composition, Method B substantially outperforms Method A and tends to require fewer frozen layers. In general, Method B shows superior performance across settings, with better results observed when fewer layers are frozen.

4.3.2 Improvement of applying TL in target cities for various dataset sizes.

The best-performing TL model for each scenario are selected and evaluated its performance gain compared to the retrained model (Method C). Figure 4.12 presents the improvement achieved by the TL models over retraining from scratch. Specifically, Figure 4.12 (a) shows the results for heating energy prediction, while Figure 4.12 (b) illustrates those for total energy prediction. Two evaluation metrics are used—R^2 and MAPE—since R^2 is bounded by 1 and MAPE is a percentage, making them more interpretable and comparable across different cases. The x-axis of Figure 4.12 represents the size of the training dataset for the target city. At the same time, the y-axis indicates the improvement in R^2 and the reduction in MAPE achieved by the best-performing TL model relative to the retrained model.

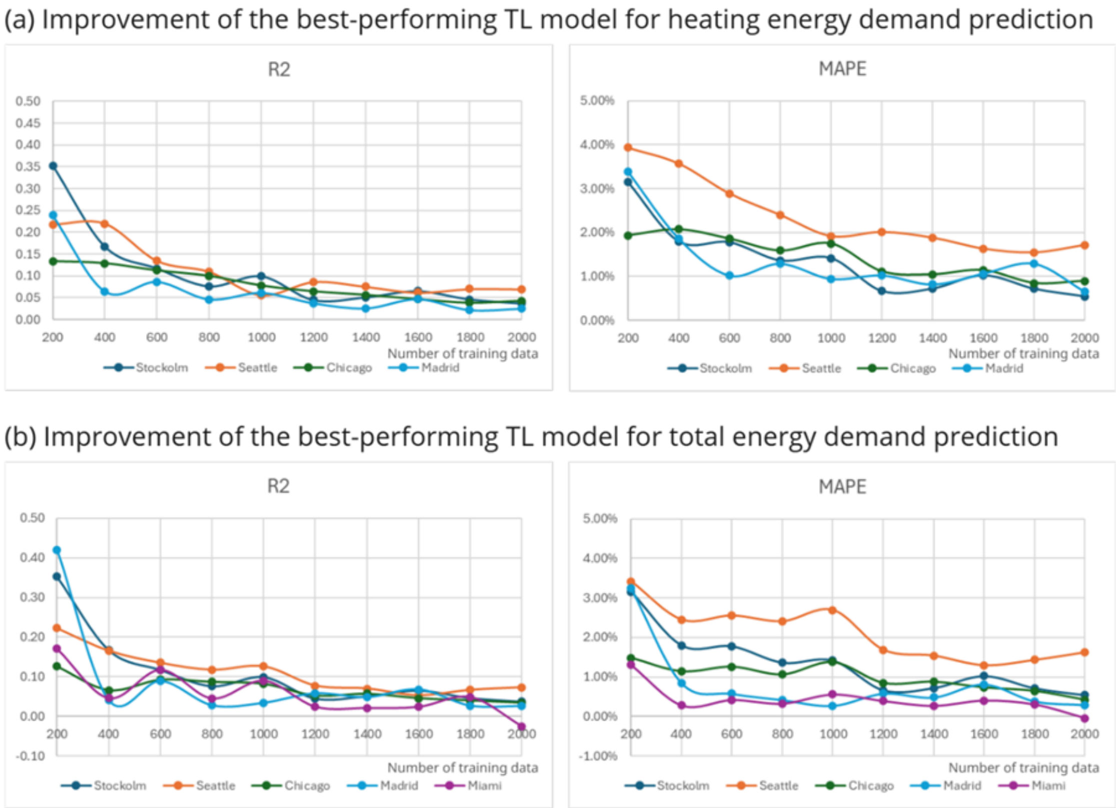


Figure 4.12. Improvement of the best-performing TL for heating and total energy demand prediction.

As shown in Figure 4.12, except for the case where the training dataset size for Miami is 2000, under which TL fails to improve prediction accuracy, TL consistently enhances performance across all other settings. The trends in Figure 4.12 are clear: in general, the smaller the training dataset size for the target city, the greater the performance improvement achieved by TL. For heating energy demand prediction, the greatest improvement is observed for Seattle, which is the city most similar to the source domain city (Gothenburg) in terms of climate (they are both in climate zone C), while the smallest improvement occurs for Madrid, which is the most dissimilar. A similar trend is observed in total energy demand prediction: TL provides the most significant gains for Seattle and the least for Miami. In general, these results highlight that TL is particularly beneficial when training data in the target city is limited and when the source and target cities share similar energy behavior characteristics.

4.3.3 Difference of TL performance in predicting heating energy demand and total energy demand for various target cities.

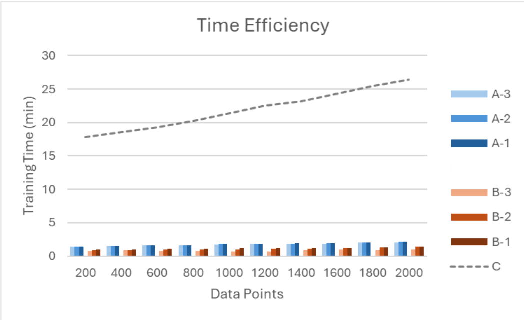
Figure 4.13 illustrates the performance improvement of the best-performing TL model over the retrained model in predicting both heating and total energy demand for the same target city, across different training dataset sizes. Only three cities—Seattle, Chicago, and Madrid—are included in this figure, as they require both heating and cooling energy. Stockholm and Miami were excluded because the former only has heating demand and the latter only has cooling demand.

Figure 4.13. Accuracy of TL models and retrained ANN models for heating and total energy.

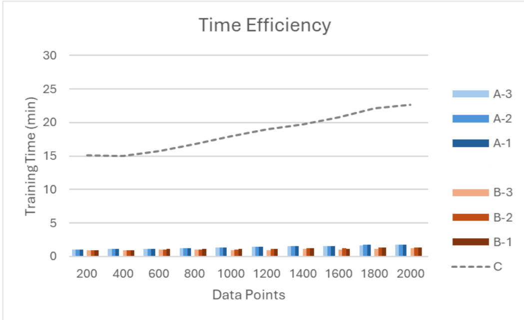
As shown in Figure 4.13, in most cases, except for the prediction of building energy in Seattle with a training dataset size of 1000, the TL model yields greater performance improvement for heating energy demand prediction, which aligns with the energy type used in the base model. This trend becomes more evident and consistent when the target city differs more significantly from the source domain city in terms of climate conditions and energy type composition.

4.3.4 Improvement of applying TL in time efficiency.

One of the key advantages of using TL is the significant reduction in model development time. Figure 4.14 presents the average training time required to predict energy demand with and without TL. As shown in Figure 4.14, the training time needed to predict heating and total energy demand in different target cities is considerably lower when using various TL models compared to retraining a full ANN model from scratch. We selected the average value to indicate the training time, as the trends for different cities are consistent.



(a) Average training time of various models predicting heating energy for Stockholm, Seattle, Chicago, and Madrid



(b) Average training time of various models predicting total energy for Seattle, Chicago, Madrid, and Miami

Figure 4.14. Average training time of TL models and retrained ANN models under different training dataset sizes for energy prediction of target cities.

It can be seen that, regardless of the TL approach used, the training time is consistently much shorter than that of full retraining. Moreover, within the TL models, freezing three layers of the original ANN requires the least training time, followed by freezing two layers, and then one layer. This trend is more pronounced in Method B. However, although freezing more layers leads to shorter training times, the differences are not substantial. Even the slowest TL model completes training in under three minutes, demonstrating the efficiency of the TL approach.

4.3.5 Reduction in required training data by applying TL.

It is well known that ML models typically require large volumes of training data to achieve high performance. However, generating such training datasets, particularly in the context of building energy simulation, can be time-consuming and computationally expensive. Figure 4.14 does not account for the time required to generate training data. Figure 4.15 shows the minimum data requirement at varying accuracy levels with and without TL.

(A). Minimum Data Requirement at Varying Accuracy Levels for Predicting Combined Heating and Cooling Energy

R2		0.7	0.75	0.8	0.85	0.9
Seattle	TL	200	200	200	400	1600
	Retrain	400	800	1200	1600	X
Chicago	TL	200	200	400	600	1200
	Reatrain	400	800	1200	X	X
Madrid	TL	200	200	600	1000	X
	Retrain	400	800	800	1800	X
Miami	TL	600	1400	1800	X	X
	Retrain	1200	1600	2000	X	X

MAPE		5.5%	5%	4.5%	4%	3.5%	3%
Seattle	TL	600	800	1800	2000	X	X
	Retrain	2000	X	X	X	X	X
Chicago	TL	200	200	400	600	800	1800
	Reatrain	400	800	1200	2000	X	X
Madrid	TL	400	800	1400	2000	X	X
	Retrain	800	1400	2000	X	X	X
Miami	TL	200	200	200	200	1000	1600
	Retrain	200	400	400	800	1400	2000

(B). Minimum Data Requirement at Varying Accuracy Levels for Predicting Heating Energy

R2		0.7	0.75	0.8	0.85	0.9
Stockholm	TL	200	200	200	800	1000
	Retrain	600	800	1000	1800	X
Seattle	TL	200	200	200	400	1200
	Reatrain	600	800	1000	2000	X
Chicago	TL	200	200	400	600	800
	Retrain	400	600	800	1200	X
Madrid	TL	200	200	200	200	800
	Retrain	400	400	800	800	1800

MAPE		5.5%	5%	4.5%	4%	3.5%	3%
Stockholm	TL	200	200	200	400	800	1800
	Retrain	600	800	1000	1400	X	X
Seattle	TL	600	1200	1800	X	X	X
	Reatrain	X	X	X	X	X	X
Chicago	TL	200	400	600	800	1000	X
	Retrain	800	1200	1200	1800	X	X
Madrid	TL	1800	X	X	X	X	X
	Retrain	X	X	X	X	X	X

Figure 4.15. Minimum Data Requirement at Varying Accuracy Levels with and without TL.

It can be seen in Figure 4.15 that TL can use substantially smaller training datasets while achieving comparable prediction accuracy to models trained from scratch with much larger datasets. This trend appears to be more pronounced when the target city shares greater similarity with the base city. For instance, in the case of heating energy prediction, TL enables the model to achieve an R^2 of 0.85 for Seattle while using 1,600 fewer training samples compared to retraining. Similarly, when predicting the total energy demand (heating and cooling combined) for Chicago, TL achieves a MAPE of 4% with 1,400 fewer samples. However, in the case of Miami, where the energy demand is entirely cooling-based and the climatic conditions differ substantially from the source city, the maximum data savings achieved by TL is limited to 600 samples, with the corresponding R^2 reaching only 0.70. In this study, generating a single data point takes approximately seven minutes on average. By applying TL, the amount of required data can be significantly reduced, resulting in time savings of up to 186 hours, which is more than a whole week. Therefore, in addition to reducing model training time, TL significantly improves overall time efficiency from a data generation perspective. This further highlights the practical advantage of TL in accelerating the model development cycle, especially in data-scarce or resource-constrained scenarios.

4.3.6 Synthesis of Study C.

Study C explores the generalizability of the developed ML building energy prediction model by applying it to make energy predictions in different climate contexts using TL. The results suggest that TL in general can be an effective strategy, but it performs best when the target city shares similar climatic conditions and energy use patterns with the base city, such as Seattle, Stockholm, and Chicago, and its effectiveness decreases when the climate difference is large, as seen in the case of Madrid and Miami. It is also suggested that TL models' performance is more evident when the target dataset contains limited data (below 600 data points), and the improvement becomes smaller as the amount of training data of the target city increases. The effectiveness of TL also differs depending on the energy type being predicted。TL is more effective when predicting heating energy than total (heating + cooling) energy demand within the same target city. Since the base model is trained specifically for heating energy in a cold climate, its internal representations align more closely with heating-dominant patterns.

One of the most significant advantages of generalizing an existing ML model using TL instead of developing a new one is its potential to drastically reduce the time and data needed to develop reasonably accurate prediction models. In some cases, TL required 1000–1600 fewer training samples to reach comparable performance to a fully retrained model. Given that generating each training sample in this study took about 7 minutes, TL could save up to 186 hours of computational time. Moreover, applying TL can also improve decent accuracy, like an R^2 of 0.85 and an MAPE of 5% with limited data as few as 600.

In summary, this study confirms the potential of generalizing ML models by using TL to support efficient, data-light, and scalable energy prediction across geographic regions. It highlights TL as a promising tool for accelerating the development of early-stage building energy optimization tools.

4.4 Study D: Evaluation and Comparison of the ML-based Early-stage Building Energy Optimization.

4.4.1 Results for Task 1.

Task 1 is a single objective optimization task that aims to find the building design configuration with the minimum building energy demand. Table 4.4 presents the optimization results based on the settings presented in Section 3.4. Optimization workflow A utilizes Energy Plus as the simulation engine, whereas optimization workflow B integrates the ML prediction model, rather than relying on simulation. The results show that the computational time of Workflow B is 38 hours less than Workflow A, which is a prominent improvement in time efficiency. Although the energy demand of the optimal building design configuration proposed by Workflow A is a bit lower than Workflow B, the difference between them is only 0.29 kWh/m^2.

Table 4.4. Optimization results comparison of Task 1.

		Optimization Workflow A	Optimization Workflow B
Number of Iterations		22	28
Optimization Time		38 hours 37 minutes 42 seconds	22 minutes 47 seconds
Simulated Energy Demand of the Optimal Building Design Configuration (kWh/m^2)		72.82	73.11
Parameters of the Optimal Building Design Configuration	Building width (Meter)	9.9	9.9
	Window length/ wall length	0.8	0.7
	Storey height (Meter)	3.0	3.0
	Shading type	1	0
	Shading length (Meter)	0.2	0.2
	Wall U value (W/m^2·K)	0.1	0.1
	Roof U value (W/m^2·K)	0.1	0.1
	Floor U value (W/m^2·K)	0.1	0.1
	Window U value (W/m^2·K)	1.0	1.0

Figure 4.16 presents the optimal building design configuration proposed by the two Workflows. It can be seen from both Figure 4.16 and the parameters of optimal building design configurations from Table 4.4 that the two configurations are highly similar. The most influential ADVs are the same, including the building width, storey height, and envelope U values. This proves that the ML-based building energy optimization workflow proposes the optimal solutions in the same direction as the simulation engine-based building energy optimization workflow, but with a much higher computational efficiency.

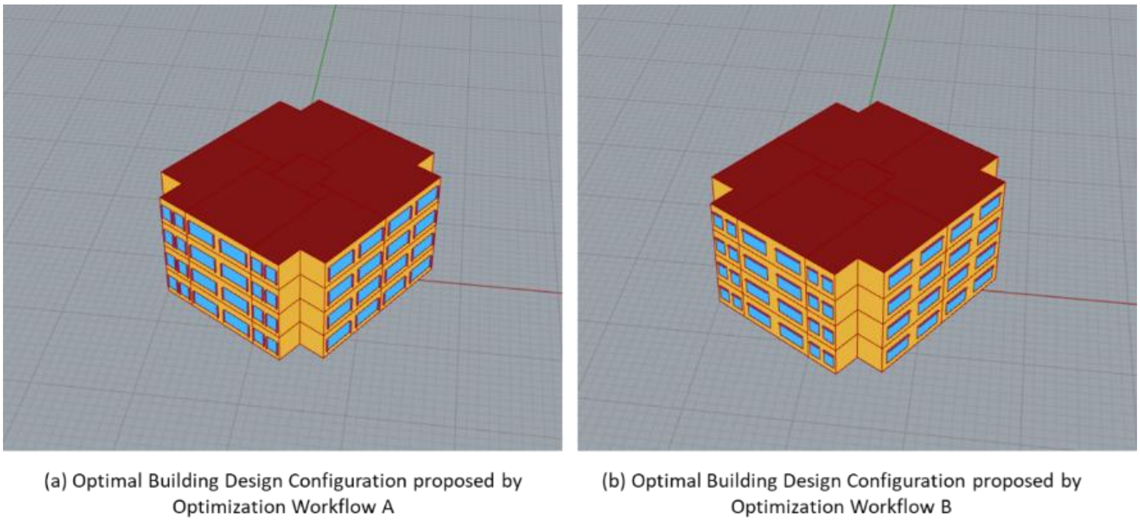


Figure 4.16. Optimal building design configurations proposed by the two optimization workflows.

4.4.2 Results for Task 2.

Task 2 is a multi-objective optimization task that aims to find the building design configuration with the minimum building energy demand and embodied carbon emissions. Figure 4.17 presents the Pareto front results of the two proposed optimization workflows. It is worth noting that in this study, the k-means clustering algorithm was applied to the obtained Pareto solutions. This procedure groups the solutions into 10 clusters, which facilitates a more structured analysis of the Pareto front and enables clearer comparisons across optimizations.

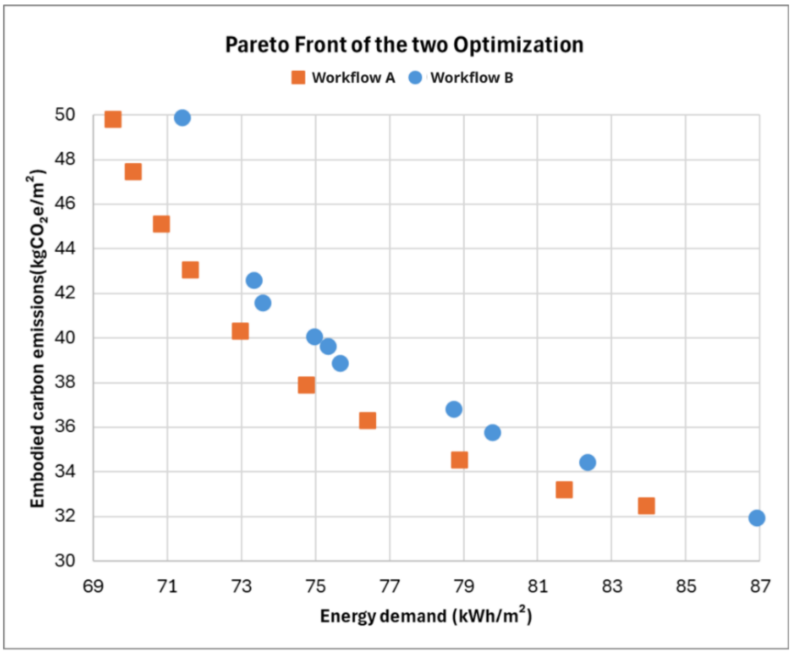


Figure 4.17. Pareto front of the solutions proposed by the two optimization workflows.

Figure 4.17 illustrates that Workflow A exhibits a broader coverage of the objective space, with its solutions extending closer to both axes. In contrast, Workflow B produces a more concentrated set of solutions within the mid-range of energy demand (72–82 kWh/m^2), yet with relatively higher embodied carbon values in the low-energy region. This indicates that Workflow A achieves more favorable trade-offs between minimizing embodied carbon and reducing building energy demand, whereas Workflow B demonstrates competitive performance in lowering energy demand but provides less benefit in carbon reduction. Given that both optimizations employed the same algorithm, the most plausible explanation for the superior Pareto coverage of Workflow A lies in the evaluative fidelity: Workflow A was assessed using a simulation engine, while Workflow B relied on a machine learning surrogate model, with the higher fidelity of the simulation engine likely contributing to the enhanced performance of Workflow A.

Table 4.5 presents the evaluation metrics for both optimization workflows. For the Hypervolume calculation, it is necessary to define a reference point that is worse than all obtained solutions. In this study, the reference point was determined by taking the maximum values of each objective among all solutions from both workflows and then adding a margin of 5 units to each objective. Based on this procedure, the reference point was set to have 91.92 kWh/m^2 energy demand and 54.90 kgCO_2e/m^2 embodied carbon emission. Using the same reference point for both workflows ensures that the Hypervolume values are directly comparable. As mentioned in Section 3.4.4, a higher value in hypervolume indicates higher quality solutions. Table 4.5 shows that although Workflow A has a higher quality solution, Workflow B demonstrates a remarkable improvement in computational efficiency, completing the optimization nearly sixty hours faster than Workflow A.

Table 4.5. Evaluation results comparison of Task 2.

Evaluation Metric	Workflow A	Workflow B
Computational Time	62 hours, 36 minutes, and 3 seconds	1 hour, 40 minutes, and 50 seconds
Hypervolume	416.98	367.40

Chapter 5

Discussion

5.1 Towards a Holistic Perspective in Identifying Input ADVs

Study A identified the most influential ADVs for building energy optimization through a literature review and a stakeholder point of view and used them as the input for the ML energy prediction model. Section 4.1 discussed how the opinions of stakeholders and literature are not always aligned. This finding is not specific to building energy, when applying the same method to investigate the influential ADVs for other sustainability objectives, such as daylight or embodied carbon, the discrepancies between the literature and the stakeholders are observed (see appended paper).

The discrepancies between the literature and the stakeholders can represent the discrepancies between academia and industry when it comes to early-stage optimization. It could be explained by multiple reasons. First, most academic studies use computer simulation as the main method when trying to evaluate a building's performance or investigating how to optimize a building design. Computer simulation is a reasonable and feasible alternative as it has an acceptable level of accuracy and is easy to implement, yet it is also very well-known that simulation results can differ from reality [119]. There are multiple reasons leading to this. For instance, the computational simulation always needs weather data as input, which can lack accuracy; most simulation engines use simplified modeling assumptions, which may not fully capture the intricacies of real-world conditions; occupants' behaviors may be unclear, and equipment performance in real-world conditions may differ from that in a simulation. Furthermore, the simulation results are normally very specific to a certain case and context. Many studies state that their results are valid only for the particular situation and are not generalizable [120][121]. Therefore, it is crucial to include stakeholders' opinions as their experience is usually based on working with real building projects, and their knowledge could compensate for the shortcomings of computer simulation results, especially when developing a tool that is highly based on ML.

However, this does not mean that the stakeholders always hold objective answers. Through the six follow-up interviews, it was found that the respondents' answers can sometimes be very subjective. First, the stakeholders are more willing to give an ADV a lower rate and consider it to be less influential when they feel reluctant to change it for a lower building energy demand. For instance, according to section 4.1, building plan is more influential in the literature than for stakeholders, and this is because changing building plan takes much more than changing for example WWR. Thus, some stakeholders said that they tend to give it a lower rating as they are not willing to change it to improve the building's performance. Another reason is that the stakeholders could give higher ratings when they are more familiar with the subject. This statement is also supported by the fact that sustainability consultants tend to give higher ratings than architects as they are more familiar with the energy optimization process. However, although stakeholders' opinions are not entirely objective and consistent, it is still reasonable to consider them. This is because the stakeholders will be the end users of the developed optimization tools, and it is crucial to make sure that their preferred ADVs are included to guarantee that they are willing to use the tools in the first place.

It is worth noting that although this thesis only focuses on the Nordic countries, the literature review of identifying ADVs is not restricted to the Nordics. Not only did many articles not specify the geographical region, but if the focus were only on the Nordics, there would not have been a sufficient number of papers for analysis. However, it is argued here that the influential ADVs for building energy do not necessarily change across the regions. This is further explained in Appendix E.

While surveys are used as the main method to gain stakeholder insights, surveys with closed-ended questions may have a lower validity rate. Further, our study involved only 12 architects and 12 consultants. To check the validity of our results, we took answers from ten randomly chosen respondents for each stakeholder category to calculate the comparative results

(Appendix F). The difference between the mean rating for 20 respondents and that for 24 respondents is small: 98% of the difference in average rating is from 0 to 0.2, with most results around 0.05. The small difference in the rating between 20 respondents and 24 respondents indicates that an increase in the number of respondents would most likely not lead to a different result.

In general, Study A tries to include stakeholders' insight into the process of developing an ML-based building energy optimization tool by including the input ADVs the stakeholders consider as influential. Incorporating stakeholders' opinions in developing ML-based tools is crucial as they have a better contextual understanding and can offer not just field-specific knowledge but also practical insights. This study chose to include stakeholders' insight in the process of selecting input ADVs as stakeholders are also the end users of the developed tool, and including the ADVs they want to use in the ML tool can ensure usability. However, according to previous research, there are also multiple other ways to introduce stakeholders in the process of developing tools [122]. For instance, stakeholders can be included in the process of parameterizing ADVs, or in the process of tool interface design to make sure it is user-friendly enough, or in defining the optimization criterion. Other methods of integrating stakeholders' opinions can be further investigated in the future.

5.2 Reflections on Developing ML Early-stage Building Energy Prediction Models.

5.2.1 Reflections on the generation of synthetic datasets.

Study B investigated the best-performing ML algorithm depending on different characteristics of synthetic datasets and provided recommendations for selecting algorithms and generating synthetic datasets. In the ML experiments' setting, the maximum dataset size was 4800 and it is defined in this study as a large dataset. However, it can be frequently seen in previous research that a training dataset that contains more than ten thousand or even twenty thousand data points is used in developing ML energy prediction models for buildings [15][70]. Having such large training datasets can for sure guarantee the model's accuracy; however, as mentioned in Chapter 1, it is not easy to acquire large existing datasets or generate large synthetic datasets. Therefore, this study chose to specifically explore how to develop an ML energy prediction model with smaller training datasets. It is summarized in the literature review that the minimum size of the training dataset in developing an ML model in building energy prediction with high performance is around 5000 data points [123][124]. Therefore, this study only investigated small datasets with data points below 5000 to see if the ML model could achieve reasonable performance with limited data points to improve efficiency and make ML more affordable for researchers and practitioners to apply in early-stage optimization.

Nine different building types were developed in parametric modeling to investigate the influence of synthetic dataset diversity on ML models' performance. As defined in this study, low-diversity datasets only contained very few building types while high-diversity datasets could contain more building types. There are a few limitations in this regard. First, the building types here were only developed to simulate how different diversity in training datasets can influence the model's performance in a theoretical way, therefore, the building types only represent the difference between building shapes and are not always the most common building shape in reality. Second, the testing dataset here has nine building types in total and this represents the diversity of the use case. In this context, a dataset that contains eight building types can be considered a high-diversity dataset. However, it is worth noting that the definition of high-diverse datasets and low-diverse datasets highly depends on the situation of the use case. For instance, when the developed ML model is mainly used for a big city with a high diversity of building styles, even a training dataset that contains eight building types can still be considered a low-diverse dataset. The scenario of the real use case should be considered when integrating the recommendation given in section 4.2.3.

Section 4.2.3 indicates that when using the best-performing algorithm, the synthetic dataset does not need to be too big. Compared with dataset size, increasing diversity has more impact on the ML model's performance after the size reaches around 1440. Therefore, when generating synthetic datasets, before the dataset size reaches 1440, the focus should be on increasing the dataset size, after that the focus should be on increasing the diversity. However, it is also worth noting that the parametric modeling in this study is specifically developed for residential buildings in Nordic countries, and the results might not be widely applicable to a more general context. The same methods could be applied to other building types, like commercial buildings or office buildings, to see if the results are the same or different. The methods could also be applied to other regions, such as tropical climates. The general trend from the findings that can be generalized is that when developing ML building energy prediction models, the size of the training dataset impacts the model's accuracy performance more when the dataset size is very small, but after reaching a certain size, enhancing training datasets diversity is more important than increasing dataset size.

However, although the training dataset size does not necessarily need to be as large as previous studies indicate, it can not be overly small. When the dataset is overly small, the ML models might not give valid results. For instance, Section 4.2.3 shows that increasing diversity in the training dataset does not improve the model's accuracy when the dataset size is smaller than 600. In the most extreme scenario, increasing diversity for a very small dataset can make the model's performance even worse. This uncommon phenomenon is probably due to the overly small size of the training dataset. Although this study investigates ML's performance on a small dataset, in general, ML is a method that highly depends on large training data. When the dataset is overly small (smaller than 600 in this case), ML might give uncommon and surprising results.

5.2.2 Reflections on the selection of best-performing ML algorithms.

Section 4.2.3 presents the recommendations for selecting the best-performing ML algorithms for different training datasets. Most previous studies choose accuracy as the main criterion for models' performance. However, since this study focuses on developing ML energy prediction models for accelerating early-stage optimization, computational efficiency is also included in performance. For future development, more criteria for models' performance could be included, such as the complexity of the model, and the difficulty of implementation. It is worth noting that in terms of accuracy, SVM is the most appropriate algorithm for almost all datasets. It might be overly ambitious to expect one single ML algorithm to effectively address the diverse range of datasets available. This could be explained by multiple reasons. Firstly, SVM might have higher adaptability than other algorithms as it is found that the optimal hyperparameters for each SVM model can be very different even if the training datasets are similar. Secondly, SVM is known as the most accurate and robust ML algorithm [125]. Thirdly, previous studies confirmed that SVM produces better outcomes in small datasets [126] [127] [128], and all training datasets used in this study can be generally considered small in size. Figure 5.1 presents the accuracy of the five models on a training dataset containing eight building types and 19000 data points. It is presented that SVM does not have the best performance when the training dataset is large.

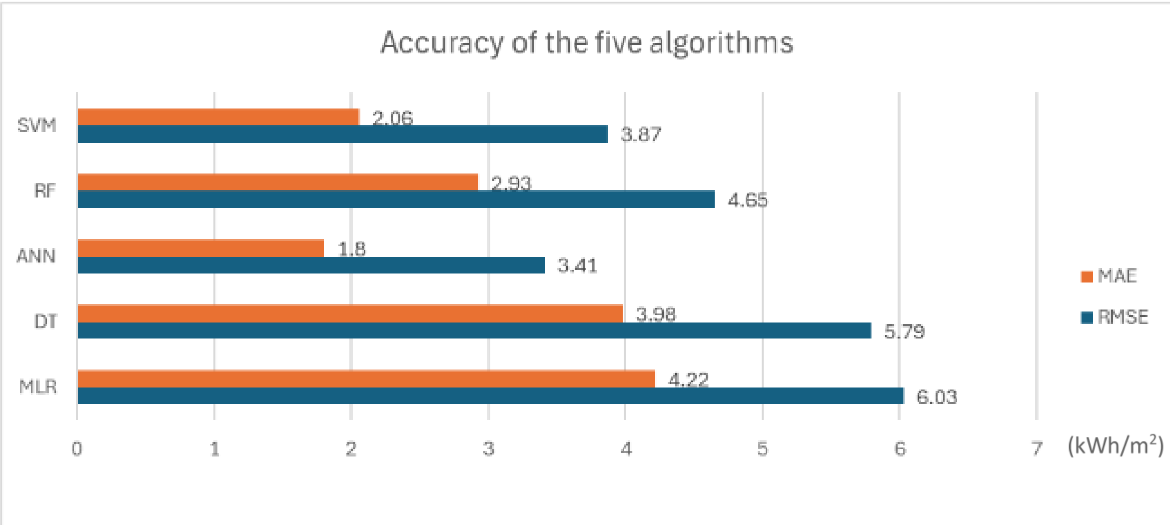


Figure 5.1. Accuracy of the five ML algorithms on a training dataset containing eight building types and 19000 data points

The limitations in the simplified hyperparameter tuning process can also influence the models' performance. The hyperparameter tuning in this study uses RandomizedSearchCV instead of GridSearchCV for less computational time. The drawback of this method is that the most optimal hyperparameter might not be found as the search process is not exhaustive. Moreover, the hyperparameter tuning process for RF is even more simplified in this study as RF's tuning time is extremely long. Therefore, RF could perform better in accuracy as the hyperparameters used in this study might not be the most optimal ones.

Among all the ML models developed with various training datasets in this study, most do not have a very high accuracy. Even the best models developed in this study can only reach an RMSE of 5.5 kWh/m^2 (around 7% on average) and an MAE of around 4 kWh/m^2 (around 5% on average), which are not as high as the ML prediction models developed in some previous studies [15][23] using large datasets in training. However, the balance between achieving sufficient accuracy and the ability to provide highly flexible and fast feedback to architects is still today's base for discussion [129]. It is argued that when the ML prediction model is developed for aiding early-stage optimization, the accuracy level does not need to be very high as it is still in the early stage where most information is not entirely in place. Therefore, it is reasonable to use the ML model trained on small datasets to acquire faster results with a compromise in accuracy. Moreover, the goal of early-stage optimization is often to find an indication of a few designs that lead to less energy demand instead of only predicting the energy demand of one building in detail. A more detailed energy simulation is usually initiated later in the design process when all information is in place and the design is complete to evaluate the energy performance.

Based on the literature review, this study only considers the five selected ML algorithms. In future studies, more ML algorithms, such as Recurrent Neural Network (RNN), Deep Neural Network (DNN), Gaussian Processes (GP), as well as hybrid methods, could be investigated to improve the accuracy.

5.3 Generalizability of Early-stage Building Energy Prediction Models to Different Scenarios.

Based on the experiments conducted in Study C and results presented in Sections 4.3.1 and 4.3.2, it is suggested that TL generally enhances model performance in accuracy when predicting energy demand in residential buildings in cities different than the base model city, compared to retraining a new ML model for the target cities from scratch. When predicting the same energy type, heating energy in this case, the improvement is particularly evident under conditions where the training dataset in the target city is more limited (below 600 data points). The benefit of TL in this context appears to be positively correlated with the similarity between the source and target cities in terms of both climate characteristics and energy demand profiles. For example, TL yielded the greatest improvement for Seattle, which shares a climate zone with Gothenburg—the source domain city—and exhibits similar heating energy demand ranges (20–60 kWh/m²/year), as shown in Figure 6. In contrast, Madrid, which lies in a distinctly different climate zone, showed the smallest performance gain.

As shown in Section 4.3.3, predicting the same energy type (heating energy) in new cities can achieve higher accuracy than predicting combined heating and cooling energy. However, when predicting building energy demand in other cities, discrepancies in energy demand composition compared to the base city are often inevitable. When the energy type changes between the source and target prediction tasks, to predict combined heating and cooling energy demand instead of heating energy in this case, TL still provides measurable improvements in most cases. However, the influence becomes smaller as the amount of training data increases. The same factors, including data availability and climate similarity, still play an important role when predicting the same type of energy use. In particular, for Miami, where energy demand is entirely cooling demand and differs climatically from Gothenburg, the improvement from TL is marginal or even negative in some cases when the data size from the target city reaches 2000. These observations suggest that while TL is a flexible approach, its effectiveness depends on the alignment of energy type and climatic context between the source and target domains. Applying TL across differing energy types is not infeasible, but it offers limited advantages compared to applying it within the same energy domain. These results align with the previous research on applying TL for image recognition tasks [130], which suggests that one of the key conditions for effective TL performance is a certain degree of similarity between the two domains.

5.4 Reflection on TL Model Development.

5.4.1 TL Model development with limited training data.

Typically, the base model for TL is trained on a large-scale dataset to ensure sufficient generalization capability across domains. However, in this study, the base model was trained on a relatively modest dataset of only 12,000 samples, and the target city training datasets are significantly smaller, ranging from just 200 to 2,000 samples. Despite this data limitation, the TL framework demonstrated notable performance improvements over the ANN models retrained from scratch. This finding highlights that even when both the source and target datasets are relatively limited in size, TL can still be effective. Moreover, by exploring the feasibility of applying TL using a base model trained on a relatively small dataset, which substantially lowers the entry barrier for implementing TL in real-world applications. In practice, acquiring large-scale labeled datasets for building energy modeling can be computationally and time-consuming, especially in early design stages when detailed information is limited. By demonstrating that TL can still be effective under limited data conditions, this study highlights the potential of lightweight and accessible TL frameworks to support building energy prediction tasks in data-scarce scenarios.

Although Study C demonstrates that a base model trained on a relatively small dataset can still be used effectively for TL model development, there are practical limitations. If the base model is trained on an extremely small dataset, its representational capacity may be insufficient to support successful knowledge transfer. In general, a larger base model dataset tends to improve TL performance. Moreover, the required size of the base model is not universal—it also depends heavily on the similarity between the base city and the target city. To investigate this aspect further, Figure 5.2 presents the TL models' performance when using three different base models of various training data sizes—6000 (Base Model 1), 9000 (Base Model 2), and 12000 samples (Base Model 3, which was used throughout the main experiments in this study)—to predict the combined heating and cooling energy demand of buildings in two target cities: Seattle and Miami. The TL approach used in all cases was Method A-3, as it has quite average performance among all TL models.

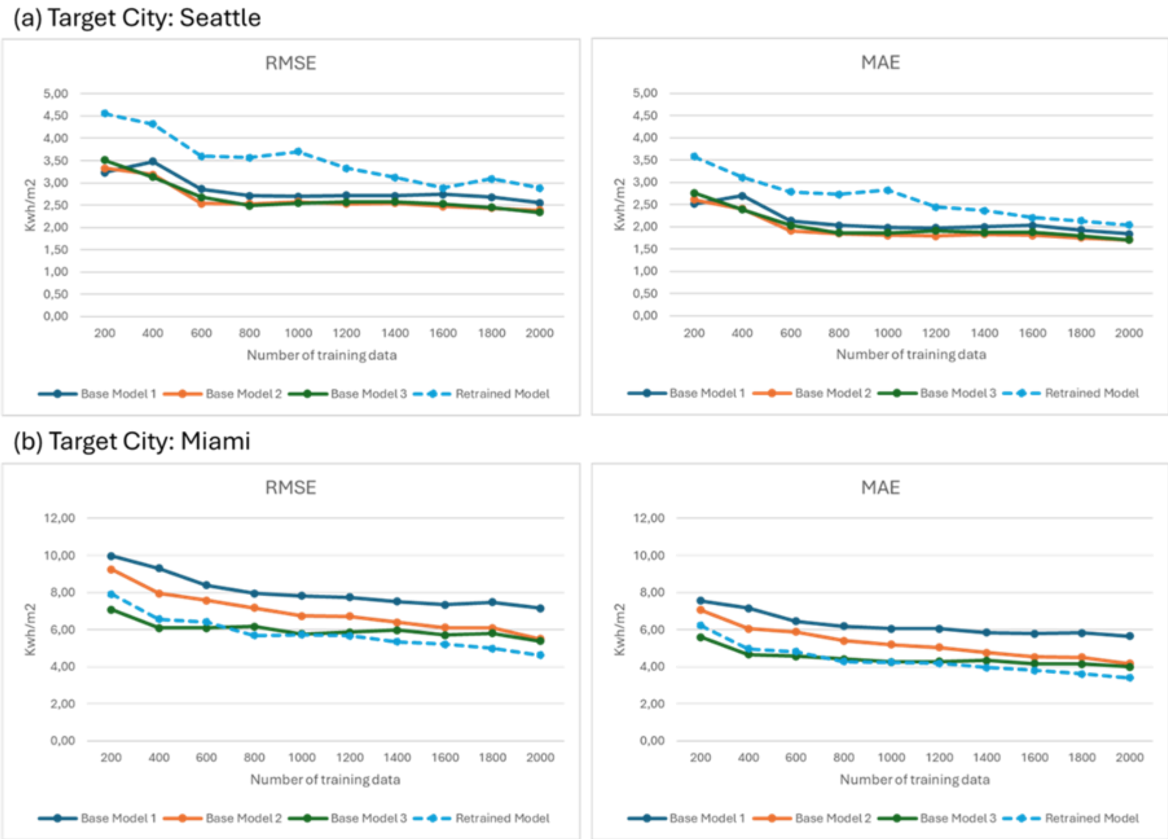


Figure 5.2. Accuracy of applying TL using base models with various data points on predicting building energy for Seattle (a) and Miami (b).

Overall, the results from Figure 5.2 suggest that the size of the base model training dataset positively correlates with TL performance. However, the degree of improvement varies with the similarity between cities. For Seattle, which is climatically closer to the base city (Gothenburg), all three base models provide performance improvement over training from scratch, though Base Model 1 performs the worst. Interestingly, Base Model 2 and Base Model 3 show similar outcomes, suggesting that beyond a certain point (9000 samples in this case), additional base data yield diminishing returns.

In contrast, for Miami, which differs substantially from Gothenburg in both climate and energy use composition, only the largest base model (Base Model 3) offers clear benefits, and only when the target training dataset is smaller than 800 samples. From the trend observed across the three models, it can be inferred that increasing the base model size beyond 12,000 samples could potentially lead to even better TL outcomes in dissimilar cities like Miami. This would even be potentially beneficial in cases where the target city dataset is larger (over 1000 samples), as the current base models struggle to outperform the retrained model in those scenarios.

These findings suggest that while TL with a base model with a limited training dataset is feasible and effective in many situations, especially for climatically similar target cities, the relationship between base model size, target city similarity, and available training data must be carefully considered. A one-size-fits-all strategy does not apply in this context. Instead, practitioners should assess the compatibility between source and target domains when deciding the necessary scale and complexity of the base model for TL.

5.4.2 TL model development with a simplified framework.

The TL framework adopted in this study is intentionally kept simple and interpretable, primarily by applying a basic layer-freezing strategy and incorporating a GRL with more new layers instead of relying on complex architecture modifications (e.g., multi-source transfer learning) [131] or advanced meta-learning techniques (e.g., Model-agnostic Meta-learning) [132]. This streamlined design contributes to significant computational efficiency, as evidenced by the short fine-tuning times observed even when only a single layer of the base model is frozen. It also facilitates easy implementation and adaptation in professional practice, especially in scenarios where model developers may not have access to high-performance computing resources. Additionally, the use of a straightforward TL strategy enhances the interpretability of the model architecture. Enhanced interpretability fosters trust and transparency, which are essential for stakeholders to understand and trust the model's predictions [133], [134]. This interpretability and simplicity not only facilitate more transparent model behavior but also help to bridge the gap between tool developers and stakeholders such as architects and consultants, enabling them to better understand, trust, and adopt ML tools in design workflows [135]. This is particularly important in multidisciplinary contexts where model transparency and ease of communication are essential for successful integration [136].

This work provides opportunities for future research in several directions. First, a more systematic analysis could be conducted to determine how the size and quality of source and target datasets interact with TL performance, particularly under residential building energy prediction. Second, integrating layer selection or adaptive freezing strategies could further enhance the robustness of the TL framework. Finally, extending this lightweight TL strategy to more diverse building types, climates, and policy contexts would help validate its broader applicability and strengthen its relevance for practical deployment.

5.4.3 TL Model for early-stage building energy optimization.

Study C demonstrates that TL can effectively transfer knowledge from one dataset comprising diverse building configurations to another with different configurations while maintaining strong performance, which makes it feasible for an energy prediction model for building early-stage optimization. This significantly enhances the feasibility of applying ML models for early-stage energy performance prediction in buildings, particularly in the context of building energy optimization during the design phase.

In contrast to most prior research on TL in building energy prediction, which typically focused on transferring models trained on one or a few buildings to one another or several similar buildings, this study emphasizes a broader scope. Specifically, early-stage energy prediction for building energy optimization relies heavily on architectural design variables such as building shape, orientation, facade configuration, including window-to-wall ratio. As such, developing a reliable prediction model for this phase requires a dataset that represents a wide diversity of building configurations rather than repeated samples from a limited number of buildings.

Another important consideration is the level of predictive accuracy required for early-stage design. Unlike detailed energy audits or operational phase simulations, early-stage energy modeling primarily serves as a decision-support tool to compare different design alternatives, rather than to generate exact energy demand values. Therefore, it does not demand the same level of precision as post-occupancy models. In this vein, Table 5.1 outlines the minimum number of training samples from the target city required to achieve acceptable prediction accuracy using TL in this study, defined as RMSE < 5 kWh/m², MAE < 4 kWh/m², MAPE < 5.5%, and $R^2 > 0.8$. These thresholds are informed by benchmarks proposed in previous research specifically for early-stage energy modeling applications [137].

Table 5.1. Minimum data requirement of developing a TL model in each target city for early-stage building energy optimization

City	Stockholm	Seattle		Chicago		Madrid		Miami
Energy Type	Heating	Heating	Heating + Cooling	Heating	Heating + Cooling	Heating	Heating + Cooling	Cooling
Required Data	800	600	600	600	600	1400	600	1800

As shown in Table 5.1, applying TL allows the development of reasonably accurate energy prediction models for early-stage building energy optimization using only a limited amount of training data from the target city.

5.5 Applying ML Prediction Model in Optimization Workflow.

In this thesis, two optimization tasks are conducted to exemplify how to implement ML models in building energy optimization workflows. Task 1 is a single objective optimization, which focuses on whether the ML model can direct the optimization process in a similar direction; Task 2 represents a multi-objective optimization problem, which is more closely aligned with real-world application scenarios. As shown in Figure 4.16, in task 1, the optimal building design configurations proposed by the two workflows are nearly identical, which proves that the ML model is capable of directing the optimization in the same direction. Figure 4.17 shows that although the simulation-based optimization proposes solutions with slightly higher quality, the difference between them and the solutions proposed by ML-based optimization is not very prominent. The improvement of computational efficiency when replacing the simulation engine with the ML model is significant. This demonstrates the ML approach's strength as a rapid and efficient tool for early-stage design support. Since design decisions will continue to evolve after the early stage, and since accurate energy performance can only be reliably simulated in later design stages, the priority at early stages is not perfect accuracy, but rather fast and informed guidance. Moreover, real-world energy use will always deviate from predictions due to factors such as weather and occupant behaviour, reinforcing the value of quick, data-driven decision support over exact estimates at this phase.

Although being a promising tool, the ML-based optimization tool does have its limitations in accuracy: the predicted value and the simulated value are not the same. This could be further seen in Appendix H.

5.6 Open-Source Code and Data Sharing.

The method developed in this thesis is built upon open-source tools, including simulation and modelling tools such as EnergyPlus, Honeybee, Grasshopper, and coding tools such as GHPython and its scientific libraries. In alignment with the values of open science, this thesis also seeks to contribute back to the academic community. All core materials developed in the project, including the synthetic datasets, the Grasshopper scripts used for parametric modeling, and supporting documentation, have been uploaded to a publicly accessible GitHub repository [138]. These resources are intended to facilitate replication of the experiments, provide reference examples for other researchers, and support methodological adaptation in future studies. In addition, the developed ML models will be further before their public release.

By openly sharing both the computational framework and the synthetic datasets, this thesis aims to contribute not only methodological insights, but also durable and practical resources that can support reproducibility and enable cumulative knowledge-building within the community. These efforts reflect a broader commitment to collaborative scientific development and the advancement of data-driven approaches in sustainable building design

5.7 Limitations and Outlook

This thesis investigates applying ML and related methods within early-stage building energy optimization workflows. Although previous chapters have demonstrated that ML models are effective for time-efficient energy prediction and can be successfully integrated into optimization workflows, ML as a data-driven method still has several limitations.

First, ML models require a large volume of training data to achieve acceptable accuracy. In the context of building energy prediction, this means generating a substantial number of building design configurations along with their corresponding energy demand results. While Study B explored the feasibility of training ML models using a limited dataset, the data generation process still requires a lot of time. Furthermore, for the ML model to perform accurately, the target building must exhibit similar characteristics to those in the training data. For instance, in Study D, the layout of the test building differed from any of the building shape in Study B's training dataset. Although the ML model was still able to perform the optimization effectively, it showed noticeable discrepancies in prediction accuracy when compared with actual simulation results (see Appendix G).

Another limitation lies in the rigidity of the ML prediction model. If end users wish to include new ADVs or modify the energy settings, the model needs to be retrained. Although Study C explored the generalizability of the ML model through TL, the results show that its effectiveness is highly dependent on the similarity between the new target domain and the source domain used for training. As it is well-established that TL performs best when the source and target domains share common features, Study C deliberately restricted the experimental scope to cities where mid-rise residential buildings dominate the housing stock. Cities like Stockholm, Seattle, Chicago, and Madrid were selected due to their comparable building typologies. Future work could expand the range of building typologies to explore how much geometric variability a TL framework can tolerate before prediction accuracy deteriorates. Additionally, further research could assess whether TL models can generalize across building types (e.g., from residential to commercial buildings) within the same city or across different urban contexts. It is also important to note that even with TL, a small amount of new data is still required for fine-tuning, meaning that data acquisition and dataset generation remain unavoidable challenges in ML-based workflows, especially when compared to traditional simulation-based methods.

Another limitation of this thesis is related to the ML development process itself. Both the ML model developed in Study B and the TL model in Study C were trained using relatively small datasets. The rationale for this choice is discussed in Section 5.2.1. However, this also introduces a constraint: the findings of this thesis are primarily applicable to ML models trained on limited data, and may not generalize to models developed using larger datasets. Moreover, the ML algorithms adopted in this thesis were intentionally basic. The five selected algorithms in Study B were chosen for their simplicity, as simpler models are computationally more efficient and faster to train and test—an important consideration when data and time are limited [139]. Additionally, complex ML models often require larger datasets to perform effectively [140], which was beyond the scope of this study. Simpler models also offer greater explainability and interpretability compared to more complex "black-box" models, allowing for better integration with design workflows and higher user trust, especially in the early design stage [141]. These models are also easier to implement in tools like Grasshopper, streamlining collaboration between designers and engineers [142]. Similarly, Study C adopted a relatively simple TL framework with minimal tuning (as explained in Section 5.4) to prioritize efficiency and clarity. However, future studies may explore more advanced TL architectures—such as progressive layer unfreezing, domain-adaptive regularization, or attention-based feature transfer—to enhance prediction accuracy, particularly in cities with energy use patterns that diverge from the original training context.

Lastly, in Study D, the ML model was embedded into the Grasshopper workflow using Flask and GHPython. While this setup is relatively lightweight and effective, it does require manual operations from users, making it less user-friendly compared to traditional optimization plug-ins that can be directly installed in Grasshopper. Future work could involve developing the ML prediction model into a dedicated plug-in for Grasshopper, allowing for seamless integration and broader adoption. Furthermore, the optimization task in Study D only investigated energy demand and embodied carbon emissions. The approach developed in this thesis could also be adapted to other early-stage sustainability objectives, such as daylighting or thermal comfort. The appended study identifies the most influential ADVs for these objectives, which can serve as inputs for developing specialized ML prediction models. Similarly, the methods used in this thesis—both in identifying suitable algorithms and in generating synthetic datasets—can be applied to develop ML models for different sustainability goals. Multi-objective optimization workflows could also benefit from ML-based prediction if developers train multiple ML models tailored to different objectives and include the corresponding ADVs within the design workflow.

Chapter 6

Conclusion

This thesis explores applying ML in early-stage building energy optimization to improve efficiency, and provides deliverable ML models for residential building energy predictions in Gothenburg, multiple TL models for residential building energy predictions in different climates, and an ML-based early-stage optimization workflow.

Study A and B together investigate how to develop an ML building energy prediction model to replace the time-consuming energy simulation in optimization processes. Recommendations for developing ML energy prediction models are provided from the perspective of identifying input architectural design variables (ADVs), selecting the most appropriate ML algorithm, and generating the training dataset. The influential ADVs for building energy optimization are defined by conducting a literature review and a stakeholder survey. The best-performing ML algorithms, as well as the acquisition of proper synthetic datasets, are investigated through multiple ML experiments. Through the literature review, it is found that building plan, window-to-wall-ratio (WWR), and wall material are considered influential in early-stage building energy optimization. However, stakeholders also consider building plan, building orientation, shading device, storey number, storey height, roof type, and roof material as influential ADVs. To get a holistic point of view, this thesis recommends including influential ADVs from both perspectives for ML energy prediction models, as it is essential to consider stakeholders' subjective points of view in developing ML-based tools. Bringing in a stakeholder's perspective is beneficial as they are experts in the fields with practical knowledge and experiences that can greatly enhance the development and effectiveness of ML-based tools. In terms of selecting the ML algorithm, Support Vector Machine (SVM) is recommended in general when developing a building energy prediction model with a training dataset smaller than 5000 as it performs well in terms of accuracy for all training datasets in the ML experiments and has an acceptable performance in computational efficiency. When the training dataset is small in size and has low diversity, Multiple Linear Regression (MLR) is recommended. Artificial Neural Network (ANN) is recommended for training datasets with larger sizes and higher diversity. Although RF does perform well in accuracy, its training time is extremely long. It is only recommended to select Random Forrest (RF) when not considering computational efficiency. In terms of generating synthetic training datasets, to achieve a reasonable accuracy performance, the dataset needs to have more than 1440 data points and a diversity that covers around 67% of the diversity in the testing dataset.

Study B developed the ML model for the Swedish context, focusing on residential buildings in Sweden, while Study C investigates the developed ML model's generalizability by applying the model to five different climatic contexts using TL. The results suggest that TL in general can be an effective strategy, but it performs best when the target city shares similar climatic conditions and energy use patterns with the base city, such as Seattle, Stockholm, and Chicago, and its effectiveness decreases when the climate difference is large, as seen in the case of Madrid and Miami. It is also suggested that TL models' performance is more evident when the target dataset contains limited data (below 600 data points), and the improvement becomes smaller as the amount of training data of the target city increases. The effectiveness of TL also differs depending on the energy type being predicted. TL is more effective when predicting heating energy than total (heating + cooling) energy demand within the same target city. Since the base model is trained specifically for heating energy in a cold climate, its internal representations align more closely with heating-dominant patterns. One of the most significant advantages of TL highlighted in Study C is its potential to drastically reduce the time and data needed to develop reasonably accurate prediction models. In some cases, TL required 1000–1600 fewer training samples to reach comparable performance to a fully retrained model. Given that generating each training sample in this study took about 7 minutes, TL could save up to 186 hours of computational time.

Study D integrates the developed ML model into an optimization workflow in Grasshopper and exemplifies the workflow using a case study. The results show that although the energy demand of the optimal building design configuration proposed by the ML-based optimization workflow is slightly higher than the simulation-based optimization workflow, the required time is reduced significantly, and the final proposals are almost identical. This proves that using an ML prediction model to replace the time-consuming energy simulation engine is feasible for guiding early-stage design decisions.

Overall, this thesis provides recommendations on how to develop the ML building energy prediction model from the perspective of selecting input ADVs, ML algorithms, and generating synthetic datasets. Moreover, this thesis also proves the feasibility of applying existing ML models to other climates with limited data. These outcomes can support researchers and software developers who want to integrate ML into the building energy optimization workflow in the early stage, and architects and consultants who want to accelerate the design optimization process.

References.

[1] C. K. Chau, T. M. Leung, and W. Y. Ng, “A review on life cycle assessment, life cycle energy assessment and life cycle carbon emissions assessment on buildings,” *Appl Energy*, vol. 143, no. 1, pp. 395–413, 2015, doi: 10.1016/j.apenergy.2015.01.023.

[2] U. N. E. Programme, “Buildings and Climate Change: Summary for Decision Makers,” 2009.

[3] D. Li, G. Huang, G. Zhang, and J. Wang, “Driving factors of total carbon emissions from the construction industry in Jiangsu Province, China,” *J Clean Prod*, vol. 276, p. 123179, 2020, doi: 10.1016/j.jclepro.2020.123179.

[4] H. Yu, W. Yang, Q. Li, and J. Li, “Optimizing Buildings’ Life Cycle Performance While Allowing Diversity in the Early Design Stage,” *Sustainability (Switzerland)*, vol. 14, no. 14, 2022, doi: 10.3390/su14148316.

[5] I. Kovacic and V. Zoller, “Building life cycle optimization tools for early design phases,” 2015. doi: 10.1016/j.energy.2015.03.027.

[6] H. Wu and T. Zhang, “Multi-objective optimization of energy, visual, and thermal performance for building envelopes in China’s hot summer and cold winter climate zone,” *Journal of Building Engineering*, vol. 59, no. April, p. 105034, 2022, doi: 10.1016/j.jobe.2022.105034.

[7] N. Delgarm, B. Sajadi, and S. Delgarm, “Multi-objective optimization of building energy performance and indoor thermal comfort: A new method using artificial bee colony (ABC),” *Energy Build*, vol. 131, pp. 42–53, 2016, doi: 10.1016/j.enbuild.2016.09.003.

[8] B. Welle, J. Haymaker, and Z. Rogers, “ThermalOpt: A methodology for automated BIM-based multidisciplinary thermal simulation for use in optimization environments,” *Build Simul*, vol. 4, no. 4, pp. 293–313, 2011, doi: 10.1007/s12273-011-0052-5.

[9] B. Lartigue, B. Lasternas, and V. Loftness, “Multi-objective optimization of building envelope for energy consumption and daylight,” *Indoor and Built Environment*, vol. 23, no. 1, pp. 70–80, 2014, doi: 10.1177/1420326X13480224.

[10] A. Li, F. Xiao, C. Fan, and M. Hu, “Development of an ANN-based building energy model for information-poor buildings using transfer learning,” *Build Simul*, vol. 14, no. 1, pp. 89–101, Feb. 2021, doi: 10.1007/s12273-020-0711-5.

[11] I. Costa-carrapiço, R. Raslan, and J. Neila, “Energy & Buildings A systematic review of genetic algorithm-based multi-objective optimisation for building retrofitting strategies towards energy efficiency,” vol. 210, 2020, doi: 10.1016/j.enbuild.2019.109690.

[12] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review," Oct. 01, 2022, *Elsevier Ltd*. doi: 10.1016/j.engappai.2022.105287.

[13] E. K. Burke and G. Kendall, "Search Methodologies Introductory Tutorials in Optimization and Decision Support Techniques Second Edition."

[14] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," *Journal of Building Engineering*, vol. 45, no. October 2021, p. 103406, 2022, doi: 10.1016/j.jobe.2021.103406.

[15] Y. Sheng, W. O. C. Ward, H. Arbabi, M. Álvarez, and M. Mayfield, "Deep multimodal learning for residential building energy prediction," *IOP Conf Ser Earth Environ Sci*, vol. 1078, no. 1, 2022, doi: 10.1088/1755-1315/1078/1/012038.

[16] S. Seyedzadeh, F. P. Rahimian, I. Glesk, and M. Roper, "Machine learning for estimation of building energy consumption and performance: a review," *Visualization in Engineering*, vol. 6, no. 1, 2018, doi: 10.1186/s40327-018-0064-7.

[17] G. Aruta, F. Ascione, O. Boettcher, R. F. De Masi, G. M. Mauro, and G. P. Vanoli, "Machine learning to predict building energy performance in different climates," *IOP Conf Ser Earth Environ Sci*, vol. 1078, no. 1, 2022, doi: 10.1088/1755-1315/1078/1/012137.

[18] K. Feng, W. Lu, and Y. Wang, "Assessing environmental performance in early building design stage: An integrated parametric design and machine learning method," *Sustain Cities Soc*, vol. 50, no. September 2018, p. 101596, 2019, doi: 10.1016/j.scs.2019.101596.

[19] N. T. Ngo, "Early predicting cooling loads for energy-efficient design in office buildings by machine learning," *Energy Build*, vol. 182, pp. 264–273, 2019, doi: 10.1016/j.enbuild.2018.10.004.

[20] A.-T. Nguyen, S. Reiter, and P. Rigo, "A review on simulation-based optimization methods applied to building performance analysis," *Appl Energy*, vol. 113, pp. 1043–1058, Aug. 2014, doi: 10.1016/j.apenergy.2013.08.061.

[21] S. Mohammed *et al.*, "The Effects of Data Quality on Machine Learning Performance on Tabular Data," May 2025, doi: 10.1016/j.is.2025.102549.

[22] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Machine learning for energy performance prediction at the design stage of buildings," *Energy for Sustainable Development*, vol. 66, pp. 12–25, 2022, doi: 10.1016/j.esd.2021.11.002.

[23] E. Elbeltagi and H. Wefki, "Predicting energy consumption for residential buildings using ANN through parametric modeling," *Energy Reports*, vol. 7, pp. 2534–2545, 2021, doi: 10.1016/j.egyr.2021.04.053.

[24] M. Płoszaj-Mazurek, E. Ryńska, and M. Grochulska-Salak, “Methods to Optimize Carbon Footprint of Buildings in Regenerative Architectural Design with the Use of Machine Learning, Convolutional Neural Network, and Parametric Design,” *Energies (Basel)*, no. Figure 1, 2020.

[25] R. Pino-Mejías, A. Pérez-Fargallo, C. Rubio-Bellido, and J. A. Pulido-Arcas, “Comparison of linear regression and artificial neural networks models to predict heating and cooling energy demand, energy consumption and CO2 emissions,” *Energy*, vol. 118, pp. 24–36, 2017, doi: 10.1016/j.energy.2016.12.022.

[26] M. W. Ahmad, M. Mourshed, and Y. Rezgui, “Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption,” *Energy Build*, vol. 147, pp. 77–89, 2017, doi: 10.1016/j.enbuild.2017.04.038.

[27] Z. Liu *et al.*, “Accuracy analyses and model comparison of machine learning adopted in building energy consumption prediction,” *Energy Exploration and Exploitation*, vol. 37, no. 4, pp. 1426–1451, 2019, doi: 10.1177/0144598718822400.

[28] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, “Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques,” *Journal of Building Engineering*, vol. 45, no. September 2021, p. 103406, 2022, doi: 10.1016/j.jobe.2021.103406.

[29] M. M. Singh, S. Singaravel, and P. Geyer, “Machine learning for early stage building energy prediction: Increment and enrichment,” *Appl Energy*, vol. 304, no. May, p. 117787, 2021, doi: 10.1016/j.apenergy.2021.117787.

[30] F. Zhuang *et al.*, “A Comprehensive Survey on Transfer Learning,” Jan. 01, 2021, *Institute of Electrical and Electronics Engineers Inc.* doi: 10.1109/JPROC.2020.3004555.

[31] P. Agee, W. O’brien, J. Day, and C. Brackley, “Toward a user-centered built environment,” *Sci Technol Built Environ*, vol. 26, no. 9, pp. 1163–1164, 2020, doi: 10.1080/23744731.2020.1810380.

[32] Y. Zhang, Y. Chen, and X. Li, “Integrated framework of knowledge-based decision support system for user-centered residential design,” *Expert Syst Appl*, vol. 216, no. December 2022, p. 119412, 2023, doi: 10.1016/j.eswa.2022.119412.

[33] S. R. 3 and M. J. K. 1 Da Yeon Park 1, Jungsik Choi 2,*, “A User-Centered Approach to the Application of BIM in Smart Working Environments,” *Sensors*, no. 106, pp. 22–25, 2013, doi: 10.2105/ajph.27.4.378.

[34] X. Xie and Z. Gou, “Building Performance Simulation as an Early Intervention or Late Verification in Architectural Design : Same Performance Outcome but Different Design Solutions,” pp. 45–62, 2015.

[35] J. Khan, B. Johansson, and R. Hildingsson, “Strategies for greening the economy in three Nordic countries,” *Environmental Policy and Governance*, vol. 31, no. 6, pp. 592–604, 2021, doi: 10.1002/eet.1967.

[36] O. Mont, E. Heiskanen, K. Power, and H. Kuusi, *Improving Nordic policymaking by dispelling myths on sustainable consumption*. 2013.

[37] I. Siksnelyte-Butkiene, D. Streimikiene, and T. Balezentis, “Multi-criteria analysis of heating sector sustainability in selected North European countries,” *Sustain Cities Soc*, vol. 69, no. February, p. 102826, 2021, doi: 10.1016/j.scs.2021.102826.

[38] Royal Institute of British Architects, “RIBA plan of work 2020,” pp. 10–11, 2020, doi: 10.4324/9780429347177-2.

[39] T. Säwén, “Early Stage Architectural Design Practice Perspectives on Life Cycle Building Performance Assessment,” 2023.

[40] B. C. Paulson Jr, *Designing to reduce construction costs*. Journal of the construction division 102.4 (1976): 587-592., 1976.

[41] Jr. , B. C. Paulson and M. Asce, “Designing to reduce construction costs,” *Journal of the construction division*, vol. 102, no. 4, pp. 587–592, Dec. 1976.

[42] X. Shi, Z. Tian, W. Chen, B. Si, and X. Jin, “A review on building energy efficient design optimization rom the perspective of architects,” *Renewable and Sustainable Energy Reviews*, vol. 65, pp. 872–884, Aug. 2016, doi: 10.1016/j.rser.2016.07.050.

[43] M. Marwan, “The effect of wall material on energy cost reduction in building,” *Case Studies in Thermal Engineering*, vol. 17, no. September 2019, p. 100573, 2020, doi: 10.1016/j.csite.2019.100573.

[44] Y. Shahbazi, M. Heydari, and F. Haghparast, “An early-stage design optimization for office buildings’ façade providing high-energy performance and daylight,” *Indoor and Built Environment*, vol. 28, no. 10, pp. 1350–1367, 2019, doi: 10.1177/1420326X19840761.

[45] S. Alsaadani, C. Bleil, and D. Souza, “Energy Research & Social Science Of collaboration or condemnation ? Exploring the promise and pitfalls of architect-consultant collaborations for building performance simulation,” *Chem Phys Lett*, vol. 19, pp. 21–36, 2016, doi: 10.1016/j.erss.2016.04.016.

[46] K. Dooley and K. Dooley, “Capturing the Stakeholder Values of a Construction Project Capturing the Stakeholder Values of a Construction Project,” no. September 2010, 2014.

[47] T. Häkkinen, M. Kuittinen, A. Ruuska, and N. Jung, “Reducing embodied carbon during the design process of buildings,” *Journal of Building Engineering*, vol. 4, pp. 1–13, 2015.

[48] S. H. Kim and J. Nam, “Can both the economic value and energy performance of small-and mid-sized buildings be satisfied? Development of a design expert system in the context of Korea,” *Sustainability (Switzerland)*, vol. 12, no. 12, 2020, doi: 10.3390/SU12124946.

[49] M. Balouktsi, T. Lützkendorf, and G. Foliente, “Embodied impacts in stakeholder decision - making in the construction sector,” *SBE16 Hamburg*, no. March, 2016.

[50] H. X. Li, D. Patel, M. Al-Hussein, H. Yu, and M. Gül, “Stakeholder studies and the social networks of NetZero energy homes (NZEHs),” *Sustain Cities Soc*, vol. 38, no. June 2017, pp. 9–17, 2018, doi: 10.1016/j.scs.2017.12.014.

[51] R. J. Yang and P. X. W. Zou, “Stakeholder-associated risks and their interactions in complex green building projects: A social network model,” *Build Environ*, vol. 73, pp. 208–222, 2014, doi: 10.1016/j.buildenv.2013.12.014.

[52] T. Østergård, R. L. Jensen, and S. E. Maagaard, “Building simulations supporting decision making in early design – A review,” *Renewable and Sustainable Energy Reviews*, vol. 61, pp. 187–201, Jul. 2016, doi: 10.1016/j.rser.2016.03.045.

[53] A. Hollberg, “Parametric Life Cycle Assessment: Introducing a time-efficient method for environmental building design optimization”, doi: 10.25643/bauhaus-universitaet.3800.

[54] A. Ciardiello, F. Rosso, J. Dell’Olmo, V. Ciancio, M. Ferrero, and F. Salata, “Multi-objective approach to the optimization of shape and envelope in building energy design,” *Appl Energy*, vol. 280, no. October, p. 115984, 2020, doi: 10.1016/j.apenergy.2020.115984.

[55] S. Lu, X. Tang, L. Ji, and D. Tu, “Research on energy-saving optimization for the performance parameters of rural-building shape and envelope by trnsys-genopt in hot summer and cold winter zone of China,” *Sustainability (Switzerland)*, vol. 9, no. 2, 2017, doi: 10.3390/su9020294.

[56] A. Sonta, T. R. Dougherty, and R. K. Jain, “Data-driven optimization of building layouts for energy efficiency,” *Energy Build*, vol. 238, p. 110815, 2021, doi: 10.1016/j.enbuild.2021.110815.

[57] V. J. L. Gan, H. K. Wong, K. T. Tse, J. C. P. Cheng, I. M. C. Lo, and C. M. Chan, “Simulation-based evolutionary optimization for energy-efficient layout plan design of high-rise residential buildings,” *J Clean Prod*, vol. 231, pp. 1375–1388, 2019, doi: 10.1016/j.jclepro.2019.05.324.

[58] A. Kusiak, M. Li, and F. Tang, “Modeling and optimization of HVAC energy consumption,” *Appl Energy*, vol. 87, no. 10, pp. 3092–3102, 2010, doi: 10.1016/j.apenergy.2010.04.008.

[59] Z. Liu and A. Guo, "Application of green building materials and multi-objective energy-saving optimization design," *International Journal of Heat and Technology*, vol. 39, no. 1, pp. 299–308, 2021, doi: 10.18280/ijht.390133.

[60] A. Heydari, S. E. Sadati, and M. R. Gharib, "Effects of different window configurations on energy consumption in building: Optimization and economic analysis," *Journal of Building Engineering*, vol. 35, no. May 2020, 2021, doi: 10.1016/j.jobe.2020.102099.

[61] Y. Zhai, Y. Wang, Y. Huang, and X. Meng, "A multi-objective optimization methodology for window design considering energy consumption, thermal environment and visual performance," *Renew Energy*, vol. 134, pp. 1190–1199, 2019, doi: 10.1016/j.renene.2018.09.024.

[62] F. Kheiri, "A review on optimization methods applied in energy-efficient building geometry and envelope design," *Renewable and Sustainable Energy Reviews*, vol. 92, no. May, pp. 897–920, 2018, doi: 10.1016/j.rser.2018.04.080.

[63] E. Asadi, M. G. da Silva, C. H. Antunes, and L. Dias, "A multi-objective optimization model for building retrofit strategies using TRNSYS simulations, GenOpt and MATLAB," *Build Environ*, vol. 56, pp. 370–378, 2012, doi: 10.1016/j.buildenv.2012.04.005.

[64] Y. Bhanu Sandhya, B. Vikram Anand, and G. R. K. D. Satya Prasad, "Energy-efficient design optimization of a building envelope using DOE-2," *International Journal of Recent Technology and Engineering*, vol. 8, no. 3, pp. 1157–1162, 2019, doi: 10.35940/ijrte.C4271.098319.

[65] R. Kamal, F. Moloney, C. Wickramaratne, A. Narasimhan, and D. Y. Goswami, "Strategic control and cost optimization of thermal energy storage in buildings using EnergyPlus," *Appl Energy*, vol. 246, no. April, pp. 77–90, 2019, doi: 10.1016/j.apenergy.2019.04.017.

[66] P. C. FISHER D, LIESEN R, STRAND R, "ASHRAE Toolkit for Building Load Calculations," *ASHRAE Transactions 109*, 2003.

[67] R. K. Arora, *Optimization: algorithms and applications*. CRC press, 2015.

[68] D. E. Golberg, "Genetic algorithms in search, optimization, and machine learning. Addion Wesley.," *Reading*, vol. 673, 1989.

[69] S. Seyedzadeh, F. P. Rahimian, S. Oliver, I. Glesk, and B. Kumar, "Data driven model improved by multi-objective optimisation for prediction of building energy loads," *Autom Constr*, vol. 116, no. November 2019, p. 103188, 2020, doi: 10.1016/j.autcon.2020.103188.

[70] E. Elbeltagi and H. Wefki, "Predicting energy consumption for residential buildings using ANN through parametric modeling," *Energy Reports*, vol. 7, pp. 2534–2545, 2021, doi: 10.1016/j.egyr.2021.04.053.

[71] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1998.

[72] K. Gurney, *An Introduction to Neural Networks*. CRC press, 2018.

[73] Z. Li, J. Dai, H. Chen, and B. Lin, “An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage,” *Build Simul*, vol. 12, no. 4, pp. 665–681, 2019, doi: 10.1007/s12273-019-0538-0.

[74] B. A. Sonkamble and D. D. Doye, “An overview of speech recognition system based on the support vector machines,” *Proceedings of the International Conference on Computer and Communication Engineering 2008, ICCCE08: Global Links for Human Development*, pp. 768–771, 2008, doi: 10.1109/ICCCE.2008.4580709.

[75] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, “Machine learning for energy performance prediction at the design stage of buildings,” *Energy for Sustainable Development*, vol. 66, pp. 12–25, 2022, doi: 10.1016/j.esd.2021.11.002.

[76] X. Fu, Z. Lei, K. Wang, and J. Yan, “Low cost carrier competition and route entry in an emerging but regulated aviation market - The case of China,” *Transp Res Part A Policy Pract*, vol. 79, pp. 3–16, 2015, doi: 10.1016/j.tra.2015.03.020.

[77] B. Dong, C. Cao, and S. E. Lee, “Applying support vector machines to predict building energy consumption in tropical region,” *Energy Build*, vol. 37, no. 5, pp. 545–553, 2005, doi: 10.1016/j.enbuild.2004.09.009.

[78] W. Tian, S. Yang, J. Zuo, Z. Y. Li, and Y. L. Liu, “Relationship between built form and energy performance of office buildings in a severe cold Chinese region,” *Build Simul*, vol. 10, no. 1, pp. 11–24, 2017, doi: 10.1007/s12273-016-0314-3.

[79] Y. Fang, X. Lu, and H. Li, “A random forest-based model for the prediction of construction-stage carbon emissions at the early design stage,” *J Clean Prod*, vol. 328, no. October, p. 129657, 2021, doi: 10.1016/j.jclepro.2021.129657.

[80] M. M. Singh, C. Deb, and P. Geyer, “Early-stage design support combining machine learning and building information modelling,” *Autom Constr*, vol. 136, no. December 2021, p. 104147, 2022, doi: 10.1016/j.autcon.2022.104147.

[81] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, “Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review,” *IEEE J Sel Top Appl Earth Obs Remote Sens*, vol. 13, pp. 6308–6325, 2020, doi: 10.1109/JSTARS.2020.3026724.

[82] C. Robinson *et al.*, “Machine learning approaches for estimating commercial building energy consumption,” *Appl Energy*, vol. 208, no. September, pp. 889–904, 2017, doi: 10.1016/j.apenergy.2017.09.060.

[83] F. Tahmasebinia, R. Jiang, S. Sepasgozar, J. Wei, Y. Ding, and H. Ma, "Using Regression Model to Develop Green Building Energy Simulation by BIM Tools," *Sustainability (Switzerland)*, vol. 14, no. 10, 2022, doi: 10.3390/su14106262.

[84] S. Papadopoulos and E. Azar, "Integrating building performance simulation in agent-based modeling using regression surrogate models: A novel human-in-the-loop energy modeling approach," *Energy Build*, vol. 128, pp. 214–223, Aug. 2016, doi: 10.1016/j.enbuild.2016.06.079.

[85] P. Domingos, "A few useful things to know about machine learning," *Commun ACM*, vol. 55, no. 10, pp. 78–87, 2012, doi: 10.1145/2347736.2347755.

[86] G. K. F. Tso and K. K. W. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy*, vol. 32, no. 9, pp. 1761–1768, 2007, doi: 10.1016/j.energy.2006.11.010.

[87] Z. Yu, F. Haghighat, B. C. M. Fung, and H. Yoshino, "A decision tree method for building energy demand modeling," *Energy Build*, vol. 42, no. 10, pp. 1637–1646, 2010, doi: 10.1016/j.enbuild.2010.04.006.

[88] M. Raghu, C. Zhang, G. Brain, J. Kleinberg, and S. Bengio, "Transfusion: Understanding Transfer Learning for Medical Imaging."

[89] X. Fang, G. Gong, G. Li, L. Chun, W. Li, and P. Peng, "A hybrid deep transfer learning strategy for short term cross-building energy prediction," *Energy*, vol. 215, Jan. 2021, doi: 10.1016/j.energy.2020.119208.

[90] C. Fan *et al.*, "Statistical investigations of transfer learning-based methodology for short-term building energy predictions," *Appl Energy*, vol. 262, Mar. 2020, doi: 10.1016/j.apenergy.2020.114499.

[91] L. Cai, H. Wen, J. Gu, J. Ma, and Z. Jin, "Forecasting customers' response to incentives during peak periods: A transfer learning approach," in *International Transactions on Electrical Energy Systems*, John Wiley and Sons Ltd, Jul. 2020. doi: 10.1002/2050-7038.12251.

[92] S. M. Jung, S. Park, S. W. Jung, and E. Hwang, "Monthly electric load forecasting using transfer learning for smart cities," *Sustainability (Switzerland)*, vol. 12, no. 16, Aug. 2020, doi: 10.3390/SU12166364.

[93] I. Ozer, S. B. Efe, and H. Ozbay, "A combined deep learning application for short term load forecasting," *Alexandria Engineering Journal*, vol. 60, no. 4, pp. 3807–3818, Aug. 2021, doi: 10.1016/j.aej.2021.02.050.

[94] F. Qian, W. Gao, Y. Yang, and D. Yu, "Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption," *Energy*, vol. 193, Feb. 2020, doi: 10.1016/j.energy.2019.116724.

[95] D. Ridley, *The Literature Review: A Step-by-Step Guide for Students*. SAGE, 2012.

[96] A. Acharya and A. Acharya, “Systematic approaches to a successful literature review,” *Perspectives on Terrorism*, vol. 7, no. 3, pp. 158–159, 2013.

[97] M. A. C. Richard A. Krueger, *Focus groups: A practical guide for applied research*. SAGE, 2014.

[98] C. Hewson, *Research design and tools for online research*. SAGE, 2017.

[99] T. Dogan, E. Saratsis, and C. Reinhart, “The optimization potential of floor-plan typologies in early design energy modeling,” *14th International Conference of IBPSA - Building Simulation 2015, BS 2015, Conference Proceedings*, no. December, pp. 1853–1860, 2015, doi: 10.26868/25222708.2015.2455.

[100] A. Gonzalez-Caceres, S. Somanath, J. Forsberg, A. Galimshina, S. Theißen, and A. Hollberg, “Application and validation of a method to assess the energy reduction and environmental impact of renovation alternatives,” *IOP Conf Ser Earth Environ Sci*, vol. 1196, no. 1, 2023, doi: 10.1088/1755-1315/1196/1/012104.

[101] S. C. Ng, “Principal component analysis to reduce dimension on digital image,” *Procedia Comput Sci*, vol. 111, pp. 113–119, 2017, doi: 10.1016/j.procs.2017.06.017.

[102] J. Ma and Y. Yuan, “Dimension reduction of image deep feature using PCA,” *J Vis Commun Image Represent*, vol. 63, 2019, doi: 10.1016/j.jvcir.2019.102578.

[103] A. Janecek, W. N. W. Gansterer, M. Demel, and G. Ecker, “On the Relationship Between Feature Selection and Classification Accuracy.,” *Fsdm*, vol. 4, pp. 90–105, 2008.

[104] N. Sharma and K. Saroha, “A novel dimensionality reduction method for cancer dataset using PCA and Feature Ranking,” *2015 International Conference on Advances in Computing, Communications and Informatics, ICACCI 2015*, pp. 2261–2264, 2015, doi: 10.1109/ICACCI.2015.7275954.

[105] K. U. V. R. Teja, B. P. V. Reddy, A. Likith Preetham, H. Y. Patil, and T. Poorna Chandra, “Prediction of Diabetes at Early Stage with Supplementary Polynomial Features,” *Proceedings - 1st International Conference on Smart Technologies Communication and Robotics, STCR 2021*, no. October, pp. 1–5, 2021, doi: 10.1109/STCR51658.2021.9588849.

[106] R. N. Kentaro Ito, “Optimizing Support Vector Regression Hyperparameters Based on Cross-Validation,” *Proceedings of the International Joint Conference on Neural Networks, 2003., Portland, OR, USA, 2003, pp.*, vol. 3, pp. 2077–2082, 2003.

[107] R. M. Neal, *Pattern Recognition and Machine Learning*, vol. 49, no. 3. 2007. doi: 10.1198/tech.2007.s518.

[108] M. C. Peel, B. L. Finlayson, and T. A. McMahon, “Updated world map of the Köppen-Geiger climate classification,” *Hydrol Earth Syst Sci*, vol. 11, no. 5, pp. 1633–1644, Oct. 2007, doi: 10.5194/hess-11-1633-2007.

[109] F. Kheiri, J. S. Haberl, and J. C. Baltazar, “Impact of outdoor humidity conditions on building energy performance and environmental footprint in the degree days-based climate classification,” *Energy*, vol. 283, Nov. 2023, doi: 10.1016/j.energy.2023.128447.

[110] R. Gupta, J. Mathur, and V. Garg, “Assessment of climate classification methodologies used in building energy efficiency sector,” Nov. 01, 2023, *Elsevier Ltd*. doi: 10.1016/j.enbuild.2023.113549.

[111] F. Gerace, D. Doimo, S. S. Mannelli, L. Saglietti, and A. Laio, “Optimal transfer protocol by incremental layer defrosting,” Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.01429

[112] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, “How transferable are features in deep neural networks?”

[113] K. Osumi, T. Yamashita, and H. Fujiyoshi, “Domain Adaptation using a Gradient Reversal Layer with Instance Weighting,” 2019.

[114] X. Fang, G. Gong, L. Chun, W. Li, and P. Peng, “Selection and peer-review under responsibility of the scientific committee of CUE2020 A transfer learning method for building energy prediction using long short term memory and domain adversarial neural network.”

[115] “15978: sustainability of construction works-assessment of environmental performance of buildings-calculation method, Available at European Comittee for Standardization.”

[116] “Boverket´s mandatory provisions and general recommendations, BBR, BFS 2011:6 with amendments up to BFS 2018:4.”

[117] T. Bartz-Beielstein *et al.*, “Benchmarking in Optimization: Best Practice and Open Issues,” Dec. 2020, [Online]. Available: http://arxiv.org/abs/2007.03488

[118] G. Goos *et al.*, “LNCS 4403 - Evolutionary Multi-Criterion Optimization.”

[119] Y. Pan *et al.*, “Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies,” Jun. 01, 2023, *Elsevier Ltd*. doi: 10.1016/j.adapen.2023.100135.

[120] Y. Zhou, V. W. Tam, and K. N. Le, “Sensitivity analysis of design variables in life-cycle environmental impacts of buildings,” *Journal of Building Engineering*, vol. 65, Apr. 2023, doi: 10.1016/j.jobe.2022.105749.

[121] W. A. Mahar, G. Verbeeck, S. Reiter, and S. Attia, “Sensitivity analysis of passive design strategies for residential buildings in cold semi-arid climates,” *Sustainability (Switzerland)*, vol. 12, no. 3, Feb. 2020, doi: 10.3390/su12031091.

[122] T. Säwén, A. Sasic Kalagasidis, and A. Hollberg, "Critical perspectives on life cycle building performance assessment tool reviews," *Renewable and Sustainable Energy Reviews*, vol. 197, no. April, p. 114407, 2024, doi: 10.1016/j.rser.2024.114407.

[123] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Machine learning for energy performance prediction at the design stage of buildings," *Energy for Sustainable Development*, vol. 66, pp. 12–25, 2022, doi: 10.1016/j.esd.2021.11.002.

[124] S. S. Gilan and B. Dilkina, "Sustainable building design: A challenge at the intersection of machine learning and design optimization," *AAAI Workshop - Technical Report*, vol. WS-15-06, pp. 101–106, 2015.

[125] X. Wu *et al.*, *Top 10 algorithms in data mining*, vol. 14, no. 1. 2008. doi: 10.1007/s10115-007-0114-2.

[126] P. Aversa, A. Donatelli, G. Piccoli, and V. A. M. Luprano, "Improved Thermal Transmittance Measurement with HFM Technique on Building Envelopes in the Mediterranean Area," *Selected Scientific Papers - Journal of Civil Engineering*, vol. 11, no. 2, pp. 39–52, 2016, doi: 10.1515/sspjce-2016-0017.

[127] Q. Li, Q. Meng, J. Cai, H. Yoshino, and A. Mochida, "Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks," *Energy Convers Manag*, vol. 50, no. 1, pp. 90–96, 2009, doi: 10.1016/j.enconman.2008.08.033.

[128] Q. Li, P. Ren, and Q. Meng, "Prediction model of annual energy consumption of residential buildings," *2010 International Conference on Advances in Energy Engineering, ICAEE 2010*, pp. 223–226, 2010, doi: 10.1109/ICAEE.2010.5557576.

[129] K. Negendahl and T. R. Nielsen, "Building energy optimization in the early design stages: A simplified method," *Energy Build*, vol. 105, pp. 88–99, 2015, doi: 10.1016/j.enbuild.2015.06.087.

[130] M. Bernico, Y. Li, and D. Zhang, "Investigating the impact of data volume and domain similarity on transfer learning applications," in *Advances in Intelligent Systems and Computing*, Springer Verlag, 2019, pp. 53–62. doi: 10.1007/978-3-030-02683-7_5.

[131] A. Kalaivani, P. Jayapriya, and A. Sangeetha Devi, "Innovative Approaches in Deep Neural Networks: Enhancing Performance through Transfer Learning Techniques," in *3rd International Conference on Automation, Computing and Renewable Systems, ICACRS 2024 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 857–864. doi: 10.1109/ICACRS62842.2024.10841600.

[132] P. Zhou, Y. Zou, X.-T. Yuan, J. Feng, C. Xiong, and S. Hoi, "Task Similarity Aware Meta Learning: Theory-inspired Improvement on MAML."

[133] M. Du, N. Liu, and X. Hu, "Techniques for Interpretable Machine Learning," May 2019, [Online]. Available: http://arxiv.org/abs/1808.00033

[134] K. Rath and L. J. Gnanasigamani, “Understanding Training Behaviour during Transfer Learning by drawing parallels to Grokking,” in *2024 4th International Conference on Intelligent Technologies, CONIT 2024*, Institute of Electrical and Electronics Engineers Inc., 2024. doi: 10.1109/CONIT61985.2024.10626430.

[135] S. R. Hong, J. Hullman, and E. Bertini, “Human Factors in Model Interpretability: Industry Practices, Challenges, and Needs,” *Proc ACM Hum Comput Interact*, vol. 4, no. CSCW1, May 2020, doi: 10.1145/3392878.

[136] E. Biliri *et al.*, “Explainable Artificial Intelligence Bundles for Algorithm Lifecycle Management in the Manufacturing Domain,” in *Proceedings of the 29th International Conference on Engineering, Technology, and Innovation: Shaping the Future, ICE 2023*, Institute of Electrical and Electronics Engineers Inc., 2023. doi: 10.1109/ICE/ITMC58018.2023.10332317.

[137] X. Wang, Y. Yu, R. Teigland, and A. Hollberg, “Size or diversity? Synthetic dataset recommendations for machine learning heating energy prediction models in early design stages for residential buildings,” *Energy and AI*, vol. 21, p. 100557, Sep. 2025, doi: 10.1016/j.egyai.2025.100557.

[138] Xinyue Wang, “Early-stage building energy dataset,” https://github.com/SB-Chalmers/early-stage-building-energy-dataset.

[139] J. Ashraf, R. Azam, A. A. Rifa, and M. J. Rana, “Multiple machine learning models for predicting annual energy consumption and demand of office buildings in subtropical monsoon climate,” *Asian Journal of Civil Engineering*, Jan. 2024, doi: 10.1007/s42107-024-01190-x.

[140] M. Adegoke, A. Hafiz, S. Ajayi, and R. Olu-Ajayi, “Application of Multilayer Extreme Learning Machine for Efficient Building Energy Prediction,” *Energies (Basel)*, vol. 15, no. 24, Dec. 2022, doi: 10.3390/en15249512.

[141] C. Fan, F. Xiao, C. Yan, C. Liu, Z. Li, and J. Wang, “A novel methodology to explain and evaluate data-driven building energy performance models based on interpretable machine learning,” *Appl Energy*, vol. 235, pp. 1551–1560, Feb. 2019, doi: 10.1016/j.apenergy.2018.11.081.

[142] H. Guo, D. Duan, J. Yan, K. Ding, F. Xiang, and R. Peng, “Machine Learning-Based Method for Detached Energy-Saving Residential Form Generation,” *Buildings*, vol. 12, no. 10, Oct. 2022, doi: 10.3390/buildings12101504.

[143] K. Amasyali and N. M. El-Gohary, “A review of data-driven building energy consumption prediction studies,” *Renewable and Sustainable Energy Reviews*, vol. 81, no. March 2017, pp. 1192–1205, 2018, doi: 10.1016/j.rser.2017.04.095.

[144] Z. Wang, J. Liu, Y. Zhang, H. Yuan, R. Zhang, and R. S. Srinivasan, “Practical issues in implementing machine-learning models for building energy efficiency: Moving

beyond obstacles," *Renewable and Sustainable Energy Reviews*, vol. 143, no. August 2020, p. 110929, 2021, doi: 10.1016/j.rser.2021.110929.

[145] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine Learning, Deep Learning and Statistical Analysis for forecasting building energy consumption — A systematic review," *Eng Appl Artif Intell*, vol. 115, no. August, p. 105287, 2022, doi: 10.1016/j.engappai.2022.105287.

[146] L. Zhang *et al.*, "A review of machine learning in building load prediction," *Appl Energy*, vol. 285, no. July 2020, p. 116452, 2021, doi: 10.1016/j.apenergy.2021.116452.

[147] P. W. Tien, S. Wei, J. Darkwa, C. Wood, and J. K. Calautit, "Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality – A Review," *Energy and AI*, vol. 10, no. August, p. 100198, 2022, doi: 10.1016/j.egyai.2022.100198.

[148] M. Bourdeau, X. qiang Zhai, E. Nefzaoui, X. Guo, and P. Chatellier, "Modeling and forecasting building energy consumption: A review of data-driven techniques," *Sustain Cities Soc*, vol. 48, no. November 2018, p. 101533, 2019, doi: 10.1016/j.scs.2019.101533.

[149] Y. Sun, F. Haghighat, and B. C. M. Fung, "A review of the-state-of-the-art in data-driven approaches for building energy prediction," *Energy Build*, vol. 221, p. 110022, 2020, doi: 10.1016/j.enbuild.2020.110022.

[150] A. Mosavi, M. Salimi, S. F. Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies (Basel)*, vol. 12, no. 7, 2019, doi: 10.3390/en12071301.

[151] J. Runge and R. Zmeureanu, "Forecasting energy use in buildings using artificial neural networks: A review," *Energies (Basel)*, vol. 12, no. 17, 2019, doi: 10.3390/en12173254.

[152] S. Seyedzadeh, F. Pour Rahimian, P. Rastogi, and I. Glesk, "Tuning machine learning models for prediction of building energy loads," *Sustain Cities Soc*, vol. 47, no. January, 2019, doi: 10.1016/j.scs.2019.101484.

[153] A. D. Pham, N. T. Ngo, T. T. Ha Truong, N. T. Huynh, and N. S. Truong, "Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability," *J Clean Prod*, vol. 260, p. 121082, 2020, doi: 10.1016/j.jclepro.2020.121082.

[154] Y. Zhao, C. Zhang, Y. Zhang, Z. Wang, and J. Li, "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis," *Energy and Built Environment*, vol. 1, no. 2, pp. 149–164, 2020, doi: 10.1016/j.enbenv.2019.11.003.

[155] M. Elsisi, M. Amer, A. Dababat, and C. L. Su, "A comprehensive review of machine learning and IoT solutions for demand side energy management, conservation, and resilient operation," *Energy*, vol. 281, no. March, p. 128256, 2023, doi: 10.1016/j.energy.2023.128256.

[156] R. Olu-Ajayi, H. Alaka, H. Owolabi, L. Akanbi, and S. Ganiyu, "Data-Driven Tools for Building Energy Consumption Prediction: A Review," *Energies (Basel)*, vol. 16, no. 6, pp. 1–20, 2023, doi: 10.3390/en16062574.

[157] H. Liu, J. Liang, Y. Liu, and H. Wu, "A Review of Data-Driven Building Energy Prediction," *Buildings*, vol. 13, no. 2, 2023, doi: 10.3390/buildings13020532.

[158] S. Seyedzadeh and F. Pour Rahimian, *Machine learning for building energy forecasting*. 2021. doi: 10.1007/978-3-030-64751-3_4.

[159] Q. Shi, C. Liu, and C. Xiao, "Machine learning in building energy management: A critical review and future directions," *Frontiers of Engineering Management*, vol. 9, no. 2, pp. 239–256, 2022, doi: 10.1007/s42524-021-0181-1.

[160] T. C. Brito and M. A. Brito, "Forecasting of Energy Consumption Artificial Intelligence Methods," *Iberian Conference on Information Systems and Technologies, CISTI*, vol. 2022-June, no. June, pp. 1–4, 2022, doi: 10.23919/CISTI54924.2022.9820078.

[161] R. Gaamouche, M. Chinnici, M. Lahby, Y. Abakarim, and A. E. E. Hasnaoui, *Machine Learning Techniques for Renewable Energy Forecasting: A Comprehensive Review*. Springer International Publishing, 2022. doi: 10.1007/978-3-030-96429-0_1.

[162] A. Moradzadeh, B. Mohammadi-Ivatloo, M. Abapour, A. Anvari-Moghaddam, and S. S. Roy, "Heating and Cooling Loads Forecasting for Residential Buildings Based on Hybrid Machine Learning Applications: A Comprehensive Review and Comparative Analysis," *IEEE Access*, vol. 10, pp. 2196–2215, 2022, doi: 10.1109/ACCESS.2021.3136091.

[163] L. F. J. Alvarez, S. R. Gonzalez, A. D. Lopez, D. A. H. Delgado, R. Espinosa, and S. Gutierrez, "Renewable energy prediction through machine learning algorithms," *2020 Ieee Andescon, Andescon 2020*, 2020, doi: 10.1109/ANDESCON50619.2020.9272029.

[164] A. Batish and A. Agrawal, "Building energy prediction for early-design-stage decision support: A review of data-driven techniques," *Building Simulation Conference Proceedings*, vol. 3, pp. 1514–1521, 2019, doi: 10.26868/25222708.2019.211032.

[165] C. Lu, S. Li, and Z. Lu, "Building energy prediction using artificial neural networks: A literature survey," *Energy Build*, vol. 262, p. 111718, 2022, doi: 10.1016/j.enbuild.2021.111718.

[166] J. Runge and R. Zmeureanu, "A review of deep learning techniques for forecasting energy use in buildings," *Energies (Basel)*, vol. 14, no. 3, 2021, doi: 10.3390/en14030608.

[167] M. L. Abdulrahman *et al.*, “A Review on Deep Learning with Focus on Deep Recurrent Neural Network for Electricity Forecasting in Residential Building,” *Procedia Comput Sci*, vol. 193, pp. 141–154, 2021, doi: 10.1016/j.procs.2021.10.014.

[168] S. Ardabili, L. Abdolalizadeh, C. Mako, B. Torok, and A. Mosavi, “Systematic Review of Deep Learning and Machine Learning for Building Energy,” *Front Energy Res*, vol. 10, no. March, pp. 1–19, 2022, doi: 10.3389/fenrg.2022.786027.

[169] R. Panchalingam and K. C. Chan, “A state-of-the-art review on artificial intelligence for Smart Buildings,” *Intelligent Buildings International*, vol. 13, no. 4, pp. 203–226, 2021, doi: 10.1080/17508975.2019.1613219.

[170] A. Alrobaie and M. Krarti, “A Review of Data-Driven Approaches for Measurement and Verification Analysis of Building Energy Retrofits,” *Energies (Basel)*, vol. 15, no. 21, 2022, doi: 10.3390/en15217824.

[171] T. D. Akinosho *et al.*, “Deep learning in the construction industry: A review of present status and future innovations,” *Journal of Building Engineering*, vol. 32, p. 101827, 2020, doi: 10.1016/j.jobe.2020.101827.

[172] I. Patsakos, E. Vrochidou, and G. A. Papakostas, “A Survey on Deep Learning for Building Load Forecasting,” *Math Probl Eng*, vol. 2022, 2022, doi: 10.1155/2022/1008491.

[173] D. Mazzei and R. Ramjattan, “Machine Learning for Industry 4.0: A Systematic Review Using Deep Learning-Based Topic Modelling,” *Sensors*, vol. 22, no. 22, 2022, doi: 10.3390/s22228641.

[174] Sveby, “Brukarindata bostäder (input data for residental buildings),” pp. 0–37, 2012.

[175] O. Pombo, K. Allacker, B. Rivela, and J. Neila, “Sustainability assessment of energy saving measures: A multi-criteria approach for residential buildings retrofitting - A case study of the Spanish housing stock,” *Energy Build*, vol. 116, pp. 384–394, Mar. 2016, doi: 10.1016/j.enbuild.2016.01.019.

Appendix A. ML algorithms mentioned in the literature

Table A. ML algorithms mentioned in the literature and their occurrence

Reference	ML Algorithms										Year
	ANN	**SVM**	**MLR**	**DT**	**RF**	**GP**	**DNN**	**XGBoost**	**ELM**	**CNN**	
[16]	√	√				√					2018
[143]	√	√		√							2018
[144]	√	√	√	√	√				√		2021
[145]	√	√			√					√	2022
[146]	√	√		√	√						2021
[147]	√	√			√				√		2022
[148]	√	√	√	√							2019
[149]	√	√	√								2020
[150]	√	√									2019
[151]	√										2019
[152]	√	√			√	√					2019
[153]					√						2020
[154]	√	√	√				√		√		2020
[155]	√	√		√		√	√			√	2023
[156]	√	√	√		√						2023
[157]	√	√	√		√			√			2023
[158]	√	√		√	√	√					2021
[159]	√	√	√	√	√	√		√			2022
[160]	√	√	√	√			√			√	2022
[161]	√	√									2022
[162]	√	√	√	√		√	√	√	√		2021
[163]		√	√		√						2020
[164]	√		√								2019
[165]	√										2022
[166]										√	2021
[167]										√	2021
[168]	√	√			√						2022
[169]	√			√							2021
[170]	√	√	√	√					√		2022
[171]							√			√	2020
[172]							√			√	2022
[173]				√	√		√			√	2022
Total	25	22	12	12	13	6	7	3	5	8	

Appendix B. Energy simulation setting

The settings for energy simulation are adapted from [174].

Table B. Energy plus setting for simulation

Energy system	Input parameter	Value and unit
Heating system	Heating setpoint for living area (apartment)	21 ℃
	Heating setpoint for the unoccupied area (circulation area)	10 ℃
Internal load	Occupancy	36 m^2/person
	Heating output	80 W/person
	Period	14 hours (0-8 o'clock, 18-24 o'clock)
	Household electricity per occupied area	3.46 W/m^2*h
	Radiance fraction	0.7
Water system	Hot water demand per occupied area	0.057 $L/h*m^2$
Ventilation	Ventilation + infiltration	0.0005 $m^3/s*m^2$

Appendix C. Energy simulation settings for other cities.

The settings for the energy simulation for Stockholm are adapted from [174]. The settings for energy simulation for Seattle, Chicago and Miami are adapted from U.S. Department of Energy (DOE). The settings for energy simulation for Madrid are adapted from [175].

Table C-1. Energy Plus Setting for Stockholm.

Energy system	Input parameter	Value and unit
Heating system	Heating setpoint for living area (apartment)	21 ℃
	Heating setpoint for the unoccupied area (circulation area)	10 ℃
Internal load	Occupancy	36 m^2/person
	Heating output	80 W/person
	Period	14 hours (0-8 o'clock, 18-24 o'clock)
	Household electricity per occupied area	3.46 W/m^2*h
	Radiance fraction	0.7
Ventilation	Ventilation + infiltration	0.0005 $m^3/s*m^2$

Table C-2. Energy Plus Setting for Seattle and Chicago.

Energy system	Input parameter	Value and unit
Heating system	Heating setpoint for living area (apartment)	20 ℃
	Heating setpoint for the unoccupied area (circulation area)	13 ℃
Cooling system	Cooling setpoint for living area (apartment)	26 ℃
	Cooling setpoint for the unoccupied area (circulation area)	28 ℃
Internal load	Occupancy	36 m^2/person
	Heating output	80 W/person
	Period	14 hours (0-8 o'clock, 18-24 o'clock)
	Household electricity per occupied area	3.46 W/m^2*h
	Radiance fraction	0.7
Ventilation	Ventilation rate per unit floor area	0.000294105 $m^3/s*m^2$
Infiltration	Flow per exterior surface area	0.00056957225 $m^3/s*m^2$

Table C-3. Energy Plus Setting for Madrid.

Energy system	Input parameter	Value and unit
Heating system	Heating setpoint for living area (apartment)	20 ℃ from 8 h to 23 h and 17 ℃ from 23 h to 7 h
	Heating setpoint for the unoccupied area (circulation area)	13 ℃
Internal load	Cooling setpoint for living area (apartment)	25 ℃ from 16 h to 23 h and 27 ℃ from 23 h to 7 h

	Cooling setpoint for the unoccupied area (circulation area)	28 ℃
	Occupancy	36 m^2/person
	Heating output	80 W/person
Ventilation	Ventilation rate	0.76 ac/h
Infiltration	Infiltration rate	0.3 ac/h

Table C-4. Energy Plus Setting for Miami.

Energy system	**Input parameter**	**Value and unit**
Cooling system	Cooling setpoint for living area (apartment)	26 ℃
	Cooling setpoint for the unoccupied area (circulation area)	28 ℃
Internal load	Occupancy	36 m^2/person
	Heating output	80 W/person
	Period	14 hours (0-8 o'clock, 18-24 o'clock)
	Household electricity per occupied area	3.46 W/m^2*h
	Radiance fraction	0.7
Water system	Hot water demand per occupied area	0.00000366 L/h*m^2
Ventilation	Ventilation rate per unit floor area	0.000294105 m^3/s*m^2
Infiltration	Flow per exterior surface area	0.00056957225 m^3/s*m^2

Appendix D. Accuracy of TL models and retrained ANN models under different training dataset sizes in predicting energy demand for four cities.

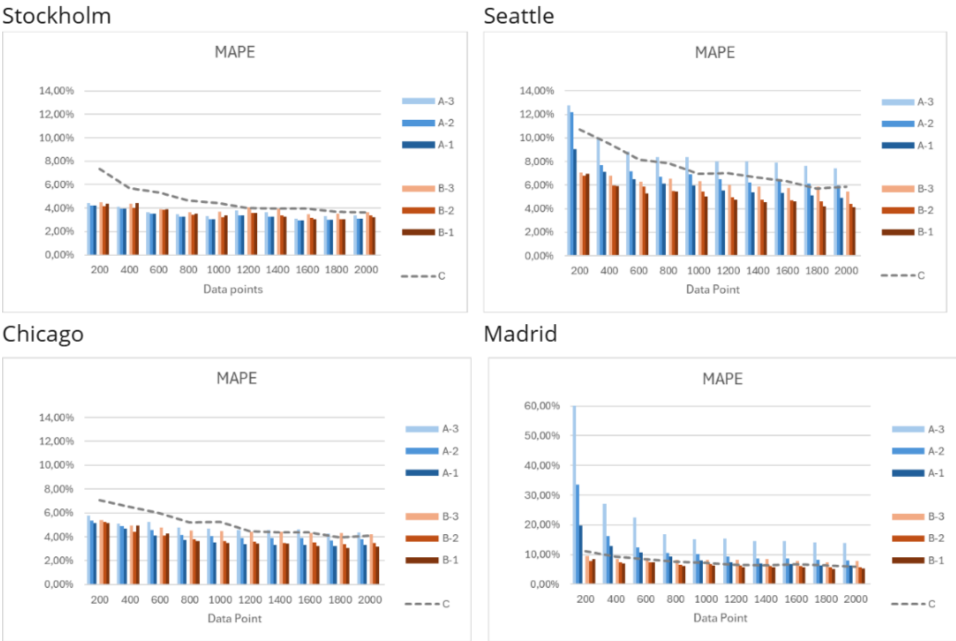


Figure D-1. MAPE of TL models and retrained ANN models under different training dataset sizes in predicting heating energy demand for four cities.

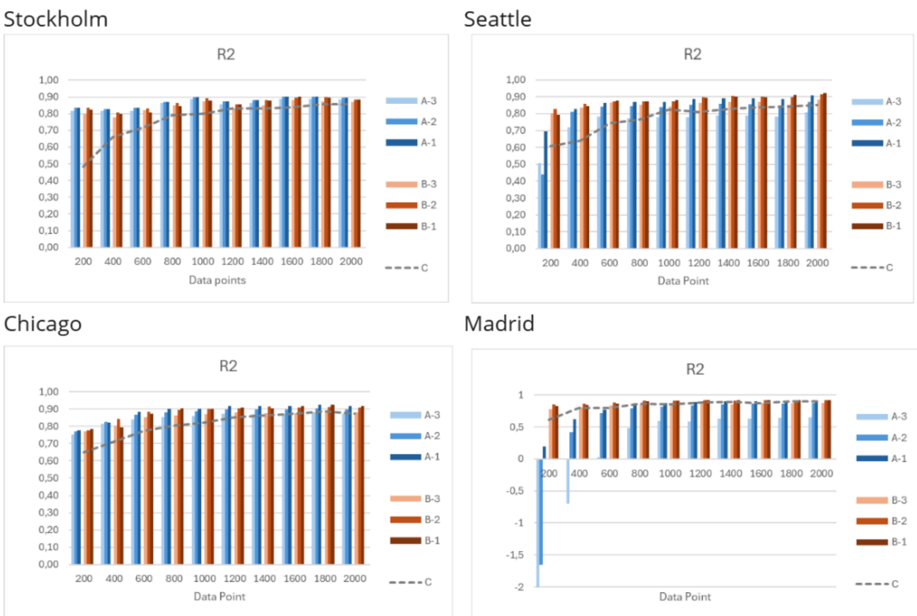


Figure D-2. R^2 of TL models and retrained ANN models under different training dataset sizes in predicting heating energy demand for four cities.

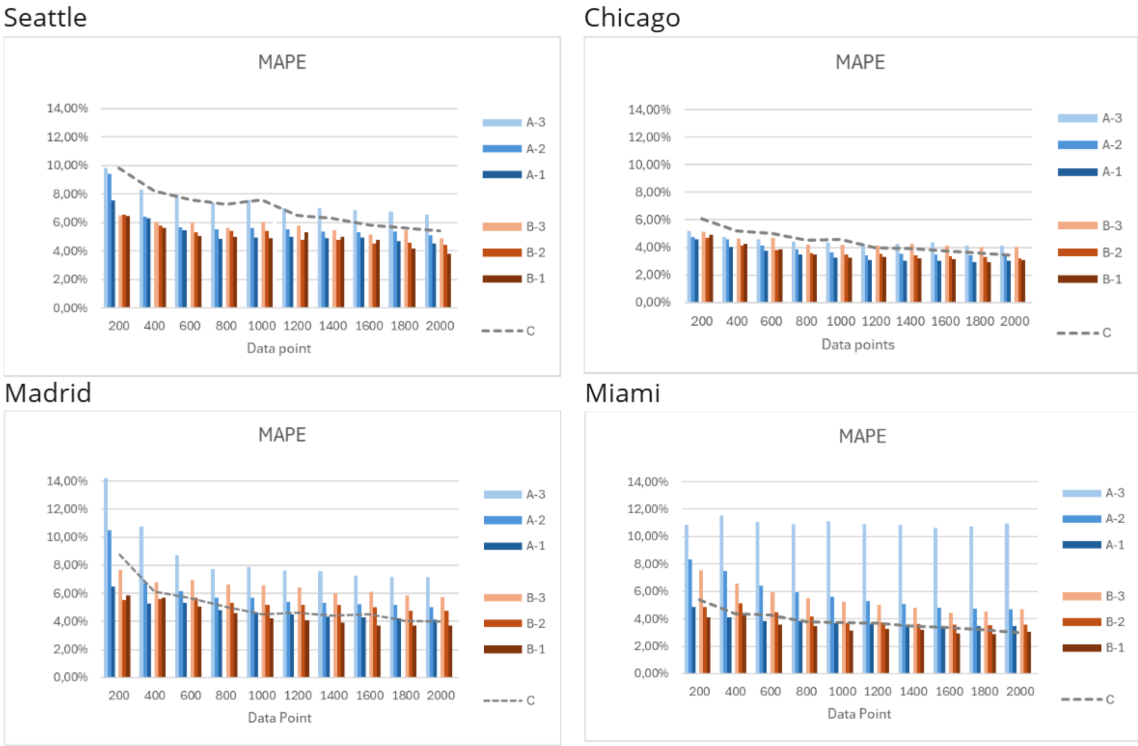


Figure D-3. MAPE of TL models and retrained ANN models under different training dataset sizes in predicting heating energy demand for four cities.

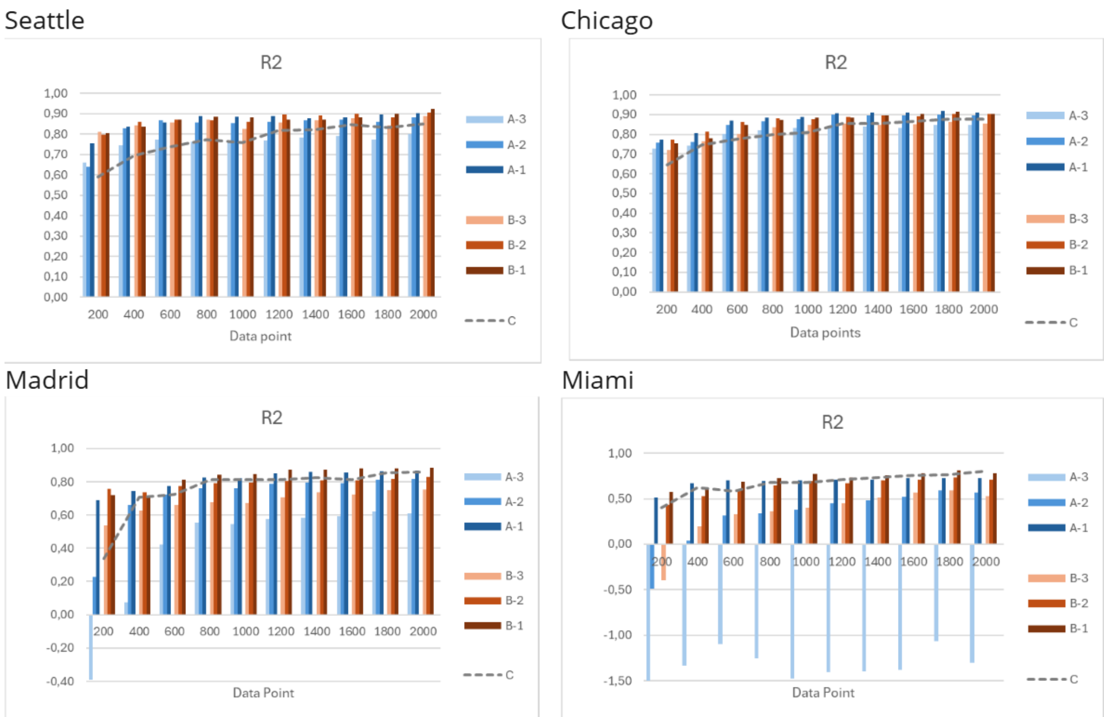


Figure D-4. R^2 of TL models and retrained ANN models under different training dataset sizes in predicting heating energy demand for four cities.

Appendix E. Influential ADVs for building energy optimization under different climates from the literature review.

Figure E further explores this by showing the percentage of the mentioned time for each ADV regarding various climate zones. The results show that most ADVs, especially the most influential ones, such as WWR and building plan, do not vary across different climate contexts. Therefore, it is reasonable to include the literature beyond the Nordics when identifying influential ADVs. However, certain ADVs do vary across geographical regions. For instance, building orientation is more influential in desert continental climate than the rest climate zones. Future research could investigate how and why certain influential ADVs differ across geographical regions. In the same vein, this study is also limited to the influential ADVs for residential buildings, thus a similar study should be conducted to identify influential ADVs for other building types such as commercial and industrial buildings.

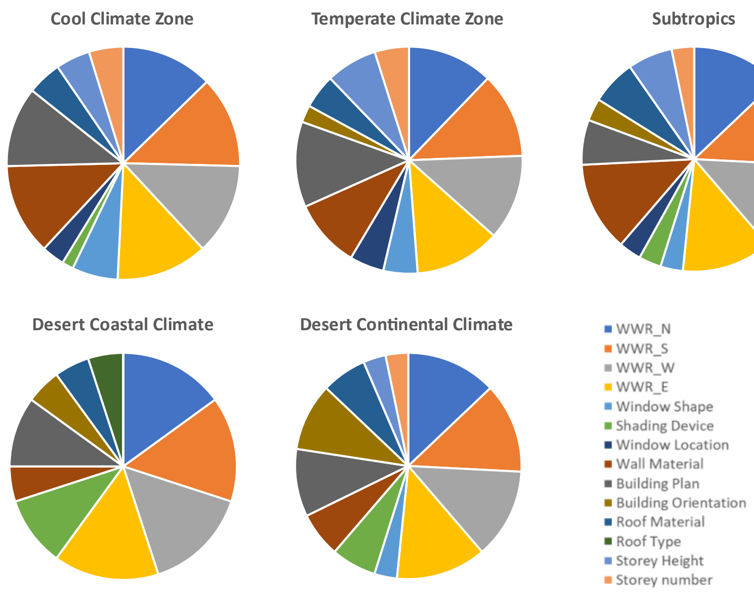


Figure E. Influential ADVs for building energy optimization under different climates from the literature review.

Appendix F. Average rating with all 24 respondents and random 20 respondents.

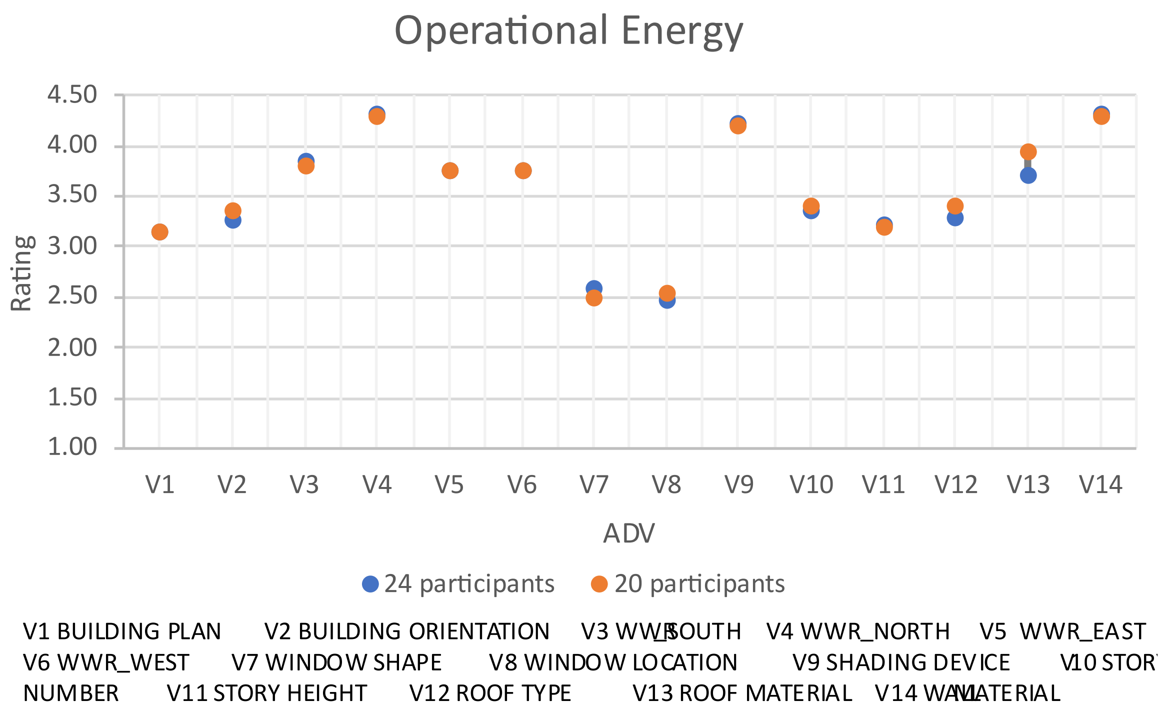


Figure F. Average rating with all 24 respondents and random 20 respondents.

Appendix G. Discrepancies in prediction accuracy compared with actual simulation results.

The final test case in Study D differs in spatial and geometric configuration from the building plans represented in the model's training dataset in Study B. As a result, the model's predictive accuracy decreases, and the gap between predicted and simulated energy values becomes more apparent. Nevertheless, as shown in Figure G, although the predicted values from the ML model do not perfectly match the actual simulation results, the overall trend of the optimal energy consumption across generations still shows a clear downward trajectory. Even though occasional increases occur due to prediction errors, the general direction of performance improvement aligns closely with that of the simulation-based optimization. This consistency indicates that the ML-based approach is capable of effectively guiding the search toward better-performing design solutions and can, to a large extent, serve as a surrogate for simulation in the context of energy optimization. To further improve the predictive accuracy of the ML model, it would be beneficial to expand and diversify the training dataset. A larger and more representative database would enhance the model's generalization ability and reduce discrepancies between predicted and simulated results in future applications.

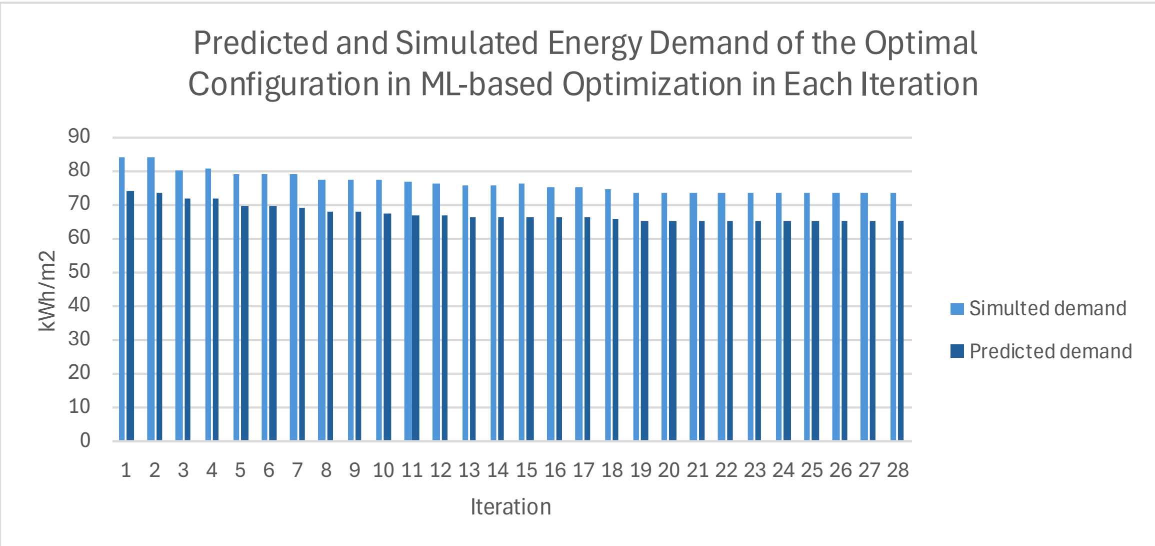


Figure G. The predicted and simulated energy demand of the optimal configuration in ML-based optimization in each iteration.