ORTA DOĞU TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜ

TEZ ŞABLONU ONAY FORMU
THESIS TEMPLATE CONFIRMATION FORM

1. Şablonda verilen yerleşim ve boşluklar değiştirilmemelidir.
2. **Jüri tarihi** Başlık Sayfası, İmza Sayfası, Abstract ve Öz'de ilgili yerlere yazılmalıdır.
3. İmza sayfasında jüri üyelerinin unvanları doğru olarak yazılmalıdır. Tüm imzalar **mavi pilot kalemle** atılmalıdır.
4. **Disiplinlerarası** programlarda görevlendirilen öğretim üyeleri için jüri üyeleri kısmında tam zamanlı olarak çalıştıkları anabilim dalı başkanlığının ismi yazılmalıdır. Örneğin: bir öğretim üyesi Biyoteknoloji programında görev yapıyor ve biyoloji bölümünde tam zamanlı çalışıyorsa, İmza sayfasına biyoloji bölümü yazılmalıdır. İstisnai olarak, disiplinler arası program başkanı ve tez danışmanı için disiplinlerarası program adı yazılmalıdır.
5. Tezin **son sayfasının sayfa** numarası Abstract ve Öz'de ilgili yerlere yazılmalıdır.
6. Bütün chapterlar, referanslar, ekler ve CV sağ sayfada başlamalıdır. Bunun için **kesmeler** kullanılmıştır. **Kesmelerin kayması** fazladan boş sayfaların oluşmasına sebep olabilir. Bu gibi durumlarda paragraf (¶) işaretine tıklayarak kesmeleri görünür hale getirin ve yerlerini **kontrol edin**.
7. Figürler ve tablolar kenar boşluklarına taşmamalıdır.
8. Şablonda yorum olarak eklenen uyarılar dikkatle okunmalı ve uygulanmalıdır.
9. Tez yazdırılmadan önce PDF olarak kaydedilmelidir. Şablonda yorum olarak eklenen uyarılar PDF dokümanında yer almamalıdır.
10. Tez taslaklarının kontrol işlemleri tamamlandığında, bu durum öğrencilere METU uzantılı öğrenci e-posta adresleri aracılığıyla duyurulacaktır.
11. Tez yazım süreci ile ilgili herhangi bir sıkıntı yaşarsanız, **Sıkça Sorulan Sorular (SSS)** sayfamızı ziyaret ederek yaşadığınız sıkıntıyla ilgili bir çözüm bulabilirsiniz.

1. Do not change the spacing and placement in the template.
2. Write **defense date** to the related places given on Title page, Approval page, Abstract and Öz.
3. Write the titles of the examining committee members correctly on Approval Page. **Blue ink** must be used for all signatures.
4. For faculty members working in **interdisciplinary programs**, the name of the department that they work full-time should be written on the Approval page. For example, if a faculty member staffs in the biotechnology program and works full-time in the biology department, the department of biology should be written on the approval page. Exceptionally, for the interdisciplinary program chair and your thesis supervisor, the interdisciplinary program name should be written.
5. Write **the page number of the last page** in the related places given on Abstract and Öz pages.
6. All chapters, references, appendices and CV must be started on the right page. **Section Breaks** were used for this. **Change in the placement** of section breaks can result in extra blank pages. In such cases, make the section breaks visible by clicking paragraph (¶) mark and **check their position.**
7. All figures and tables must be given inside the page. Nothing must appear in the margins.
8. All the warnings given on the comments section through the thesis template must be read and applied.
9. Save your thesis as pdf and Disable all the comments before taking the printout.
10. This will be announced to the students via their METU students e-mail addresses when the control of the thesis drafts has been completed.
11. If you have any problems with the thesis writing process, you may visit our **Frequently Asked Questions (FAQ)** page and find a solution to your problem.

☒ **Yukarıda bulunan tüm maddeleri okudum, anladım ve kabul ediyorum. / I have read, understand and accept all of the items above.**

Name : İlkim
Surname : Canlı
E-Mail : iilkimcanli@gmail.com
Date :
Signature : ________________________

URBAN SCALE PREDICTION OF INDOOR DAYLIGHTING ILLUMINATION FOR SUSTAINABLE BUILDINGS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

İLKİM CANLI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF ARCHITECTURE
IN
ARCHITECTURE

DECEMBER 2022

Approval of the thesis:

URBAN SCALE PREDICTION OF INDOOR DAYLIGHTING ILLUMINATION FOR SUSTAINABLE BUILDINGS

submitted by **İLKİM CANLI** in partial fulfillment of the requirements for the degree of **Master of Architecture in Architecture, Middle East Technical University** by,

Prof. Dr. Halil Kalıpçılar
Dean, Graduate School of **Natural and Applied Sciences** ________________

Prof. Dr. Cânâ Bilsel
Head of the Department, **Architecture** ________________

Assoc. Prof. Dr. İpek Gürsel Dino
Supervisor, **Architecture, METU** ________________

Assoc. Prof. Dr. Sinan Kalkan
Co-Supervisor, **Computer Engineering, METU** ________________

Examining Committee Members:

Assist. Prof. Dr. Mehmet Koray Pekeriçli
Architecture, METU ________________

Assoc. Prof. Dr. İpek Gürsel Dino
Architecture, METU ________________

Assoc. Prof. Dr. Yasemin Afacan
Interior Architecture and Environmental Design, Bilkent University ________________

Date: 26.12.2022

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name Last name : İlkim Canlı

Signature :

ABSTRACT

URBAN SCALE PREDICTION OF INDOOR DAYLIGHTING ILLUMINATION FOR SUSTAINABLE BUILDINGS

Canlı, İlkim
Master of Architecture, Architecture
Supervisor : Assoc. Prof. Dr. İpek Gürsel Dino
Co-Supervisor: Assoc. Prof. Dr. Sinan Kalkan

December 2022, 134 pages

Daylight illumination has been an essential consideration during design for architects throughout history. Daylight is a crucial design component for long-term sustainability that influences the visual and thermal comfort of the occupants and energy usage in buildings. Utilizing daylighting effectively reduces the energy required for artificial lighting and the indoor thermal loads of spaces. However, dense urban areas prevent daylight from reaching buildings. Each surrounding building acts as a shadow element obstructing the building's access to natural light. Therefore, analyzing daylight illuminances and understanding the building design characteristics and urban form parameters that affect daylight illuminance is unavoidable for sustainable building design. Simulations are one of the most preferred tools to analyze the level of illuminance in building designers. Simulations require detailed modeling knowledge and expertise to get precise results. Also, daylighting simulations performed at an urban scale take much computational time.

In contrast, machine learning (ML) models enable designers to analyze daylighting levels with less computational time and detailed knowledge. This study aims to develop a method to predict hourly indoor daylighting illuminances in an urban context using ML models. For the development of the method, three different ML models (multi-layer perceptrons, random forest, extreme gradient boosting) were developed, and their performance results were compared. The ML model with the highest performance accuracy was selected as the final model. The developed method helps designers/ architects to analyze hourly indoor daylight illuminances in an urban context. The developed methodology also calculates how much daylight-dependent electric lighting is used in buildings by analyzing hourly indoor daylighting illuminances on an urban scale. The proposed methodology enables the integration of indoor daylighting analysis with the electric lighting energy consumption calculation based on real-time estimation of daylight illuminances. Residential units in the Bahçelievler neighborhood in Ankara were simulated using various design factors, and the simulation results were utilized for training and evaluating the machine learning models. The proposed model can enhance the usage of machine learning in architectural design stages to analyze daylight illuminances, accordingly, forecast the artificial electric load in buildings, and help designers integrate daylight into the buildings.

Keywords: Daylighting, Daylighting Illuminance Prediction, Machine Learning Models, Urban Scale Daylighting Analysis

ÖZ

SÜRDÜRÜLEBİLİR BİNALAR İÇİN İÇ MEKAN GÜNIŞIĞI AYDINLANMASININ KENT ÖLÇEKLİ TAHMİNİ

Canlı, İlkim
Yüksek Lisans, Mimarlık
Tez Yöneticisi: Doç. Dr. İpek Gürsel Dino
Ortak Tez Yöneticisi: Doç. Dr. Sinan Kalkan

Aralık 2022, 134 sayfa

Tarih boyunca, gün ışığının tasarımlarla entegrasyonunu artırmak mimarlar için her zaman önemli bir konu olmuştur. Gün ışığı, kullanıcının ruh halini, bina sakinlerinin görsel ve termal konforunu ve binalarda enerji kullanımını etkileyen uzun vadeli sürdürülebilirlik için önemli bir tasarım bileşenidir. Gün ışığından faydalanmak, yapay aydınlatma için gereken enerjiyi ve yapılar üzerindeki iç yükü etkili bir şekilde azaltır. Ancak yoğun kentsel alanlar, gün ışığının binalara ulaşmasını engellemektedir. Çevredeki her bina, binanın doğal ışığa erişimini engelleyen bir gölge unsuru görevi görür. Bu nedenle, sürdürülebilir bina tasarımı için gün ışığı aydınlanmalarının analiz edilmesi ve gün ışığı aydınlanmasını etkileyen bina tasarım özelliklerinin ve kentsel form parametrelerinin anlaşılması kaçınılmazdır. Simülasyonlar, bina tasarımcılarının aydınlanma düzeyini analiz etmek için en çok tercih ettiği araçlardan biridir. Simülasyonlar, detaylı ve doğru sonuçlar elde etmek için ayrıntılı modelleme bilgisi ve uzmanlığı gerektirir. Ayrıca, kentsel ölçekte gerçekleştirilen günışığı simülasyonları çok fazla hesaplama zamanı almaktadır.

Buna karşılık, makine öğrenimi modelleri, tasarımcıların daha az hesaplama süresi ile ayrıntılı bilgi gerektirmeden günışığı seviyelerini analiz etmelerini sağlar. Bu çalışma, makine öğrenmesi (ML) modellerini kullanarak kentsel bağlamda saatlik iç mekan günışığı aydınlanmalarını tahmin etmek için bir araç geliştirmeyi amaçlamaktadır. Aracın geliştirilmesi için üç farklı ML modeli (MLP, RF, XGBoost) geliştirilmiş ve performans sonuçları karşılaştırılmıştır. Performans doğruluğu en yüksek olan makine öğrenimi modeli nihai model olarak seçilmiştir. Geliştirilen araç, tasarımcıların/mimarların kentsel bağlamda saatlik iç mekan gün ışığı aydınlatmalarını analiz etmelerine yardımcı olacaktır. Geliştirilen araç aynı zamanda kentsel ölçekte saatlik iç mekan günışığı aydınlanmalarını analiz ederek binalarda gün ışığına bağlı olarak ne kadar elektrik aydınlatmasının kullanıldığını da hesaplamaktadır. Önerilen metodoloji, gün ışığı aydınlanmalarının gerçek zamanlı tahminine dayalı elektrik aydınlatma enerji tüketimi hesaplaması ile iç mekan günışığı analizinin entegrasyonunu sağlar. Ankara Bahçelievler mahallesindeki konut birimleri çeşitli tasarım parametreleri kullanılarak simüle edilmiş ve simülasyon sonuçlarından makine öğrenmesi modellerinin eğitimi ve değerlendirilmesi için yararlanılmıştır. Önerilen model, gün ışığı aydınlanmalarını analiz etmek, buna bağlı binalarda aydınlatma için kullanılan elektrik tüketimini tahmin etmek ve tasarımcıların gün ışığını binalara entegre etmelerine yardımcı olmak için mimari tasarım aşamalarında makine öğreniminin kullanımını artıracaktır.

Anahtar Kelimeler: Günışığı, Günışığı Aydınlanma Tahminlemesi, Makine Öğrenmesi Modelleri, Kentsel Ölçekte Günışığı Analizi

Felicitâ…

ACKNOWLEDGMENTS

Firstly, I would like to express my deepest gratitude to my supervisors, Assoc. Prof. Dr. İpek Gürsel Dino and Assoc. Prof. Dr. Sinan Kalkan. In addition to my undergraduate education, the vision that Prof. Gursel Dino provided to my academic and personal life during my graduate education gave me very different perspectives on my life. She supported me throughout this process by encouraging me to think differently and explore different paths throughout my academic life. It has been a great pleasure and privilege for me to work with her. I would like to thank Assoc. Prof. Dr. Sinan Kalkan as my co-advisor for expanding my vision in interdisciplinary fields and contributing to my work with his valuable contributions in this process.

I would like to thank the examining committee members, Assist. Prof. Dr. Mehmet Koray Pekeriçli and Assoc. Prof. Dr. Yasemin Afacan for accepting to join the committee and providing invaluable feedback to my study.

I would like to thank my dear friends, Aslı Zeynep Doğan and Mehmet Oğuz Nas, for always being there for me when I needed them psychologically. In particular, my soul-sister Ezgi Cemre Ural, I am very grateful to you for being there for me day and night whenever I needed her and always inspiring me.

Finally, my biggest support in this process is my dear family. My father Serdar, my mother Nilüfer, my brother Burak, and my daughter Helen, I could not get through this process without you. Your loving and constant support is my biggest help in completing this process. I would also like to tell my dear uncle, Gökhan, how lucky I am to always be by my side throughout this process. I dedicated this thesis to my mother.

This work is partially funded by the Scientific and Technological Research Council of Turkey under grant number TUBİTAK 120M997.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

ABBREVIATIONS	
ANN	Artificial Neural Network
ASE	Annual Sunlight Exposure
CDA	Continous Daylight Autonomy
DA	Daylight Autonomy
DF	Daylight Factor
DT	Decision Tree
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multi-layer Perceptron
MSE	Mean Squared Error
R^2	Coefficient of Determination
RF	Random Forest
RMSE	Root Mean Squared Error
SDA	Spatial Daylight Autonomy
SE	Sky Exposure
SVM	Support Vector Machine
VT	Visible Transmittance
WWR	Window-to-Wall Ratio
XGB	eXtreme Gradient Boosting

CHAPTER 1

INTRODUCTION

Throughout the history of architecture, daylight has been accepted as an essential element of design. Ancient civilizations like the Persians, Arabs, Greeks, and Romans designed their homes around courtyards that welcomed natural light and established solar zoning laws that gave people access to the sun inside their homes (Boubekri, 2014). In ancient times, daylight was considered a decorative and aesthetic design element that was integrated into spaces in different ways to reflect the spirit of the space. Today, daylight is no longer regarded solely as a decorative element but also as a critical concern for long-term sustainability that influences the mood and behavior of humans (Webb, 2006), the visual and thermal comfort of occupants, and energy usage in buildings (D. H. W. Li et al., n.d.). Daylight regulates the circadian cycle of hormone secretions and body temperature, affecting sleep/wake states, alertness, mood, and behavior (Webb, 2006). Circadian rhythm is significantly impacted by daylight, which dominates human psychology and behavior. Moreover, vision and glare affect the occupants' visual comfort and are closely related to daylight. A good daylight design aims to provide enough light for successful visual performance while ensuring an appropriate level of comfort; therefore, both the visual and non-visual aspects of daylight should be considered (Wienold & Christoffersen, 2006). Daylight has significant effects on the building as well as its effects on the occupants of the building. Buildings use nearly 40% of the energy used in the world, and a significant amount of that energy is used for lighting (Kaminska & O˙ Zadowicz, 2018). Energy used for lighting accounts for approximately 19% of global electrical energy consumption (Enkvist et al., 2010; Papinutto et al., 2022). Utilizing daylight reduces the energy required for artificial lighting and building interior load. Reported savings in lighting energy consumption obtained by increasing the use of daylight can start from 15% and go up to 80% (Waide & Tanishima, 2006). Therefore, daylight should be considered a passive design strategy to maintain occupants' physical and psychological well-being, enabling visual comfort and minimizing artificial electric lighting consumption and buildings.

However, dense urban areas stemming from rapid urbanization are a crucial barrier to free daylight access to buildings. Each building acts as context shading and leads to problems, including reflecting daylight or blocking the daylight, which requires different solutions. The reflectivity of surrounding buildings results in visual discomfort for other buildings. Conversely, daylight access can be completely blocked due to the surrounding buildings. As a result, building characteristics and urban form significantly influence how daylight penetrates the buildings. Therefore, the daylight illuminances of buildings should be analyzed by considering not only building-related parameters but also urban form-related parameters.

The impacts of design elements for efficient daylighting, such as building orientation, climate, window-to-wall ratio (WWR), glazing type, and fixed outside shade, have been comprehensively investigated in several studies over the past few decades. Several methods are used to analyze daylighting from past to present; the most recent are simulation and machine learning. Computer simulation tools enable designers to comprehensively analyze daylight access to the buildings in hourly/annual resolution. However, simulation tools are limited because of the excessive computational time of urban modeling to analyze daylight and its necessity for detailed modeling knowledge by users (Nault et al., 2017). Also, simulations require a complex set of inputs such as weather data, building, and urban form parameters (Ayoub, 2019a). The resulting computational cost and dependence on usage knowledge limit the interactivity between the tool results and the design process (Beckers & Rodríguez, 2009).

Machine learning (ML) models provide comprehensive daylight analysis by reducing computational costs without requiring detailed usage knowledge (Ayoub, 2020). The models learn the relationship between inputs and output based on the training dataset. Once the model is trained with the data, it can be adapted to different buildings when parameters are indicated. Although ML applications to daylighting estimation have increased by considering different input and output parameters in recent years, daylighting studies focus on working environments such as offices (Dogan & Park, 2019). Daylight access in residential buildings on an urban scale needs to be analyzed in detail.

Despite the increasing ML applications in daylighting studies, fewer studies focused on the real-time prediction of daylight performance metrics in an urban context. Generally, case studies focused on a singular building, excluded from its urban context. Ignoring urban parameters leads daylighting to being calculated in a biased way. At the urban scale, each element around the building can affect the building in different ways. It can reflect the daylight to the building or block the daylight reaching the building to a certain extent or completely. Therefore, ignoring the impact of urban elements on the daylighting penetration and access of buildings leads to unrealistic calculations.

Moreover, it is important to make real-time predictions by considering the urban form and building parameters to observe the daylight changes in hourly resolution. Using machine learning models, different daylighting performance metrics are estimated in daylighting prediction studies. These performance metrics can provide specific information, such as how much the space meets certain illuminance levels annually or what percentage of an area remains above a certain illuminance level. However, the real-time estimation of daylighting prediction enables users to analyze how much the space is illuminated in hourly resolution. This information can be used in different areas directly related to daylighting, such as visual comfort and electricity consumption. The illuminance results predicted by ML models can allow the estimation of electricity consumption based on daylight. On the other hand, there is limited information about how predicted daylighting illuminance results can be utilized to forecast electric lighting consumption. The real-time estimation of daylight illuminance considering the urban morphology and integrating this information to forecast the electric usage for artificial lighting provides a comprehensive analysis of daylight usage and its effect on electric lighting usage in the architectural design stages.

1.1 Problem Statement

Daylight is a free source that aids sustainable design by improving humans' psychological and physical health, the visual and thermal comfort of occupants, and decreasing electrical lighting energy usage in buildings. Integrating daylight with buildings is one of the most feasible passive design strategies supporting the sustainable design concept. However, the surrounding buildings block daylight access to buildings on an urban scale. Daylighting illumination can be quantified utilizing daylighting simulation. However, simulations require the development of 3D models in which semantic data related to daylighting illuminance is captured. However, these model-based approaches involve a high cost of model development and a high computational cost of simulations. As a result, their use in design exploration, during which immediate feedback to designers is critical, remains limited. This also restricts the number of alternatives that can be explored during design. Therefore, there is a need for a quick analysis of indoor daylighting illuminances in an urban context for performance-driven decision-making. The analysis requires a detailed understanding of urban form parameters and their influences on indoor daylighting illuminances. Identifying the gaps in the literature, this thesis will focus on the following research problems:

- Factors affecting daylighting illuminance at the urban scale should be examined in detail, and estimation should be made using these parameters in ML models. However, most of the studies examined in the literature considered the spaces as a single unit, ignoring the urban form parameters. This leads to overestimating the estimated indoor daylighting illuminance values, resulting in miscalculations. ML models should be trained by considering urban form, building, and climate parameters.

- Indoor daylighting illuminance values analyzed on an urban scale enable different analyzes based on daylight. If the amount of daylighting in the area can be precisely calculated, visual comfort, glare, and electric lighting assessments can be made. Although ML models have attempted to forecast values for daylighting illuminance in the literature, the use of the model's data in other domains has not been extensively investigated. Uses of the results of ML models related to daylighting should be explored.
- In the literature, there are studies in which the energy used for electrical lighting in buildings is estimated with ML models using simulation data of energy models or, if available, ready consumption data. However, calculating electrical lighting consumption based on predicted indoor daylighting illuminances on an urban scale has not been studied thoroughly yet. The impact of daylighting integration on electricity consumption in buildings can be studied by calculating the electric lighting consumption based on daylighting estimation at the urban scale.
- Instead of a single ML model, different data-driven methods (ML) that support performance-based design decisions should be explored for quick analysis of indoor daylighting illuminances in an urban context. Multiple ML models to predict indoor daylighting illuminances in an urban context with high accuracy should be discovered, and the performance results of the models should be compared with analysis. Providing immediate feedback and ease of use, the prediction models have the potential to be used by architects, urban planners, and companies.

1.2 Research Objectives

This thesis aims to investigate the potential of machine learning to predict indoor daylighting illuminance at the urban scale. A comprehensive literature review on the state-of-the-art factors that affect indoor daylighting performance, daylighting performance metrics and calculation methods, and electric lighting energy consumption calculations were conducted to achieve this aim. The potential and limitations of the studies were analyzed. Based on the review, different machine-learning models were developed with analyzed parameters. The developed ML models' performances were compared, and the model that performed the best overall was selected as the final model. The developed model can be utilized to perform daylighting analysis at an hourly resolution on an urban scale at different stages of the design and to make design decisions based on the analysis.

After the methodology was developed, the potential areas that the method could be utilized were explored. The developed methodology can also calculate lighting electricity consumption based on daylight. This exploration aims to comprehensively analyze the real-time estimation of indoor daylighting illuminances at the urban scale and artificial lighting electric consumption based on daylighting illuminances.

The method was developed based on simulation data. The illuminance results of different buildings in Bahçelievler, Ankara, were obtained from simulations, and different ML models were trained based on the simulation data.

To conclude, the aims of this thesis are;

- Exploring design parameters that influence indoor daylighting illuminances at the urban scale
- Exploring possible different ML models to predict indoor daylighting illuminances at the urban scale in hourly resolution
- Developing a design method that can support decision-making in the architectural design stages based on daylighting illuminances using ML models
- Integrating the results of the prediction model to forecast the electric lighting consumption in buildings based on daylighting and make a comprehensive analysis between the electric lighting usage and daylighting

1.3 Research Questions

The primary research question and its supporting questions arose as follows in light of the research gap described in the introduction:

Main thesis questions:

- To which extent can data-driven (ML) methods accurately calculate the indoor daylight illuminances and associated energy use at the urban scale considering the effects of surrounding buildings?

Sub-Questions:

- Which building and urban-related input features should be used to perform the urban scale analysis in the machine learning model?
- What ML models have the best predictive capacity in predicting indoor daylighting illuminances?
- How can ML models be used to calculate electrical lighting energy consumption by analyzing daylighting performance on an urban scale?
- How does the electrical lighting energy consumption calculated based on daylighting that is predicted with ML models differ when compared to the electrical lighting energy consumption calculated with the standard schedules used in Urban Building Energy Modeling (UBEM)?

1.4 Thesis Outline

The thesis is composed of five main chapters, including the present chapter. In order to reach the objectives of this thesis, the research work went through the following chapters of the study:

Chapter 1 (Introduction)

The current chapter is called an 'Introduction' in which the thesis' motivation, aims and objectives, research questions, and contribution to the related field are explained.

Chapter 2 (Related Works & Background):

The literature will be reviewed to examine the relationship between architecture, sustainability, and daylight, factors that affect daylighting illuminances, daylighting performance metrics, calculation methods for daylighting performance metrics, and calculation of electric lighting in buildings. The objectives of this thesis are stated by identifying the knowledge gaps in the literature, and the need for the proposed methodology is highlighted.

Chapter 3 (Methodology):

The methodology of the proposed design methodology to achieve the objectives of the thesis is explained in this chapter. A method based on ML models for estimating indoor daylighting illuminances in an urban context is developed to address the need for data-driven tools that assist quick design decisions in architectural design stages. For this purpose, several buildings are modeled and simulated for dataset generation. The results of the simulations are used to train and validate the ML model. The model is evaluated based on different performance metrics. Three different ML model is trained. The model that showed the best performance result was chosen as the final model. After the development, other possible applications of the method are explored. The developed method also calculates electric lighting usage in buildings on an urban scale based on daylight.

Chapter 4 (Results):

The accuracy of the ML models in predicting indoor daylighting illuminances in an urban context is reported in this chapter. The performances of the models are compared in terms of performance metrics. Based on the developed methodology, electric lighting consumption results of different cases will be analyzed in this chapter.

Chapter 5 (Conclusion):

The study's contributions and the proposed method's applicability and limitations will be examined in this chapter.

CHAPTER 2

LITERATURE REVIEW

This chapter examines the existing literature to achieve the thesis aims. The relationship between daylighting and architecture, the visual and non-visual effects of daylighting, the relationship between daylighting and energy, factors that affect daylight illuminances, daylighting performance metrics, daylighting calculation methods, and electric lighting energy consumption calculations are examined based on the review. Figure 2.1 shows the concepts reviewed in Chapter 2. At the end of the chapter, the identified gaps in the studies are given.

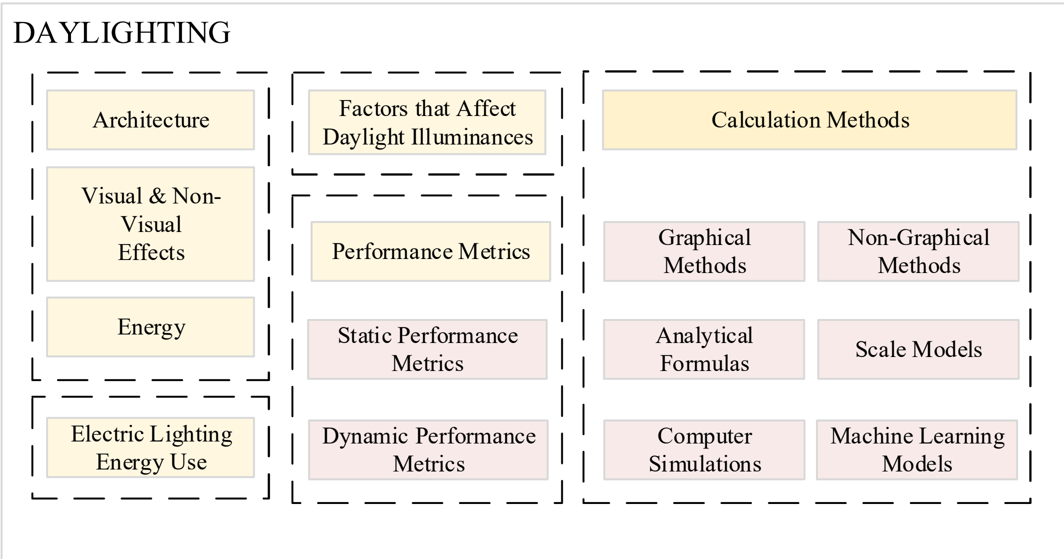


Figure 2.1. The concepts reviewed in Chapter 2

2.1 Daylighting and Architecture

Augmenting natural light in architectural designs has been an important topic for architects throughout history. It does more than give us the physical ability to see; it also gives architecture the primary energy component required for the coexistence of an integrated dualism of matter and energy, giving users a sense of aesthetic pleasure (Serra, 1998). In the architecture of ancient civilizations such as the Persians, the Arabs, the Greeks, and the Romans, courtyards were designed to penetrate daylight (Boubekri, 2014). Moreover, in some civilizations, daylight is considered a right and necessity (Turan et al., 2020). Protecting access to daylight was reflected in urban zoning laws in all major cities. The right to light of building residents is protected in the United States by legal action and legislation in some states (Pfeiffer, 1982; Davis, 1989). Many cities' urban forms today result from zoning regulations implemented throughout the 20th century to preserve individual and public rights. More recently, New York City's zoning laws specified limitations regarding buildings' exterior design to reduce shade (Willis, 1995).

While the use of natural light in city planning is pritiorized in different ways, the use of daylight at the building scale is also handled from different perspectives. Historically, daylight usage was one of the major design considerations, as artificial lighting is expensive and difficult to obtain (Bainbridge & Haggard, 2011). Natural light is a deciding design element for determining the quality of a place, whether it comes from a single entrance like the Pantheon or more intricate designs like those found in baroque, cathedral, church, and mosque architecture. Especially daylight was integrated into the buildings to reflect the holistic spirit of the sacred places.

Peter Zumthor stated that one of the nine basic elements that determine the atmosphere of space is light in his famous book ‘Atmosphere’(Zumthor, 2006). Le Corbusier carefully constructed daylight as a design element, especially in the two most well-known buildings, ‘The Chapel at Ronchamp’ and ‘Paris Church of Saint-Pierre in Firminy’. He prioritized the use of daylight not only based on buildings but also on urban design ideas. Le Corbusier also designs city master plans (La Ville

Radieuse) to increase green spaces and access to daylight. ‘Kimbell Art Museum’ is one of the most important works of Louis Kahn to merge structure with daylight. Tadao Ando is another important architect for designing light in spaces. 'Church of Light,' by Tadao Ando, merges natural light with architecture to reflect the holy spirit of the church. Although sometimes daylight has been integrated into the building employing small openings, over time, it has been integrated with the space with large openings and even almost full glass facades. One of the most famous examples of this is Glass House by Philip Johnson.

Today, daylight is no longer regarded solely as a decorative element but as a critical concern for long-term sustainability that influences the mood and behavior of humans (Webb, 2006), the visual and thermal comfort of occupants, and energy usage in buildings (D. H. W. Li et al., n.d.). Figure 2.2 shows the integration of daylighting in different types of buildings (residential, office, and public).

(a): De Syllas House, London

(b): Passmores Academy, United Kingdom (Tim Crocker)

(c): Alexander Court, Washington (Fox Architects)

Figure 2.2. Daylighting integration in a different types of buildings

Therefore, architects and designers aim to provide good daylighting in buildings. However, what good daylighting is a debatable topic. The various professions tackled this question from different perspectives. According to Reinhart & Galasiu's (Reinhart et al., 2006) survey, there are five sample definitions for good daylighting. The architecture profession describes natural light as a tool that interacts with building form to create an interior environment that is aesthetically interesting, healthy, and productive. From lighting savings design, good daylighting means replacing indoor electric illumination needs with daylight, which decreases annual energy consumption for lighting. Regarding building energy consumption, good daylighting includes fenestration systems and responsive electric lighting control devices leading to reduced building energy requirements, including lighting, heating, and cooling. Load management defines daylighting as dynamic control of fenestration and lighting to monitor and control peak electric load. Finally, from a cost perspective, daylight means strategies to minimize operating costs and maximize output, sales, or productivity. The same survey results showed that even though the definitions and relevance of these definitions of daylighting vary from profession to profession, it is inevitable that good daylighting is essential for sustainable architectural design.

Integration of daylighting with buildings requires analyzing the position of the building with its site and microclimate, designing proper building and glazing elements properties to meet certain requirements, including increasing daylight levels in the building, protecting the occupants against excessive glare, enabling a good view, minimizing solar heat gain in the summer while maximizing it in winter (Boubekri, 2014). Daylight can be integrated using passive design strategies. Figure 2.3 illustrates the passive design strategies used for integrating daylight with buildings.

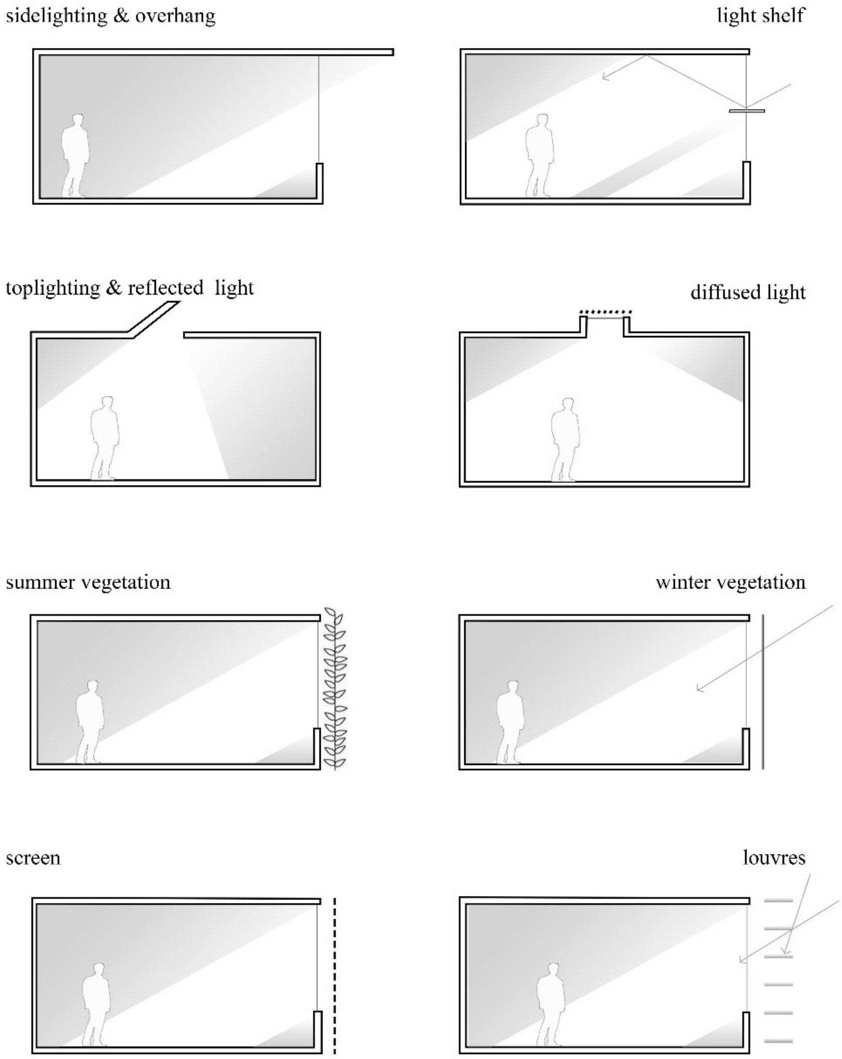


Figure 2.3. Passive design strategies to integrate daylight into the buildings

Mainly, daylight penetrates the building through toplight and sidelight (Bainbridge & Haggard, 2011). **Top lighting strategies** may vary depending on the usage and scale of the space and structure. Top lighting provides deeper penetration of daylight into the space. Generally, it enables uniform lighting distribution. Several strategies, including clerestory windows, sawtooth roofs, roof monitors, skylights, light tubes, and wells, are applied as top lighting strategies in buildings. Figure 2.4 shows the demonstrative applications for top lighting in buildings.

(a): University of New Mexico Law School, United States (Francesca Desmarais)

(b): Green Lighthouse, Denmark (Adam Mark, Courtesy of VELUX Group)

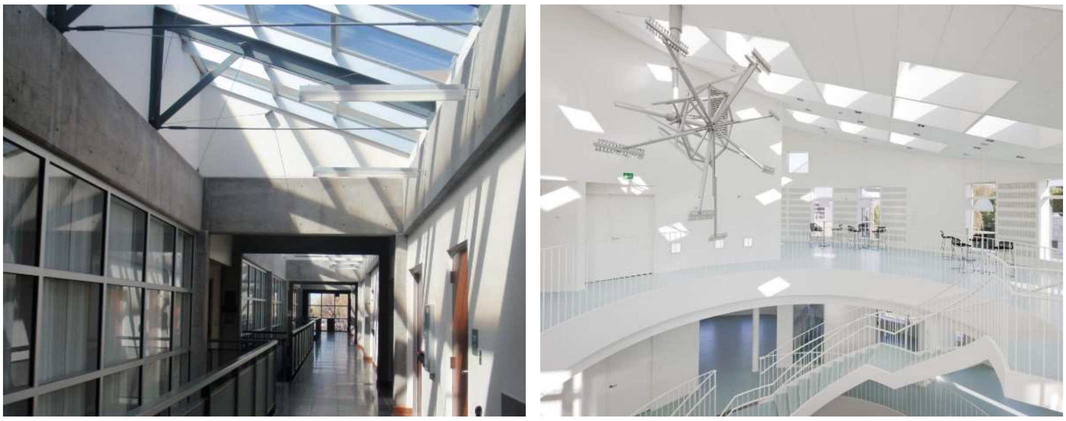
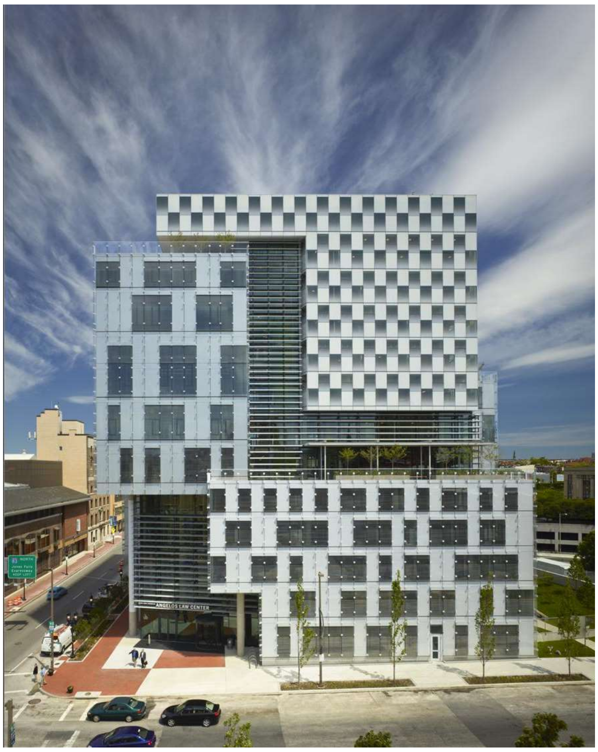


Figure 2.4. Top lighting applications in buildings

Sidelight systems are the ones that collect and direct daylight inside the spaces through the openings in the wall (Boubekri, 2014). In sidelight systems, consideration of height between floor and ceiling, implementation of light shelves, shading elements, properties of surfaces, and sloped ceilings determine the quality of the sidelight (Bainbridge & Haggard, 2011). In sidelight systems, increasing the ceiling and windows height of the space enhances daylight penetration. Light shelves reduce glare and provide a more directed distribution by reflecting the light to the ceiling and further away. A light shelf separates a side window into the top portion and the lower portion day (Boubekri, 2014). The top part is called a clerestory window, and the lower part is called the view window. They can be located in both the exterior and interior of the spaces. In order to reduce glare, exterior and interior lighting shelves can be combined. According to the desired effect in the space, the glare can be adjusted by changing the shelve in reflectance feature. Adding shading elements such as louvers, overhangs, and blind systems is also the strategy to adjust daylight and minimize the glare used in side lighting. The essential point in implementing these systems is to design to control the heat gain and daylight penetration before or after entering the building by considering the climatic conditions and long-term usability (Phillips & Gardner, 2012). Increasing the inclination of the ceiling in the use of light shelve may result in the ceiling being lower than the windows and cause more light to reflect into the space. Figure 2.5 shows the applications of side lighting strategies in the buildings.

(a): John and Frances Angelos Law Center, Baltimore (Behnisch Architekten/Ayers Saint Gross)



(b): Lewis Integrative Science Building, USA (HDR/THA Architecture; © 2013 Lara Swimmer)

(c): Research SupportFacility at the National Renewable Energy Laboratory, USA (LightLouver™)

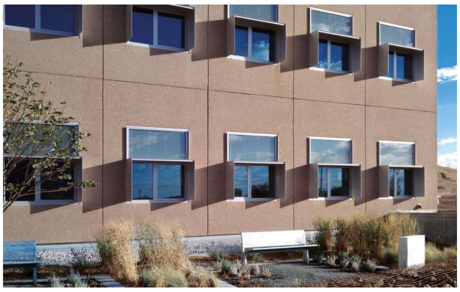


Figure 2.5. Side lighting applications in buildings

Another significant passive design technique is landscaping. Some vegetation offers building shade, cooling, and solar control during various seasons (Bainbridge & Haggard, 2011). Leafy and tall plants provide shading in summer and block sunlight, while deciduous plants in winter allow daylight to reach places more easily.

Implementation of passive design strategies increases the use of daylight, reducing the need for artificial lighting. However, it does not only provide control for electric consumption but also its visual and non-visual effects affect occupants' physical health, psychology, and productivity. In the next chapters, the visual and non-visual effects of daylighting on occupants and energy usage will be analyzed.

2.2 Daylighting and Non-visual/ Visual Effects on Occupants

2.2.1 Non-visual Effects of Daylighting on Occupants

Health and well-being are significantly impacted by 24-hour rhythms (light-dark cycle), which dominate many areas of human psychology and behavior (Lockley, 2009). This cycle affects people in many different ways, from the state of being awake or not to the secretion of the hormone melatonin. Furthermore, the cardiovascular, metabolic, immune, and skin systems depend on the light and dark rhythm to maintain physical and mental health (Veitch & Galasiu, 2012). Not only 24-hour rhythms but also seasonal changes in daylight is effective on hormone cycles (Webb, 2006). Therefore, the non-visual effects of the daily and periodic changes in daylight should be considered while considering the daylight design. Daylight should be used in spaces, especially where efficiency in occupant production should be at the forefront (such as offices). Daylight has been proven effective in users' creativity and productivity. Poor lighting quality may negatively impact occupants' health, leading to stress and different problems, including visual discomfort, vision, or posture (Phillips & Gardner, 2012).

2.2.2 Visual Effects of Daylighting on Occupants

The visual effects of daylighting are associated with visual comfort, including vision and glare. Visual comfort means being free from sensitivity and distraction and is directly related to glare protection and view outdoors or indoors (Nasrollahi &

Shokri, 2016; Tabadkani et al., 2021). The CIE defines glare as a condition that may be uncomfortable or impair visual performance and visibility stemming from a lack of proper luminance distribution or high contrasts in the visual area (CIE, 1995). Glare does not necessarily have to cause discomfort or disability. If the glare results in a decrease in visibility, it is called 'disability glare'. However, if it only causes discomfort without affecting visual performance, it is called 'discomfort glare' (Iuliano et al., n.d.). Characteristics of daylight, including the sky conditions, intensity, and distribution, color may lead to glare problems (Nazzal, 2005). If daylight is integrated into the excessive interior, sunlight may create a glaring problem. Therefore, design phases should consider the visual effects of daylight since a good daylight design aim to offer adequate light for effective visual performance while guaranteeing adequate comfort (Wienold & Christoffersen, 2006).

2.3 Daylighting and Energy

Rapid urbanization stems from a significant population shift from rural areas to cities, one of the century's significant problems (Ali et al., 2021; Ang et al., 2022). United Nations reported that whereas the people who live in urban areas are 30% of the total population, this ratio is expected to reach 68% by 2050 (United, 2018). It is predicted that the number of people living in cities will rise to 5 billion by 2030. While rapid urbanization promotes economic and social growth, it also increases energy demand and greenhouse gas emissions (W. Li et al., 2017). The rise in greenhouse gases will lead to drought, an increase in sea levels, climate change, and dangering living life and the existing ecosystem (Olabi & Abdelkareem, 2022). In developed and developing countries, buildings are one of the most significant greenhouse gas sources, as buildings’ energy demand represents almost 40% of the total energy of the cities (Argonne et al., 2017). In order to reduce GHG, the existing buildings’ energy demand should be reduced by considering energy-efficient strategies, and the new buildings should be designed to meet sustainable requirements. Integrating multiple performance variables, including daylight, energy consumption, visual and thermal comfort with buildings, are necessary to meet sustainable goals (Evins, 2013). Lighting is responsible for energy usage, representing 20 percent of global electricity consumption (Waide & Tanishima, 2006). Lighting energy consumption buildings account for approximately 11% of the total energy used in the building (Department of Energy, 2015). Several techniques are used to reduce energy demand, mainly changing the window type and shading elements and replacing the existing artificial lighting with more efficient lamps. (Mangan & Oral, 2016) applied different window alternatives for five climatic zones of Turkey to assess the residential building performance in Turkey. Similarly, Krarti (Bichiou & Krarti, 2011) changed the window types to optimize the building envelope. (Brandão De Vasconcelos et al., 2016) identified the retrofit scenarios, including changing window type, and calculated the primary energy demand for each scenario. Studies showed that changing the window types is an efficient retrofit scenario that decreases the energy usage for artificial lighting as it enhances the quality of the glazing element.

The second strategy is the usage of shading elements. (Hamdy et al., 2013) tested different shading types to prevent summer overheating. Using efficient shading elements can be a beneficial strategy in terms of energy and cost. (Corrado et al., 2014) applied different types of shading based on energy-efficient measures. Again, literature studies showed that the usage of efficient shading decreases overheating and energy used for cooling. Also, different types of shading (such as dynamic) can be used to adjust the daylight. Figure 2.6 demonstrates the different types of shading elements used in buildings.

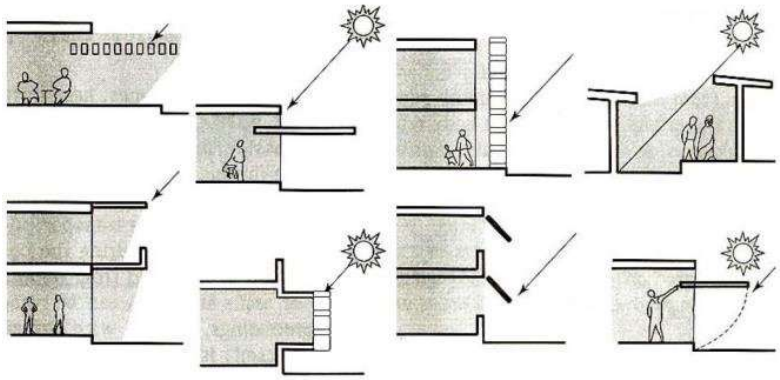


Figure 2.6. Different types of shading elements (Kamal, 2010)

The last strategy is changing the lighting equipment load by changing the lighting element. Standard incandescent heat lamps consume much more electricity than efficient LED lamps. Therefore, most studies deal with retrofit scenarios that change the lighting load per square meter in the literature. (S. Yang et al., 2016) showed that lighting power density is one of the most dominant factors affecting annual electricity usage. (Rackes & Waring, 2017) used different distributions for lighting power density to simulate. Lighting power density is one of the most significant indicators of economizer savings. Literature studies showed that lighting retrofits are meaningful in carbon emission and energy demand reduction. Although alternative lighting retrofits are applied, the energy used for lighting is still substantial (Boubekri, 2014). Except for mentioned strategies, daylighting is a passive design strategy to reduce the lighting energy demand. The proper use of daylight as a free local energy source is essential (Dogan et al., 2012). Daylight penetrating the buildings is an effective energy-efficient strategy in terms of its influence on reducing electric lighting demand (Zhou & Liu, 2015). The reduction of electric lighting demand leads to less heat in the building provided by electric lighting (Tzempelikos & Athienitis, 2007); accordingly, the cooling demand of the building can be reduced. Daylight integration with buildings reduces the energy demand of the building and increases the health, productivity, and visual comfort of the occupants (Lockley, 2009; Webb, 2006).

However, dense urbanization prevents daylight access to the buildings. Each building act as a shadow element inheriting the daylight penetration of the buildings. Urban form parameters (street orientation and width, surrounding building heights and distances to the building, etc.) are highly influential on daylight access to each building (Saratsis et al., 2017). In the design stages, there is a need to analyze the relationship between context and buildings to benefit from daylight as much as possible.

2.4 Factors that Affect Daylighting Illuminances

Many factors affect daylight illuminance, which can be divided into external and internal parameters (Ayoub, 2020). External parameters are divided into climate conditions, temporal settings, and external obstruction parameters. Internal parameters are divided into physical features, openings, and shading devices. These parameters must be examined in detail and correctly given to the analysis models to make accurate and high-resolution daylighting analyses at an urban scale.

External Parameters

Climate condition refers to the location-dependent parameters. Global horizontal illuminance, direct normal, and diffuse radiation are one of the main climate parameters that affect daylight illuminances. The higher values of them result in higher daylight penetration to the buildings.

Temporal settings refer to the time of the day and the day of the year. Depending on the month, the day, and the time of the day, the illumination level of the spaces changes with the change of the angle of incidence of the sunlight.

External obstruction parameters (urban form parameters) include context buildings' height, width, reflectance, and the distance between the context and the building itself**.** Urban morphology influences daylight penetration, solar radiation, and the view. Therefore, some environmental assessment schemes such as HK-BEAM (HK-BEAM Society, 2004) and LEED (U.S. Green Building Council, 2008) evaluate the influence of neighboring buildings on access to daylight and views by combining view and overshadowing performance (Fung & Lee, 2012). Especially the location of the surrounding buildings determines how much daylight penetrates the buildings. If the surrounding buildings are parallel to the examined building, the daylight reaching the building is largely blocked (Munoz et al., 2014). As the angle of the surrounding buildings with the analyzed building increases, the daylight effects on that building decrease. Daylight penetrates much more into buildings that do not have close-range context buildings. The closer the surrounding buildings are to the analyzed building, the more they block the daylight. As the surrounding buildings move away from the examined building, their effects on the examined building decrease. The distance between buildings and the height difference between the surrounding building and the investigated building affects daylight penetration. An increase in the floor height of the surrounding buildings casts more shadow on the examined building and prevents daylight. *'Obstruction Angle'* is calculated when the distance between the buildings and the relationship between floor heights are wanted to be examined simultaneously. The obstruction angle is the angle between the two lines indicating the distance between the two buildings and the difference between the end of the opposite building from the middle of the building's window. When the distance between buildings is constant, increasing the obstruction angle increases the shading effect of the opposite building. Southeast and southwest orientations are crucial for solar access in urban settings; therefore, the obstruction angle should preferably not exceed certain degrees by obstructions in this zone (Littlefair, 2001). Also, the reflectance of the surrounding elements, including trees, buildings, and ground surfaces, is important as they act as reflective surfaces. A significant chunk of the overall radiation is received from the interreflections between the surfaces (Bugeat et al., 2019). The interreflections in the exterior surfaces with reasonable values improve the daylight conditions. Darker surfaces absorb sunlight, while lighter surfaces reflect sunlight. The materials used on exterior surfaces in urban areas should also be carefully designed to prevent urban heat island formations and to control the sunlight reflected on the buildings.

Internal Parameters

Building physical features parameters include geometrical properties of the building/room (width, length, height, width/depth ratio, area), the orientation of the building, and reflectances of building elements (wall, floor, ceiling). Geometrical properties of the building/room are important for penetrating daylight through the interior. The higher heights of the space mean higher daylight penetration to the building (O'Connor, 1997). Furthermore, long and narrow footprints are preferable to square ones for better daylight penetration. The shallower floor plans require less artificial lighting (Gadelhak, 2015). Also, for daylighting control strategies, deep facades should be designed. A facade with a certain depth can act as a shading element to block the sun (Figure 2.7).

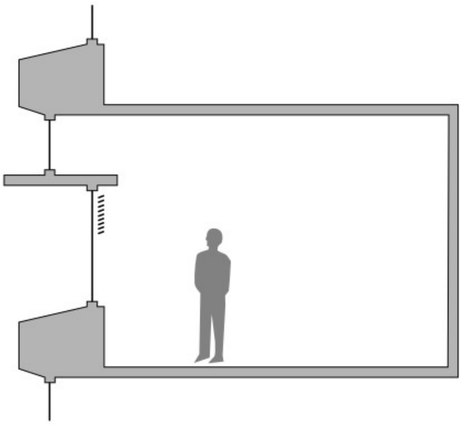


Figure 2.7. Deep-wall section to provide self-shading (O'Connor, 1997)

Orientation of the building has a significant impact on both building energy consumption and daylighting control. Therefore, it must be considered in the early design stages to maximize the amount of beneficial natural light and sunlight penetrating the interior (Phillips & Gardner, 2012). Deciding on the proper orientation is essential for minimizing the amount of electricity used by artificial lighting, which is responsible for about 20% of the total energy consumption of buildings (Halonen et al., n.d.). North-facing windows are an efficient source of daylighting as most of the sunlight that enters a room from the north is indirect throughout the year leading to a reduction in glare and heating. Therefore, the spaces should face north when constant lighting levels are needed. If adequate shading precautions are not taken, the excessive sunlight entering through south-facing windows can cause glare and discomfort to the eyes.

Additionally, it may cause overheating, raising the need for cooling. The proper horizontal shading elements (overhang) should be implemented on the south faces to prevent glare and overheating. While east-facing windows get direct morning sun, west-facing windows receive sunshine in the afternoon. For east and west facades, louvers should be implemented as daylighting control strategies. Whereas the amount of south- and north-facing facade area are maximized, east, and especially west exposure should be minimized for the controllable daylight fenestration. Another important factor for daylight penetration is the reflectance of the surface elements; therefore, the color and texture of the interior and exterior surfaces should be designed properly. Surfaces with a light color reflect more light than dark-colored surfaces. Specular surfaces like mirrors may lead to glare and visual discomfort. If the aim is to distribute daylight uniformly, reflected materials should be used in interiors. However, reflected materials for the exterior can cause visual discomfort to surrounding buildings. Therefore, exterior and interior materials' reflectances should be properly considered.

Openings and shading device parameters refer to the parameters related to glazing and shading elements. The properties of openings include geometrical properties of the window (WWR, window height, window width), orientation and location of the window, number of windows, visible transmittance value (VT), and Solar Heat Gain Coefficient (SHGC). WWR is the ratio between the window and wall area. Window area affects daylight penetration, heat gain, and loss. In addition to increasing the amount of daylight, larger windows also result in more heat gain and loss. Therefore, windows with less visible transmittance should be preferred if larger windows are used. Windows can be vertical or horizontal, depending on the aspect ratio. Continuous strip windows provide uniformly distributed and adequate daylighting for a space (O’Connor, 1997).

On the other hand, punched windows prevent uniform daylighting distribution due to the gaps between them (Figure 2.8). Depending on the function of the space, such windows can be preferred, especially in working areas. Although vertical windows provide deeper daylighting access, they do not provide equal light distribution to the space. The aspect ratio of the windows should be carefully chosen according to the function of the space and the intended light intake.

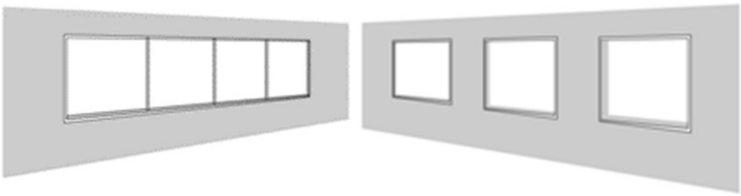


Figure 2.8. Strip windows vs. punched windows (O'Connor, 1997)

The location of the windows is highly influential on daylight distribution, view, and glare. A daylighted zone's practical depth is normally constrained to 1.5 times the window's height, and it can be extended up to 2.5 times using a reflecting light shelf (O'Connor, 1997). High windows let more daylighted zone into the spaces for deeper spaces than the lower ones (Chartered Institution of Building Services Engineers [CIBSE], 1999). High windows can result in glare problems stemming from the skylight. At this point, reducing window size reduces glare and energy costs. Visible transmittance (VT) is a fraction of the visible light transmitted through the glazing. A higher VT means more visible light. As the number of layers of the glazing increases (double, triple), the visible transmittance values decrease. The decrease in visible transmittance reduces the light transmitted into the space and affects the view quality. The solar heat gain coefficient (SHGC) measures how much solar radiation is transmitted or absorbed through a window. Lower SHGC means less heat transmitted. More heat is transported when the SHGC is higher, which is advantageous for capturing solar heat in the winter when it is undesirable in the summer (Feuermann & Novoplansky, 1998). The appropriate values of VT and SHGC vary depending on the space's location, orientation, and window size.

Shadings are systems that shade spaces from sunlight, guide natural light, shield the interiors from glare, and avoid overheating (Gherri, 2015). The usage of adequate shading devices decreases energy consumption by preserving the needed daylighting illuminance levels for the space avoiding glare (Ye et al., 2016). A high amount of solar radiation can be prevented with the proper use of shading devices, which decreases the cooling load in summer and allows the penetration of the needed solar gain in winter. Shadings can be classified based on their implementation: internal or external, and their operation type: fixed and mobile. Internal shading systems are mainly in the form of rollers, Venetian blinds, and easily adjustable curtains. External shading systems can be in the form of external solid or perforated shade devices. External shading systems can prevent most of the outside heat from entering a space; however, internal shading systems distribute the heat that has already entered the space (Figure 2.9).

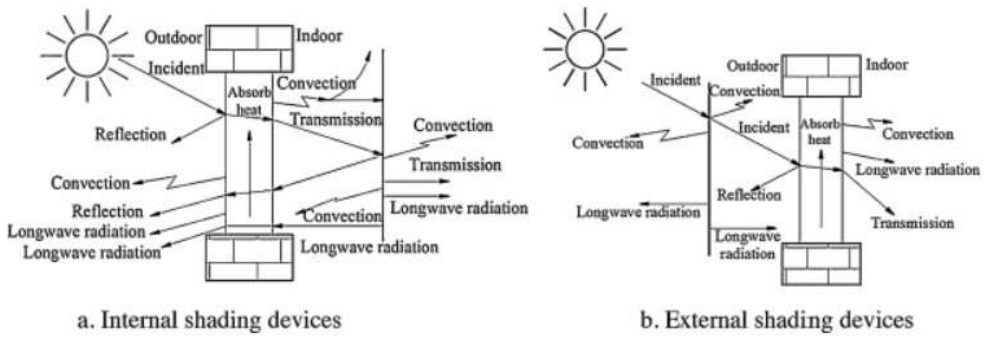


Figure 2.9. External shading systems vs. internal shading systems (Ye et al., 2016)

Mobile shading devices include Venetian blinds, vertical blinds, and roller shades that can control sunlight based on the adaptation of variable factors, including solar radiation, indoor temperature, and illuminance level (Kirimtat et al., 2016). Fixed devices, including overhangs, horizontal/vertical louvers, and egg-crate, can not protect the space according to the dynamic relations arising from environmental factors such as daylight and solar radiation.

Shading devices provide daylight by preserving the necessary heat gain for spaces, and some of them can provide a view by preventing glare and leading to visual comfort for the occupants. Therefore, the proper integration of daylight with spaces should be carefully designed based on daylight requirements.

2.5 Daylighting Performance Metrics

This section examines the literature to evaluate daylight performance indicators. Daylight performance metrics are important tools for assessing the energy savings and the potential of natural light in a space. There are several daylighting performance metrics shown in Table 2.1. However, a few metrics most commonly used to analyze indoor daylighting in the studies will be examined in detail. These metrics can be grouped under two main categories in terms of calculation methods which are Static Performance Metrics (Daylight Factor) and Dynamic Performance Metrics (Daylight Autonomy, Useful Daylight Illuminance, Daylight Availability, and Annual Sunlight Exposure (Bellia et al., 2017)).

Table 2.1 Daylighting Performance Metrics Classification based on the (Ayoub, 2019a) study

	Daylight Metrics		***Direct Sunlight Metrics***		***Glare Indices***
S	Daylight Factor	**D**	Annual Light Exposure	**D**	Glare Index
D	Daylight Autonomy	**D**	Sunlight Duration	**D**	Visual Comfort Probability
D	Continuous Daylight Autonomy	**D**	Sunlight Beam Index	**D**	CIE Glare Index
D	Maximum Daylight Autonomy	**D**	Annual Sunlight Beam Index	**D**	Daylight Glare Index
D	Minimum Daylight Autonomy	**D**	Annual Sunlight Exposure	**D**	CIE Unified Glare Index
D	Useful Daylight Illuminance			**D**	Daylight Glare Probability (DGP)
D	Daylight Availability			**D**	Simplified DGP
D	Frequency of Visual Comfort			**D**	Enhance Simplified DGP
D	The intensity of Visual Comfort			**D**	Unified Glare Rating
D	Spatial Daylight Autonomy			**D**	Annual Visual Discomfort Frequency

S : static, D: dynamic

2.5.1 Static Performance Metrics

Daylight Factor

The oldest performance indicator is the 'Daylight Factor' introduced by Trotter in 1895 (Walsh, 1951). Daylight Factor (*DF*) is the ratio of internal illuminance to external horizontal illuminance under an overcast sky defined by the CIE (Commission Internationale de l´Eclairage) luminance distribution (Tregenza, 1980). Thus, the daylight factor (DF) is calculated by dividing the inside illuminance level by the outside illuminance level (Equation 2.1).

$$DF = \frac{Indoor\ illuminance\ from\ daylight}{Horizontal\ unobstructed\ outdoor\ illuminance} \times 100\%$$

Equation 2.1. Daylight factor calculation

Also, as DF is composed of three main components, it can be calculated by summing the Sky Component, Externally Reflected Component, and Internally Reflected Component (Acosta et al., 2013). A representation of these components is given in Figure 2.10. Equation 2.2 shows the calculation of DF based on these three components.

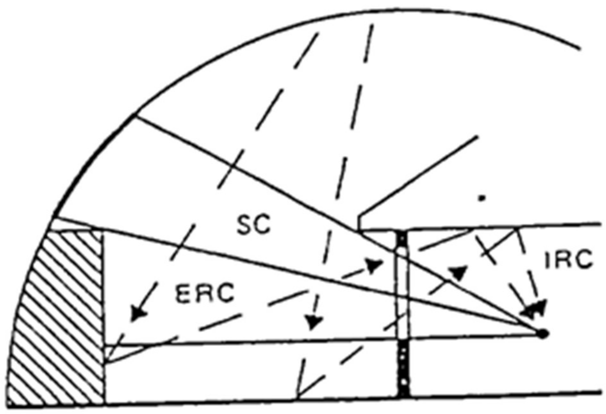


Figure 2.10. Components of natural light (Szokolay & Brisbin, 2004)

$$DF = SC + ERC + IRC$$

SC – Sky Component

ERC – Externally Reflected Component

IRC – Internally Reflected Component

Equation 2.2. Daylight factor calculation based on SC, ERC, IRC

CIE defined the sky component as the daylight factor produced only by the sky luminance, excluding the reflectance of the inner surfaces (CIE, 2011). In other words, the sky component represents the ratio of daylight falling on a vertical surface to the daylight available under an unobstructed sky.

Externally reflected component refers to the light reflected from external surfaces such as surrounding buildings, trees, and grounds (Fakra et al., 2011). The reflectance of the surroundings, including the ground surface and external obstructions, directly affects the externally reflected component (Mohelnikova & Hirs, 2016).

Internally reflected component refers to the light that reaches a reference point from reflections inside the space. Daylighting in interiors is affected by room geometry, colors, and surface patterns leading to inter-reflections directly influencing internally reflected components (Mohelnikova & Hirs, 2016; Singh & Rawal, 2011).

Daylight Factor provides quick feedback on how much daylight can be expected for the space. However, there are several limitations of DF. It can not represent the change in illumination levels indoors stemming from temporal variation of sky luminance and ignores the effects of building orientations, which directly influence daylighting illuminances (Kota & Haberl, 2009). Even though there are modified versions of the DF method to analyze the effect of different sky conditions (clear sky & sunlight), these methods ignore the dual effect of reflected sunlight from both ground and obstruction (Alshaibani, 1999). Furthermore, it does not give an idea about the glare caused due to daylight.

2.5.2 Dynamic Performance Metrics

Illuminance

Daylighting refers to the illumination of spaces by daylight delivered through openings, and this illumination may stem from direct sunlight, skylight, or reflected light (Knoop et al., 2020). Figure 2.11 illustrates the components of daylighting.

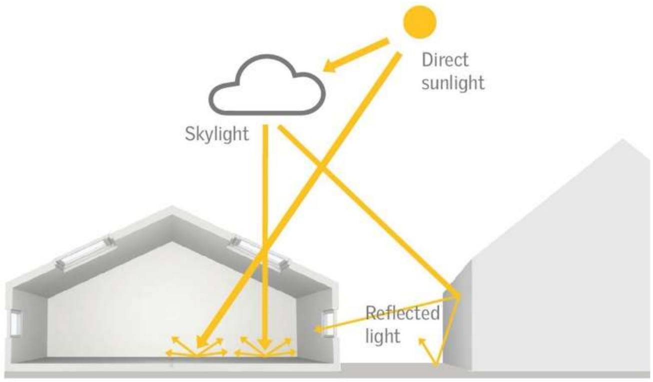


Figure 2.11. Components of daylighting (Roy, 2014)

Except for the feeling of the space, daylight should be considered a design element to achieve a certain illuminance level that which function of the space requires. Illuminance can be described as the amount of light falling on a surface. Each different activity type requires different illuminance levels. Table 2.2 shows recommendations of the IESNA (Illuminating Engineering Society of North America) for the different activities and their required illuminance values (IESNA, 2000).

Table 2.2 IESNA recommendations for specific tasks

Task	*Illuminance*
Reading	30
Meeting Rooms	300
Classrooms	300
Office Spaces (intensive computer tasks/ variety of tasks)	300/500
Fine Machine Work	3000-10000

Also, IESNA and CIBSE (Chartered Institute of Building Services Engineering) have recommendations for the illuminance values of residential spaces (CIBSE, 2013; IESNA, 2007). Table 2.3 shows the recommended illuminances for residential spaces.

Table 2.3 IESNA and CIBSE recommendations for illuminance values of residential spaces (lux)

Task	*IESNA 2007*	*CIBSE 2013*
Hall	100	200
Lounge	300	-
Kitchen	300	200
Dining/Living Room	300	200
Bathroom	300	150
Bedroom	300	100
Stairs	300	100

Daylight Autonomy

Daylight Autonomy (DA) is a dynamic performance metric proposed in 1989 by the Association Suisse des Electriciens (Association Suisse des Electriciens, 1989). It represents the percentage of the annual occupied timesteps when the minimum illuminance threshold is met by daylight alone (Reinhart & Walkenhorst, 2001). It is a climate-based metric that considers the various sky conditions and facade orientations. Higher DA value results in less usage of electric lighting. DA_{300}lx [50%] can be expressed as the percentage of the work plane area that receives daylight above 300 lux for at least 50% of the annual time (Bian & Ma, 2017). Required minimum illuminance levels (thresholds) for different spaces can be directly taken from various documents, including the IESNA Lighting Handbook (IESNA, 2000).

Later on, several modified versions of DA were developed. Continuous Daylight Autonomy (*cDA*) indicates the illuminance falls below the minimum illuminance threshold considering a partial credit linearly to values below the threshold defined (Reinhart et al., 2006). Minimum Daylight Autonomy is the percentage of occupied time when an illuminance threshold can be met by daylight alone under continuous overcast sky conditions (Acosta et al., 2019). Maximum Daylight Autonomy is the percentage of the annual occupied timesteps when the illuminance level exceeds ten times a predefined threshold (Rogers & Goldman, 2006).

The usage of modified versions of DA is not wide as the use of DA; however, they can be helpful in certain cases. Recently, a new metric was proposed called Spatial Daylight Autonomy (sDA) by IESNA (Illuminating Engineering Society of North America). Spatial Daylight Autonomy (sDA) is a measure of daylight illuminance sufficiency for a given area, reporting a percentage of floor area that exceeds a specified illuminance level (e.g., 300 lux) for a specified amount of annual hours (IESNA, 2012). Illuminating Engineering Society of North America (IESNA) defined this threshold as 300 lux for 50% of all occupied hours ($sDA_{300/50\%}$), and the $sDA_{300/50\%}$ threshold is referenced in both the LEED (Leadership in Energy and

Environmental Design) (Council, 2013) and WELL (Institute, 2020) building certification systems (Turan et al., 2020). Therefore, it is an important daylighting performance indicator for sustainable design. It allows for evaluating the space for one year regarding the sufficiency of daylight illuminance. sDA considers both temporal and spatial dynamics of daylight in buildings (Kazanasmaz et al., 2016).

Useful Daylight Illuminance

Useful Daylight Illuminance (*UDI*) is defined as the percentage of the annual occurrence of illuminances across the work plane where all the illuminances are within the range of 100-2000 lux (Nabil & Mardaljevic, 2005). Later, UDI is redefined as a percentage of the occupied time during the year when the illuminance value is between 100 and 3000 lux (Mardaljevic et al., 2012). Then, the UDI range is further subdivided into two main categories: UDI-supplementary and UDI-autonomous. UDI-supplementary shows the presence of daylight illuminances in the range of 100 to 300 lux. In this range of illumination, additional artificial lighting can be required to ensure the necessary daylight for common tasks such as reading and drawing. UDI- autonomous gives the occurrence of daylight illuminances in the range of 300 to 3000 lux, where extra artificial lighting will probably not be required. Including these two main categories, the UDI range is divided into five sub-categories:

- UDI fell-short (UDI-f) represents the percentage of the occupied time where illuminance is less than 100 lux.
- UDI supplementary (UDI-s) represents the percentage of the occupied time where illuminance is more than 100 lux and less than 300 lux.
- UDI autonomous (UDI-a) represents the percentage of the occupied time where illuminance is higher than 300 lux and less than 3000 lux.
- UDI combined (UDI-c) represents the percentage of the occupied time where illuminance is higher than 100 lux and less than 3000 lux.

- UDI exceeded (UDI-e) represents the percentage of the occupied time where illuminance exceeds 3000 lux.

UDI gives more information about assessing visual comfort and thermal discomfort than daylight autonomy metrics (Ayoub, 2019a). Furthermore, UDI gives an idea about the percentage of occupancy hours in which oversupply of daylight happens, and insufficient daylight occurs by indicating lower and upper thresholds (Yu & Su, 2015).

Daylight Availability

Daylight Availability is a dynamic metric proposed by (Reinhart & Wienold, 2011) to combine Daylight Autonomy and Useful Daylight Illuminance as a single indicator. It is represented as fully daylit, partially daylit, non-daylit, and overlit (Ayoub, 2019a). If illuminances exceed 300 lux for at least 50% of the annual occupied timesteps, it is indicated as fully daylit ($DAv_{300/50\%}$) (Reinhart, Rakha, & Weismann, 2014). Partially daylit reports the illuminances above 150 lux for at least 50% of the annual occupied timesteps ($DAv_{150/50\%}$). Overlit calculates the illuminances that exceed ten times the target illuminance for at least 5% of the annual occupied timesteps ($DAv_{3000/5\%}$), giving an idea about possible visual discomforts (Reinhart & Wienold, 2011).

Annual Sunlight Exposure

Annual Sunlight Exposure (*ASE*) is a direct sunlight metric that provides a second dimension of daylight analysis, looking at one potential source of visual discomfort: direct sunlight (IESNA, 2012). ASE indicates visual discomfort and the amount of direct sunlight the workplace receives annually. When ASE is used with spatial daylight autonomy, it provides a meaningful understanding of how a space should be designed for adequate and good daylighting in terms of visual comfort. According to the LEED v4 ((2022 U.S. Green Building Council, n.d.), no more than 10% of space should have direct sunlight of more than 1000 lux for a maximum period of 250 hours per year ($ASE_{1000/250}$).

In summary, each daylighting performance metric provides information about daylighting performance from different perspectives; some represent how much the space is illuminated above certain values during the year, and some represent the space area that is illuminated as much as certain illumination values. Dynamic performance metrics are specific analyzes obtained by calculating illuminance values. Illuminance is the rawest data among these metrics, and its calculation is also associated with areas such as energy use related to daylighting.

2.6 Daylighting Calculation Methodologies

In this section, the daylighting calculations used in the literature were given in detail. Throughout history, several methods have been proposed to calculate daylighting performance metrics. Mainly, these can be categorized into six groups: graphical methods (diagrams, charts, tables), non-graphical methods (protractors, daylight factor calculator), analytical formulas, scale models, computer simulations, and machine learning (ML) models (Kazanasmaz et al., 2009; Kota & Haberl, 2009).

2.6.1 Graphical Methods

Waldram and Waldram developed one of the first graphical methods attempts in 1923 to estimate the 'Sky Component' (Waldram & Waldram, 1923). Waldram Diagram graphically enables to representation of the amount of light from the sky at any particular point within a room (Defoe, 2016). In 1954, the pepper-dot chart method was produced by Pleijel to calculate the sky component for the standard CIE sky (Hopkinson, Petherbridge, & Longmore, 1966). Turner developed a similar method in 1969 to estimate sky components by using charts consisting of a pattern of dots (Fuller, 1985). In 1980, the 'Graphic Daylight Design Method' was proposed by Millet to represent the daylight factor by contours for the overcast sky (Millet, Adams, & Bedrick, 1980). Later, the proposed methodology was extended to represent a clear sky (Fuller, 1985).

2.6.2 Non-graphical Methods

Non-graphical methods include protractors and daylight factor calculators (Ayoub, 2019a). Protractors calculate the sky component of the daylight factor (Dufton, 1946) by dividing the sun path projections into solar hours (Ne'eman et al., 1976). It was developed to simplify daylight calculation and make it possible to measure daylight directly from architectural drawings (Bryan & Carlberg, 1985).

2.6.3 Analytical Formulas

Analytical formulas calculate different daylight components such as daylight factor, internally reflected component, and externally reflected component (Kota & Haberl, 2009). In 1928, Frühling developed the Lumen Method to forecast the DF with mathematical formulas (Frühling, 1928). However, the biggest shortcoming of this method was ignoring the daylight stemming from an externally reflected component. In 1954, Dresler extended this methodology to integrate with internally reflected components requiring intensive measurements (Dresler, 1954). To develop this idea, Arndt developed a new formula to consider internally reflected components without requiring intensive measurements. Later on, based on Arndt's formula, the split-flux method was developed by Hopkins, Longmore, and Petherbridge (Hopkinson et al., 1954). The split-flux method considers the reflections from external obstructions. Tragenza modified this formula to involve large vertical obstructions (Tregenza, 1989). The split-flux method works well with a certain geometry but is unsuitable for complicated geometries. In 1979, the 'Average Daylight Factor' was proposed by Lynes (Lynes, 1979), requiring fewer parameters than the daylight factor, including opening size, shape, and position to extract information about the ratio between average internal and external horizontal illuminance for an overcast sky. The calculation methodologies mentioned above evaluate static metrics such as DF ignoring the changing weather conditions, orientation, and location. Therefore, in 1983, Tregenza & Waters (Tregenza & Waters, 1983) proposed 'Daylight

Coefficient' as a dynamic method to reflect the effects of variable conditions. The proposed methodology can consider different sky conditions, orientations, and locations.

2.6.4 Scale Models

Other attempts to calculate daylight inside buildings are scale models, which are the physical representation of buildings to assess daylighting performance. This method examines how daylight behaves within a small-scale model of a planned building under different sky circumstances to provide insight into how daylight would probably behave if the building in question were built (Ngarambe et al., 2022). The ratio of the scale models varies from 1:8 to 1:32 (Kazanasmaz et al., 2009). One of the biggest challenges of this method is to represent the building realistically. Finishing materials, interiors, and reflections of surfaces should be matched with the real buildings (Littlefair, 2002). In order to avoid unwanted light penetration, the location for measurements should be chosen correctly. Ensuring the mentioned guidelines is challenging; therefore, these models have lost their usefulness over time (Thanachareonkit et al., 2005). Figure 2.12 shows examples of scale models.

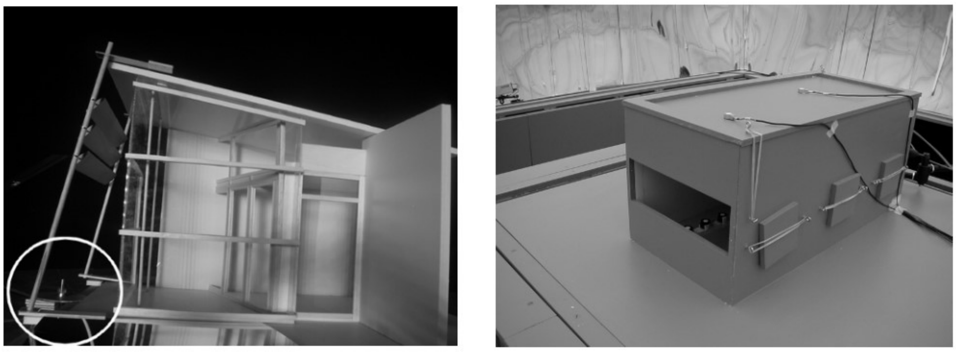


Figure 2.12. Scale models for daylight penetration (Bodart et al., 2007)

2.6.5 Computer Simulations

Traditional daylight calculation methods only consider the standard overcast sky without accounting for the various sky conditions. Furthermore, they are not useful for evaluating different design alternatives, including orientation, window types, and shading elements. It is worth noting that the methodologies described above (graphical methods, non-graphical methods, analytical formulas, and scale models) are insufficient for providing extensive assessments of daylighting performance. On the other hand, computer simulations allow designers/architects to evaluate a more extensive analysis to estimate daylighting illuminance inside buildings. Furthermore, they are easy-to-use design tools as they help optimize the use of daylight in buildings. Various design alternatives can be evaluated by using simulations in terms of daylight. Evaluating the visual performance and energy efficiency with the use of daylight considering the changing conditions (location, orientation, climate) is possible with computer simulations. Climate-based daylight modeling (CBDM) helps to predict any luminous quantity (illuminance/luminance) on grids using realistic sky conditions retrieved from representative climate data of the specific location at specific temporal resolutions (Mardaljevic, 2006; Reinhart & Herkel, 2000). Computer simulations enable the calculation of daylighting metrics based on climate-based modeling. There are many computer simulation tools to analyze daylighting performance metrics. These can be categorized under two groups based on illuminance calculation methods: radiosity and ray-tracing techniques (Kazanasmaz et al., 2009).

The radiosity method is based on dividing the room surfaces into a mesh of polygons, and for each divided part, the direct light is calculated (Chartered Institution of Building Services Engineers [CIBSE], 1999). This process is continued until the reflections between each polygon are measured. In this method, each element can receive light from it, which means each element act as a light source (Ashmore & Richens, 2001). This method assumes all surfaces are perfectly diffusing. There are some limitations to the usage of the radiosity method. Firstly, this method requires a meshing process to divide each surface into smaller ones, which takes much computational time. Secondly, even though specular reflectance and complex models can be simulated, they require much memory. Finally, as it assumes all surfaces are perfectly diffusing, the transparency effects of surfaces cannot be modeled easily. Delight (Hitchcock & Carroll, 2003) and Lightscape (*Autodesk*, 2000) are computer simulation programs based on the radiosity method.

The ray-tracing method calculates the visibility of surfaces by following light rays from the viewer's eye to the rendered scene's elements (Muneer & Beliveau, n.d.). A projection center, also called the viewer's eye, and an arbitrary view plane is chosen to render the scene on a plane. The ray follows the geometry of the reflection or transmission of the surface to the following surface if the surface is specular or transparent. The light received at the intersection point from the sources is calculated if other elements do not hinder the ray, and this process continues until the specified number of rays is attained (Chartered Institution of Building Services Engineers [CIBSE], 1999). There are two different ray-tracing methods: backward ray tracing and forward ray tracing. In a forward ray-tracing technique, rays are traced from the light source to the eye position. On the other hand, in a backward ray-tracing technique, a ray is followed back from the eye position until it reaches a surface.

The developed computer algorithms enable to trace of millions of rays to manage high-resolution renders. Unlike the radiosity technique, the ray-tracing technique can consider the specular reflectances and transparency effects and deal with complex geometries (Muneer & Beliveau, n.d.; Ward & Rubinstein, 1988). Figure 2.13 shows the rendering based on the raytracing (left) and radiosity (right) methods.

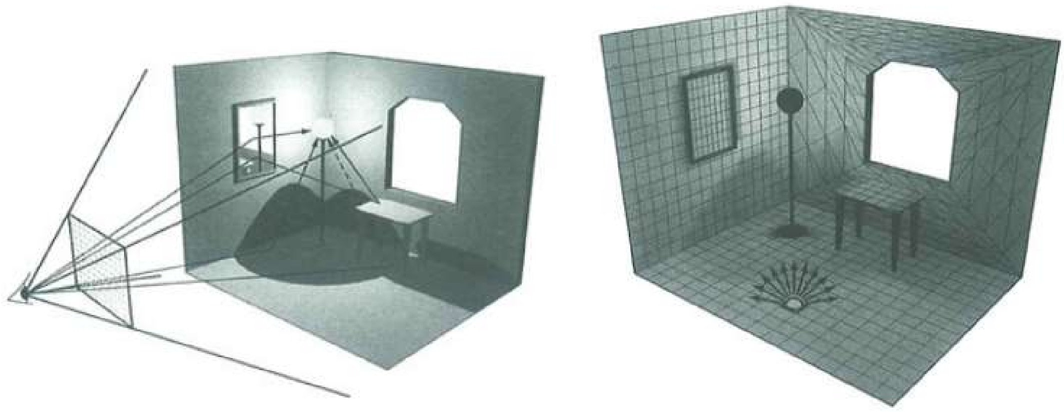


Figure 2.13. The rendering of a room based on the raytracing (left) and radiosity (right) method (Niedenthal, 2008)

Most simulation programs, including RADIANCE and GENELUX, are used for daylighting and electric calculation based on the ray-tracing technique (Muneer & Beliveau, n.d.).

There are several simulation programs used for daylighting and energy analysis of buildings. The most common simulation programs used will be covered.

RADIANCE is essentially a rendering engine developed by Lawrence Berkeley Laboratory (LBL) in California and the Ecole Polytechnique Federale de Lausanne (EPFL) in Switzerland (Ward, 1994). It uses ray tracing and offers comprehensive lighting analysis. The software employs backward ray tracing, which works with specular, diffuse reflections and complex, curvilinear geometry to produce effective rendering outcomes under many circumstances. Additionally, it mixes different ray-tracing methods, such as deterministic and stochastic, to balance between speed and accuracy of the rendering results. It is free software and a plugin that can be embedded with another program. RADIANCE is used by architects/designers for extensive daylighting analysis, including indoor illuminations, visual quality, and implementation of new lighting technologies. However, even though the software provides comprehensive lighting analysis, the use of the program is not easy to learn.

DAYSIM is a RADIANCE-based daylight simulation tool that performs dynamic simulations under various sky conditions using daylight coefficients (Reinhart & Walkenhorst, 2001). Unlike RADIANCE, DAYSIM uses daylighting coefficients based on Tregenza (Tregenza & Waters, 1983). With a single run, the program can easily deal with complex geometries. The outcome of the simulation results is extensive, including annual illuminance results, dynamic performance metrics, and glare indices (Ayoub, 2019a). Researchers prefer DAYSIM to predict indoor illuminances under different sky conditions. It is a stand-alone program; however, it can work with other programs such as EnergyPlus, Ladybug, and Honeybee. It is an open-source and accessible tool.

DIVA (Design Iterate Validate Adapt) is a plugin for the Rhinoceros 3D Nurbs modeling (Becker & Golay, 1999). Also, it is a simulation tool to integrate daylighting and thermal simulations (Jakubiec & Reinhart, 2011). Based on daylighting analysis, lighting schedules are created and can be shared with the energy simulation to calculate lighting energy demand based on daylight. It can be integrated with RADIANCE and DAYSIM (Reinhart & Walkenhorst, 2001). The plugin is mostly used for research as it has a cost except for educational purposes.

Energy Plus is a building software program to calculate the energy demand of buildings (Drury et al., 2000). It gives information about the buildings' heating, cooling, and lighting demand. The outcome can be given daily, monthly, seasonally, or annually based on user selection (Crawley et al., 2001). The program can be integrated with other programs, such as DAYSIM, and RADIANCE, to provide a more comprehensive daylighting analysis. For example, for accurate lighting energy demand analysis based on daylighting, the schedules created by DAYSIM are shared with Energy Plus, and Energy Plus calculated lighting energy demands based on these schedules.

Ladybug & Honeybee, developed by Ladybug Tools, are parametric plugins that can be integrated with 3D programs such as Rhino. They also can be integrated with RADIANCE, DAYSIM, and Energy Plus. As they are free and have straightforward usage, they easily enable designers to estimate energy predictions, daylighting calculations, and their visualizations.

Climate Studio has been recently developed by Solemma LLC for daylighting and energy modeling. It enables users to perform different analyses, including calculating climate-based daylighting metrics, glare analysis with varying time solutions, thermal zone energy, and load estimations.

2.6.6 Machine Learning Models

During the early design stages, computer-based simulations were a crucial part of daylight modeling. As mentioned above, numerous daylight simulation tools have been created over the years and are now commonly used in study and practice. Compared to the above-discussed scale models and analytical formulas, the current technologies offer a more detailed analysis. However, users must run simulations for several combinations of variables, especially for multi-objective optimization analyses, making it time-consuming to simulate complex buildings and select various parameters to maximize daylighting efficiency (Ngarambe et al., 2022). The simulations' results may vary depending on the level of modeling detail and the users' expertise. Machine learning algorithms employ data to learn correlations and patterns between the output and input parameters. After learning these patterns, the model may be used to predict the output variables under different circumstances. When an ML algorithm is sufficiently trained, it can be used to determine the most suitable combination of design parameters required for better daylighting performance. This property of ML algorithms offers a practical solution to the earlier discussed methods' limitations (time-consuming, requiring modeling knowledge and expertise) of daylight estimation in the design stages.

ML models reduce computation time based on learning patterns between input and output parameters (M. I. Jordan & Mitchell, 2015). Although machine learning models have many capabilities for predicting daylighting within buildings, the use of machine learning in daylighting studies is relatively new (Ngarambe et al., 2020).

ML algorithms that vary depending on the type of training data, the order, and the method, mainly include supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, transductive inference, on-line learning, and active learning (Rostamizadeh & Talwalkar, n.d.). Supervised learning is one of the most used learning types in ML models. It requires a set of labeled data to train the model. The most common problem types in ML models, including regression, classification, and ranking, are associated with supervised learning. In classification problems, outputs are specific categories predicted from inputs, whereas regression models give output as real-valued labels (Rostamizadeh & Talwalkar, n.d.). Studies related to daylight predictions in literature can be divided into two based on task: regression and classification problems. Regression is used when the desired result consists of one or more continuous variables (M. Jordan et al., n.d.). In daylight studies, regression tasks are mostly preferred for predicting illuminance values, DF, DA, sDA, cDA, ASE, and DGP. For regression problems, ANN (Artificial Neural Network), DT (Decision Tree), SVM (Support Vector Machine), MLR (Multilinear Regression), and RF (Random Forest) algorithms are mostly chosen. The most preferred model is ANN. The ability of ANNs to learn complex relationships between features has made them the primary choice in complex problems in daylighting prediction.

Studies that used ANN evaluated their models on four performance metrics mentioned in the following chapters: R^2, MAE, MSE, and RMSE. Based on performance metrics, hyperparameter tuning is applied in some studies by changing the number of hidden layers, the number of neurons, epoch size, learning rate, and activation functions. Table 2.4 shows some of the analyzed works. Among the studies reviewed, R^2 reached a maximum of 0.99 (Ngarambe et al., 2020). However, this study only estimates the hourly illuminance values per room scale grid point. It can not predict illuminances by considering urban form parameters and spaces of different sizes and orientations. In studies where more inputs such as urban form, location, and building properties are given to the model, the problem becomes more complicated, and as a result, performance values may decrease (Ayoub, 2019b; Han et al., 2021). Therefore, studies also can be analyzed in terms of their resolution. The studies are handled at grid or space resolution. Although the grid resolution works (Kazanasmaz et al., 2009; Ngarambe et al., 2020) are in high resolution, their usability on an urban scale remains limited. Most of the analyzed studies (Lorenz & Jabi, 2017; Zhou & Liu, 2015) in the literature are on a spatial scale. However, many of them ignore urban form parameters. In the studies examined, the estimation performance of the study model, which is closest to the urban scale and considering the external obstruction form parameters, varies between $R^2 = 0.87$ and 0.9 (Ayoub, 2019b). However, in this study, although the relationship with the surrounding buildings has changed, the examined zone has always remained constant; that is, the analysis of multi-dwelling buildings on an urban scale has not been addressed. Table 2.5 shows the inputs used in daylighting prediction studies. Urban form parameters include obstruction width, height, angle, material reflectance, and distance from obstruction. However, Sky Exposure (SE), the visible sky ratio that includes the surrounding buildings, is not included in studies. It is an indicator of light availability in dense urban areas (J. Zhang et al., 2012); therefore, it can be used as an urban form parameter in daylighting prediction studies.

Most of the studies in the literature focus on public or commercial buildings' illumination instead of residential buildings. According to (Dogan & Park, 2019) study findings, 73% of the studies on climate-based daylighting metrics analyzed offices, while only 27% focused on residentials. Therefore, there is a need to increase the daylighting prediction studies and methods of residential buildings in the literature.

Studies also can be grouped into two in terms of data collection: field measurements and simulation-derived data. Field measurements take much more time than simulation to collect the necessary data. At the same time, it is not possible to quickly reproduce field measurements or record new values obtained with changing design parameters. Data generation with simulation is produced faster than field measurements. However, the accuracy of the data produced here also depends on the modeling detail and the expertise of the modeler. Daylighting prediction studies also can be evaluated in terms of their temporal resolution. There can be predicted for selected time series or annual/hourly basis.

Table 2.4 Some of the studies using ML models to predict daylighting performance metrics

Reference	**Problem**	**Model**	**Building Type**	**Output**	**Performance**
(Le-Thanh et al., 2022)	Regression	ANN	NS	UDI	R^2= 0.89* & 0.78**
(Han et al., 2021)	Regression	ANN	office	UDI & DA	R^2= 0.98 & 0.96 MAE= 1.58 & 1.37 MAPE=2.10% & 2.36%
(Ngarambe et al., 2020)	Regression	DNN RF GB	NS	Illuminance	R^2 = 0.99 & 0.92
(Ayoub, 2019b)	Regression	MLR	residential	sDA & ASE	R^2= 0.87 & 0.98 MSE = 5.28 & 8.13%
(Lorenz & Jabi, 2017)	Regression	ANN	office	DA	MSE = 0.0005
(Zhou & Liu, 2015)	Regression Classification	ANN SVM	NS	Illuminance UDI	Accuracy: 96.35* & 62.15**
(Kazanasmaz et al., 2009)	Regression	ANN	office	Illuminance	IPE = 2.17%
(Kurian et al., 2006)	Regression	ANN	NS	Illuminance	RMSE = 0.13–1.12%

* indicates the highest performance result in the study

** indicates the lowest performance result in the study

Table 2.5 Input parameters used in daylighting prediction models

Climatic Condition	Temporal Settings	External Obstruction	Building Physical Features	Openings & Shading
Global horizontal radiation	Time of the day	Obstruction width	Width	WWR
Direct normal radiation	Day of the year	Obstruction height	Length	Window height
Diffuse horizontal radiation	Solar Altitude	Obstruction length	Height	Window width
UV Index	Solar Azimuth	Obstruction angle	Orientation	Orientation of the window
UV dose	Solar Declination	Obstruction material reflectance	Sensor point identification	Number of windows
Exterior Horizontal Beam Illuminance	Solar Hour	Distance from obstruction	Distance from windows	Window sill height
Horizontal Infrared Radiation Intensity			Reflectances of surface materials	Transmittance value (VT)
			Work plane height	Solar Heat Gain Coefficient (SHGC)
			Width/depth ratio	Shading devices parameters
			Area	
			Indication of floor	

2.7 Electric Lighting Energy Consumption Calculations

Most of the energy used worldwide and the CO_2 emissions produced from this energy are attributable to buildings. The quantity of energy used by buildings in America and Europe accounts for 40% of all energy utilized, and the amount of CO_2 emissions resulting from this usage accounts for about 38% of all emissions (Zhao & Magoulès, 2012). Therefore, buildings have a significant potential to reduce global energy consumption (Amasyali & El-Gohary, 2016). This leads to the prediction of building energy consumption gaining importance for improved decision-making for reducing energy demand and associated CO_2 emissions (Amasyali & El-Gohary, 2018).

Therefore, forecasting building energy use has recently received much attention. Several attempts are used to calculate building energy consumption. However, studies that estimate specifically electric lighting consumption is a few (Amasyali & El-Gohary, 2016). According to the Amasyali & El-Gohary review study, only 2% of the analyzed studies reviewed predicted lighting energy consumption, and most of the work that predicts lighting energy consumption is for non-residential buildings (Amasyali & El-Gohary, 2018). There is a substantial lack of development of prediction models for lighting energy consumption based on daylighting in residential buildings in the literature.

There are two main approaches for calculating building lighting energy consumption: physical modeling and a data-driven approach (Amasyali & El-Gohary, 2018). Physical modeling is based on detailed modeling and analysis, which can be done using simulations. Different simulation tools can calculate artificial lighting energy consumption in buildings. UMI (Urban Modelling Interface), proposed by Reinhart et al. (2013), is an urban simulation interface that can calculate daylighting illuminance and integrate it with energy simulations. The other popular approach is Building Information Modeling (BIM) enables intercorporate lighting schedules and energy simulations and performs detailed analysis (Mahgoub, M.H., 2020). Climate Studio, developed by Solemma LLC, provides a comprehensive analysis of the building's advanced daylighting, artificial lighting, and energy analysis. The integrated use of RADIANCE and DAYSIM, and EnergyPlus enables the calculation of the amount of electricity used for lighting in buildings, considering the daylight. However, even though these approaches enable users to estimate detailed performance analysis, they can be used by only experienced people. Also, getting extensive results requires a high level of modeling detail and much computational time. Especially considering variable occupant profiles and urban form parameters are computationally very high to calculate lighting energy consumption using simulations.

Contrary to the physical modeling approach, data-driven prediction models learn from available data to predict lighting consumption instead of requiring detailed energy models. Several ML models have been used to predict electric lighting energy consumption, including ANN, SVM, decision trees, and other ML algorithms. ML models make predictions using historical data, if available, measured field data, or data produced by simulations. Electricity data used for illumination is not available hourly, so it is not possible to estimate electric illumination energy consumption based on existing data in hourly resolution. Measurement data made at certain intervals is specific to where it is made. In machine learning models based on data produced by simulation, data can be produced parametrically by considering variable parameters. Studies generally used the simulation data obtained by creating certain simulation setups as a data set in the ML model (Aydinalp et al., n.d.; Wong et al., 2010). However, the simulation data obtained are the lighting energy consumption data provided over the energy models. Even though a few studies take daylighting parameters into account, they ignore the use of artificial light depending on daylighting illuminance levels and the calculation of lighting energy consumption, taking this into account.

In addition to the studies mentioned above, mathematical formulas are used to calculate energy use for artificial lighting in different temporal solutions. The hourly, daily, or annual lighting energy consumption of a place can be calculated in kWh/Wh using the information about the power of artificial lighting used, how many hours it is used, and the illuminated space area. These calculations remain basic without enabling analysis based on daylighting illuminance at an urban scale.

2.8 Gaps in the Literature

Daylighting studies in the literature have been analyzed in many aspects, such as daylighting performance metrics, factors that affect daylight illuminances, daylighting calculation methods, different ML models, and features used in prediction studies. Accordingly, the shortcomings identified in the analysis of indoor daylighting illuminance are listed as follows:

- Generally, studies treat the space where daylighting analysis is made as a single zone without considering external parameters. In the dense urban fabric, the daylighting of the buildings is hindered by the surrounding buildings. Surrounding buildings can block the light to the examined building to a certain amount, cause more light to be reflected on themselves, causing more illumination of the examined building, or completely block the daylight. Therefore, indoor daylighting analysis of an urban-scale space can not be handled independently of the surrounding buildings. Even if it is handled, it will lead to erroneous analysis results. Also, climatic parameters are highly influential on daylighting illuminance, so the studied space can not be considered independently of its location. Climate parameters vary depending on the location. The external parameters (external obstruction and climate condition parameters), together with the internal parameters, should be considered during daylighting analysis.
- Most studies deal with the daylighting analysis of non-residential buildings (offices, schools, e.g.). Indoor daylighting illuminance analysis of residential buildings is rarely included. Moreover, the analysis of multi-dwelling buildings on an urban scale has not been addressed. Units on various floors may be lit differently despite having the same physical properties. Because of the daylighting reflected from the surrounding buildings, the illumination

of the interiors may not rise in a correlated way with floors, in this case, complicating the analysis of indoor daylighting. Due to reflected lights, a downstairs area could be more well-lit than an upstairs area. Therefore, methods for analyzing indoor daylighting for multi-dwellings on an urban scale should be developed. The lack of daylighting analysis of residential buildings is also considered a shortcoming in review studies (Ayoub, 2020).

- Most studies that developed ML models considered specific daylighting performance metrics and estimated these metrics with ML models. Although these metrics give meaningful results for the daylighting analysis of the space, they are calculated differently in each, and this difference makes it impossible to compare the studies with each other. Estimating and analyzing absolute illuminance values in high resolution instead of these metrics allows comparative analysis of studies with each other (Ayoub, 2020). Other illuminance-dependent studies are also provided by estimating absolute daylighting illuminance. One can calculate the amount of electricity used during daylight. A technique for this, though, has not yet been covered in the literature. In order to examine electrical lighting usage based on predicted daylighting illuminance, a complete method that uses ML models to forecast indoor daylighting illuminance is required. Without considering daylighting, urban energy consumption estimation models provide inaccurate findings by either underestimating or overestimating electricity demand. Inaccurate calculations of electrical illumination consumption in urban energy models indirectly result in incorrect calculations of heating and cooling energies.

CHAPTER 3

A METHOD FOR THE PREDICTION OF INDOOR DAYLIGHTING ILLUMINANCE ON AN URBAN SCALE

This chapter presents the development of a method for the urban scale prediction of indoor daylighting illuminance. The method can estimate the indoor daylighting illuminances on an urban scale at an hourly solution using ML models. The developed method enables daylighting analysis at hourly resolution considering building and external parameters. The method was developed by exploring different machine learning models, revealing how data-driven models can be used in urban-scale daylight analysis. After the method was developed, possible areas that can be used on the urban scale depending on daylighting were also discussed as a research topic. It has also been explored that the developed method can calculate the lighting electricity consumption of residential buildings on an urban scale depending on daylighting and different occupant profiles. The method is based on four steps that are (i) 3D model development, (ii) simulation-based data generation, (iii) development of prediction models, and (iv) lighting energy use calculations. In Step (0), *3D model development*, the studied area is modeled to be ready for simulation by gathering data from different sources and producing them when there is no available data. In Step (ii), *simulation-based data generation*, hourly illuminance values are obtained from annual simulations. Step (iii), *development of prediction models*, presents how different ML models were trained with simulated data and the comparison results of models' performances. In Step (iv), *electric lighting energy use calculations*, the methodology to calculate electric lighting energy use is described based on predicted illuminance results and different occupant profiles. Figure 3.1 shows the proposed methodology for the development of the method.

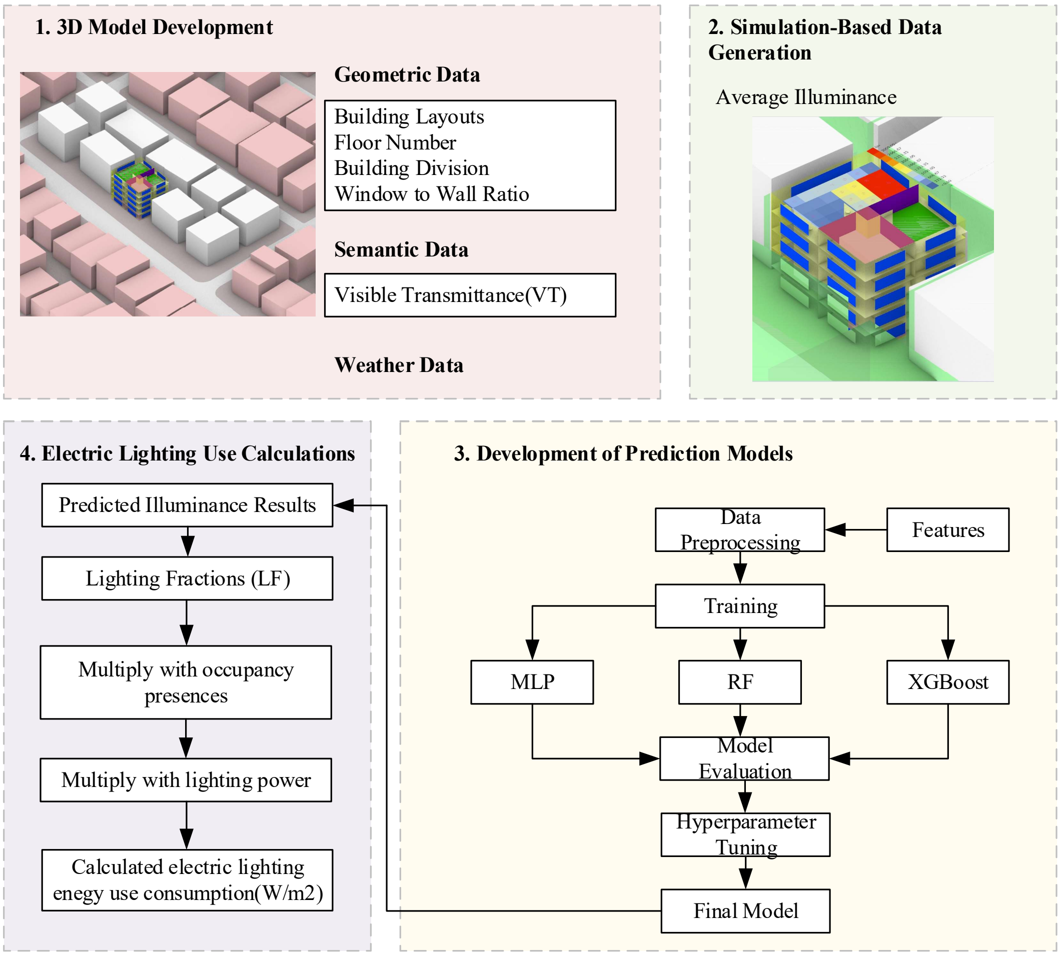


Figure 3.1. An overview of the proposed methodology

3.1 3D Model Development

The data required for 3D modeling is divided into three main groups: geometric, non-geometric, and climate data. Figure 3.2 shows the 3D model development process. These data exist in different formats from different sources. It is aimed to keep the data in a single database to be given to the model.

Geometric data includes building layouts, floor numbers, unit division of buildings, and window/wall ratio information. The building layout information of the Bahçelievler district was obtained from the Turkish Ministry of Environment and Urbanization as a 2D file. Floor numbers were recorded by looking at Google Images of the buildings examined. It is necessary to know how many units are on each floor to divide the buildings into units. Information on how many units are in the buildings was obtained from the Civil Registration and Citizenship Affairs (NVI) website. The total number of units obtained for each building was divided by the number of floors, and the number of flats per floor was calculated. Window/wall ratio information was requested to be obtained from the EPC (Energy Performance Certificate) queried from the BEP-TR page of the Turkish Ministry of Environment and Urbanization. However, there is no EPC for every building. Therefore, the existing EPCs were collected, and the window/wall ratios obtained from them were reconstructed for all buildings by applying the distribution. Distributions are obtained from the project funded by the Scientific and Technological Research Council of Turkey under grant number TUBITAK 120M997.

Semantic data includes only Visible Transmittance (VT) information under this model. General data of VT are available in different ranges for different glass types in the ISICAM catalog. From these ranges, they are reproduced for all buildings by distribution based on the VT values of the glass types suitable for use in residential buildings.

Climate data is taken from Climate One Building (https://climate.onebuilding.org/) with .epw extension. The typical meteorological data of Ankara, Central location covering the years 2004-2018 were taken.

Box plots showing the distribution of input features are given in Appendix A.

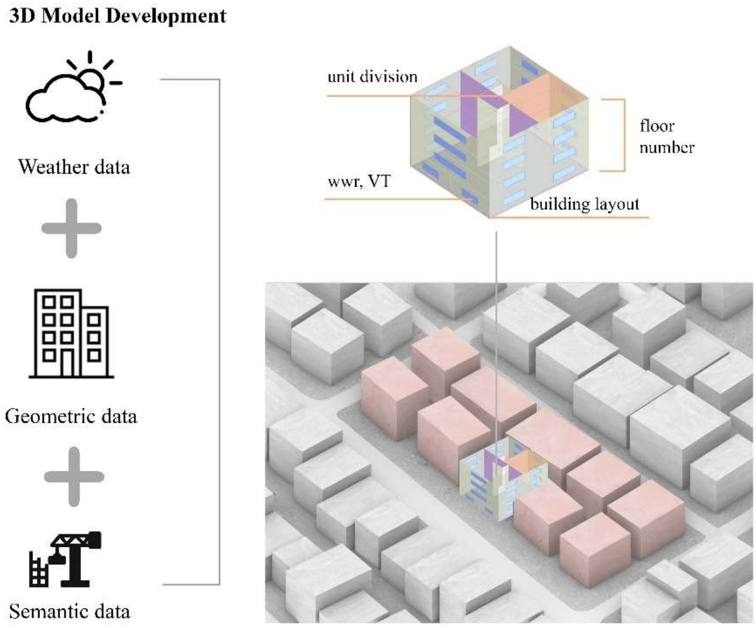


Figure 3.2 3D model development process

3.2 Simulation-Based Data Generation

The average daylighting illuminance data to be used in ML models were produced by simulations. Also, simulations are utilized to calculate one of the input parameters used in ML models: ' Sky Exposure'.

3.2.1 Daylighting Illuminance Calculation

Building simulations are performed using Honeybee plugins of DAYSIM and RADIANCE to obtain daylighting illuminance values. Information regarding the zone to be examined, the surrounding buildings, and the resolution at which the analysis will be conducted (grid, annual, etc.) must be provided to calculate the daylighting illuminance values. In order to enhance the accuracy of the results, various settings of the simulation can be edited. In this study, RADIANCE settings were set up for this purpose. Mainly, there are five settings for radiance parameters: ab (ambient bounces), aa (ambient accuracy), ar (ambient resolution), ad (ambient divisions), and as (ambient super-samples). These values are described by RADIANCE developed by Lawrence Berkeley National Lab for four conditions: minimum values, fast simulations, accurate results, and maximum values that input can be taken. For the scope of the study, parameters are chosen for accurate results. Table 3.1 shows these parameters and their values.

Table 3.1 Radiance settings used in the study

	Accurate
Ambient Bounces(ab)	2
Ambient Accuracy(aa)	0.15
Ambient Resolution(ar)	128
Ambient Divisions(ad)	512
Ambient Super-Samples(as)	256

Moreover, weather files and analyzed test points must be given to the model. The weather file is given as input in .epw format. Test points should be generated for each simulated zone. In the study, ten buildings in a parcel of Bahçelievler, Ankara, a total of 119 units are simulated. Since daylighting simulations are computationally expensive, each unit is considered a single zone without room division. Each zone is divided into 2 m grids. These grid points are elevated on the z-axis of 0.762 m, which is generally accepted as the height of the work plane area in the literature. The illuminance values for each point were calculated hourly throughout the year. The average value of the obtained values is taken for each zone and is given to the ANN model to be trained.

3.2.2 Sky Exposure Calculation

The effect of the surrounding buildings on the examined building at the urban scale was obtained by calculating the Sky Exposure (SE). This calculation examines the visible sky ratio by including the surrounding buildings as context from the center point of the windows located in four directions in the zones. A Sky Exposure of 100 indicates that any building opposite does not block the zone. As the effect of the context buildings increases, the sky exposure, that is, the rate of visible sky decreases. This parameter is calculated in four directions for each zone and given as input to the ML models. Figure 3.3 diagrammatically shows the calculation method of SE.

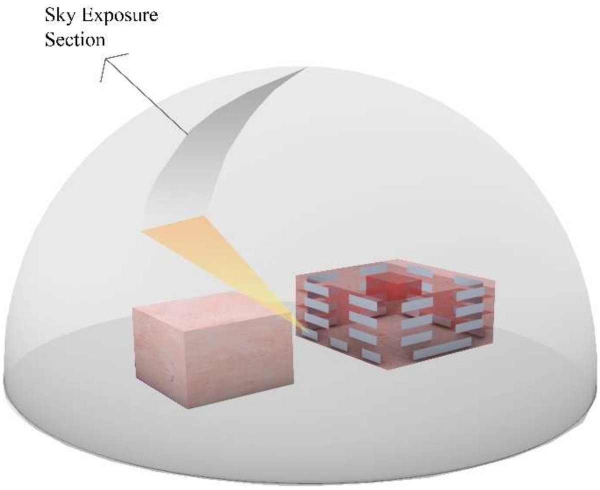


Figure 3.3 Sky exposure calculation

For the reader's information, in the study, it is assumed that surrounding buildings are made of the same material; as a result, their reflectivity is the same. Reflectivity characteristics should be considered if it is wanted to compare material variations.

3.3 Development of Prediction Models

Machine learning models require a set of predictor and target variables. Within the scope of this thesis, the regression models are used to predict the average illuminance value (lux). Climate conditions and external obstruction parameters were selected external parameters. Building physical and opening features were used as internal parameters. The inputs were selected after studying related inputs examined in the literature to forecast average illuminance accurately. A summary of the input variables and descriptions is given in Table 3.2.

Table 3.2 Summary of Input Variables and Descriptions

	Predictor Group	Description	Input Name	Unit
Internal	Building Physical Features	Indication of floor	z-axis	-
		The ratio between the width and length of the units	Aspect ratio	-
		The ratio between the total window area and total floor area of units	TotalWindow/ TotalFloorArea	-
	Opening Features	WWR in 4 directions	wwrNorth wwrWest wwrSouth wwrEast	%
External	External Obstruction	Sky Exposure in Four directions	SENorth SEWest SESouth SEEast	%
	Climate Condition	Global Horizontal Illuminance	Global Horizontal Illuminance	lux

Inputs required for the ML models are WWR in four directions, information on the unit's floor, aspect ratio (width/length ratio), sky exposure calculation in four directions, visible transmittance (VT), and global horizontal illuminance. The information of all values except Global Horizontal Illuminance is taken from the created 3D model. Hourly GHI data is obtained from the weather file with the .epw extension. A database is created by saving these data on Google Colab. After the database is created, data is given to the ML models. The flow of the development of ML learning models is also given in Figure 3.4.

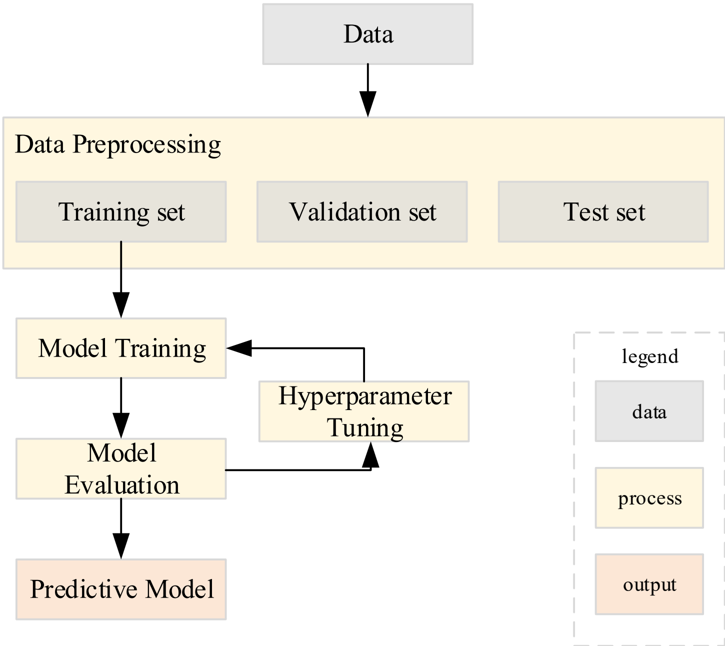


Figure 3.4. The flow of the development of ML models

3.3.1 Data Preprocessing

Preprocessing converts the original data into a format the prediction model can use. There are several data processing techniques, including handling null values of the dataset, data reduction, scaling the data, and splitting the dataset. In this study, the obtained dataset was checked for null values , and no null values were found. The input features discussed in this study are the parameters known to be effective on daylighting illuminance in the literature. Therefore, no data reduction was made in the selected parameters, and it was desired to train the estimation model using all parameters. Since each input given to the model is in very different value ranges from each other, giving these values to the model as it is will cause the model to show biased results. All data were scaled to the range of [0, 1] using the Min-Max Scaler. After the data set was scaled, it was divided into three subgroups to make it ready for the model.

ML models require three main subsets of original datasets: train data, test data, and validation data. A training set refers to examples used for model learning (D. Ripley, 2008). A validation set is a set of examples used to tune the parameters of a model. The test set is a set of examples used only for the unbiased evaluation of the final model. In the study, the data is split into 70% and 30% as train and test data. 50% of the test data is reserved as validation data.

3.3.2 Model Training

At the scope of the thesis, the ML-based design method that can analyze the hourly indoor daylighting illuminances and calculate the electric lighting use based on these illuminances is developed. For the development of the method, three different ML models are utilized: MultiLayerPerceptron (MLP) (Anil K. & Jianchang, 1996), Random Forest (RF) (Breiman, 2001), and XGBoost (XGB) (Chen & Guestrin, 2016). Models are developed and trained with the same data. The structures of the models are shown in Figure 3.5.

Artificial Neural Networks (ANNs)

ANN involves artificial neurons to identify and store information, which resembles the way neurons in the brain work (Das & Roy, 2019). Although its structure resembles that of neurological neurons in the brain, ANN has only one type of connection connecting one neuron to another. ANN consists of input and output layers and hidden layers of neurons consisting of units that convert input into output. Each neuron has different weights, and its influence on the output layer depends on their weights. The neurons can receive input signals and send them to other layers by processing the weights and biases (Z. Zhang, 2018). Activation functions determine the output of given inputs. One of the most widely used ANN designs is the Multilayer Perceptron (MLP).

Random Forest (RF)

An ensemble method called random forest combines the predictions of various tree predictors in a forest (Breiman, 2001). Each tree in the forest depends on the values of a random vector sampled individually. The method is known as 'random forests' as the ensemble's fundamental components are tree-structured predictors built with randomness (Segal, 2003). The predictions are based on combining multiple trees instead of a single tree. Several trees in the forest are combined for predictions, and the estimations are provided by averaging the results of each tree. Estimating by averaging the prediction results of several trees in the forest, instead of a single tree, results in more reliable results when compared to single tree methods, considering the diversity.

eXtreme Gradient Boosting (XGBoost)

A tree-boosting method called XGBoost constructs new models regularly depending on estimated errors and then merges them into an ensemble model (Chen & Guestrin, 2016). XGBoost is scalable, unlike conventional gradient boosting techniques, which help prevent the model from overfitting. XGBoost combines all scaled predictions from all trees by learning errors from previous trees. One of the biggest achievements of the technique is parallel and distributed computing which speeds up learning and makes model exploration possible more quickly. XGBoost can be utilized for both regression and classification problems.

The generated dataset was utilized for training three different ML models. The architectures of models vary because of the chosen parameters. In Section 3.3.4, it will be thoroughly described what each parameter utilized in each model represents. Initially, the MLP model included four layers. Each layer receives a ReLU (Sharma et al., 2020) activation function. Using the ReLU required scaling the inputs between 0 and 1. Adam was chosen with a learning rate of 0.001. The tuning of these parameters follows, as will be covered in Section 3.3.4 in more detail. The models were directly implemented using the sklearn library without specifying the RF and XGBoost parameters. Then, these models' parameters were tuned.

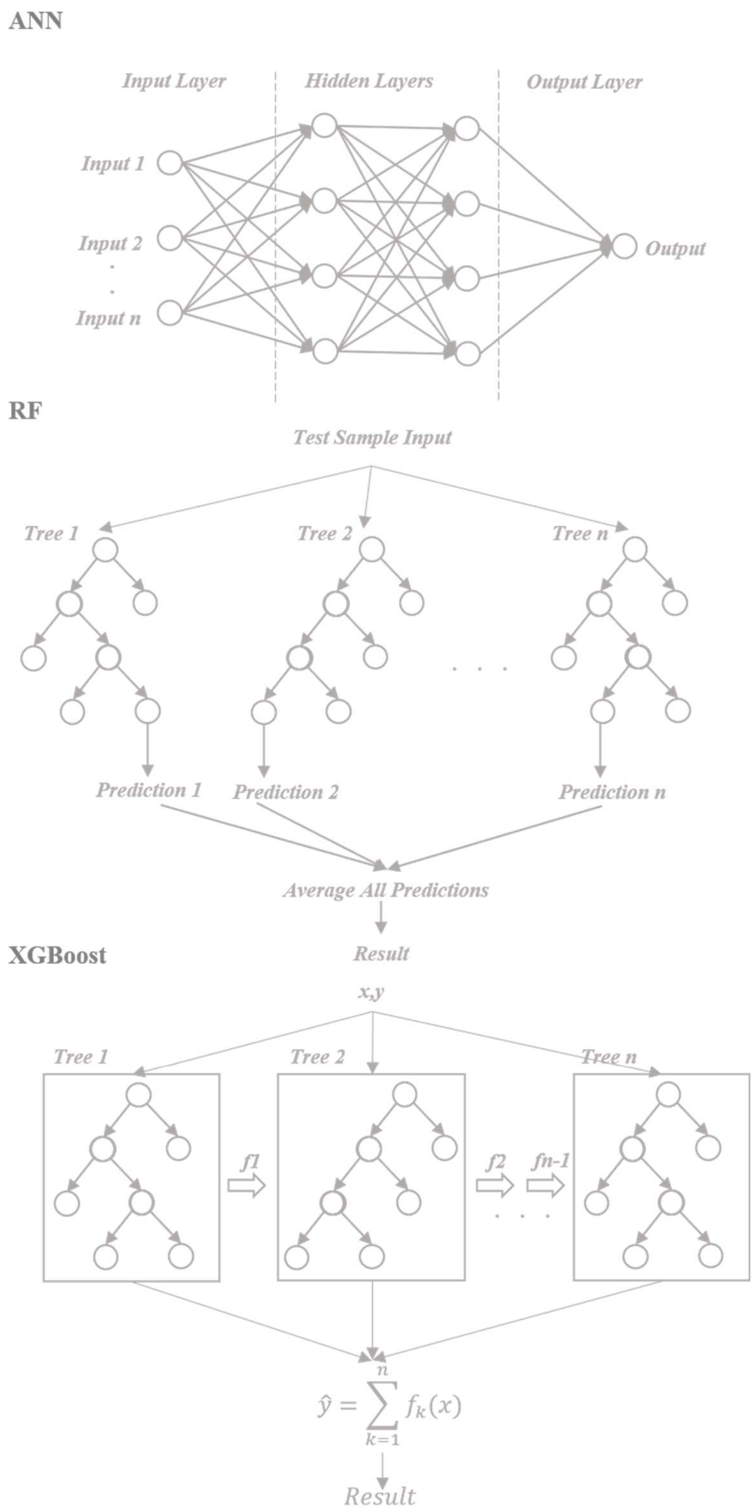


Figure 3.5 The structures of the ML models

3.3.3 Model Evaluation

In order to evaluate the performances of the ML models, four different performance metrics are calculated: Mean Absolute Error (MAE) (1), Mean Squared Error (MSE) (2), Root Mean Squared Error (RMSE) (3), and Coefficient of Determination (R2) (4). Their equations are indicated below.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}| \tag{1}$$

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2 \tag{2}$$

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2} \tag{3}$$

$$R^2 = 1 - \frac{\sum(y_i - \hat{y})^2}{\sum(y_i - \hat{y})^2} \tag{4}$$

where $\hat{y}$ represents the actual value and y_i is the predicted value.

After the ML models were trained, their performance was compared. With the highest R^2 score (0.733), the best XGBoost model performed the estimation. Then came the RF model with 0.713 and the MLP model with 0.669. Although the performance values of the models are in the range of the performance values of the studies examined in the literature, the performance of the models has remained below many studies. The graphs (Figure 3.6) of the actual and predicted values calculated by the models were analyzed, and it was seen that the MLP and XGBoost models could not predict values above a certain value (3000 lux). After re-examining the original dataset, it was determined that the non-uniform distribution of the data may lead the ML models not be able to predict values above a certain value.

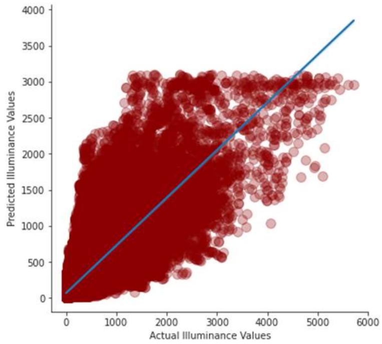
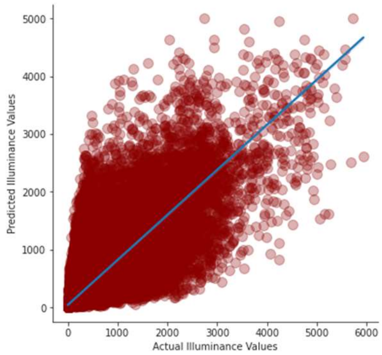
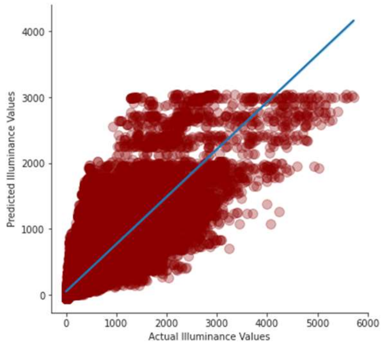
Model	
MLP $R^2 = 0.669$	
RF $R^2 = 0.713$	
XGBoost $R^2 = 0.733$	

Figure 3.6 The graphs of the actual and predicted illuminance values

It has been analyzed that almost all of the outputs in the original data set are between 0-2000 lux.

Data augmentation was implemented to increase the performances of the ML models by enhancing the diversity of the value range of data seen by ML models. Data augmentation is a method to increase the amount of data by adding newly created synthetic data from existing data. Data augmentation is proposed to increase the diversity in the data set and expand the distribution ranges so that the models can better learn the values at the extremes. For data augmentation, the new data was reproduced with simulations. The produced data are not representative of the current situation. The data representing the current situation was named the V1 (base) scenario. The V1 scenario represents the data created in Section 3.2.1. New scenarios have been produced to expand the distribution ranges of the data. Table 3.3 shows a total of four different scenarios (V1, V2, V3, V4) created and the inputs that change depending on these scenarios.

Table 3.3 Proposed scenarios for data augmentation

	V1(Base Scenario)	***V2***	***V3***	***V4***
Window-to-Wall Ratio	0- 0.36	0- 0.36	0- 0.36	0- 0.9
Visible Transmittance	0.17- 0.55	0.17- 0.55	0.6- 0.9	0.17- 0.55
Context Shading	Included	The heights of the contexts scaled to 0.5	Included	Included

In the V1 scenario, the WWRs are formed from the distributions obtained from the existing EPCs, and the windows in four directions take values between 0 and 0.36. VT is again created by distribution, and values between 0.17-0.55 are taken. In this scenario, context buildings are modeled with real floor heights.

In the V2 scenario, only the floor height parameter of the surrounding buildings is changed so that the machine learning models can better learn the relationship between the SE values and the illuminance. The illuminance values for the analyzed buildings are recalculated by 0.5 scale heights of the surrounding buildings. Other parameters (WWR, VT) are kept constant.

The V3 scenario represents the situation where the VT values are increased. The surrounding buildings are modeled with their real heights, and the WWR includes the values produced for the base scenario in the first place.

Finally, the WWR value ranges were expanded in the V4 scenario. Although there is no window/wall ratio of 0.9 in Bahçelievler, the maximum value has been increased to 0.9 for machine learning models to learn the effect of WWR on illuminance better.

Introducing all the new data generated by simulations to the model would not be suitable for increasing data distribution. Therefore, a new dataset was created by adding only the data in the less represented range (2000 and above) to the first generated data (V1 scenario). Statistical descriptions of all data are given in Appendix B. After the final dataset was created, it was given to the ML models to be trained. Figure 3.7 shows the initial estimates of the models with the generated dataset. With the new dataset produced, the performance of all three models has increased significantly.

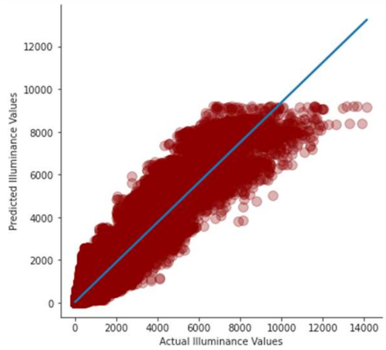
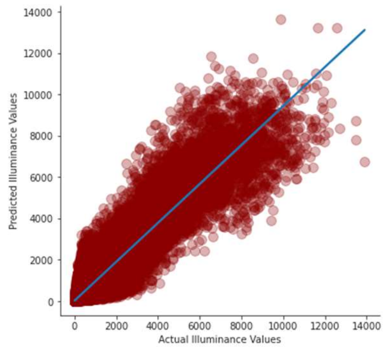
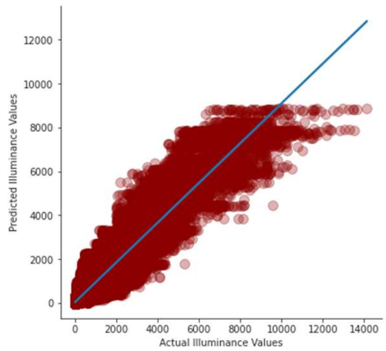
Model	
MLP $R^2 = 0.923$ MAE = 0.008 MSE = 0.001 RMSE = 0.022	
RF $R^2 = 0.916$ MAE = 0.008 MSE = 0.001 RMSE = 0.023	
XGBoost $R^2 = 0.914$ MAE = 0.01 MSE = 0.001 RMSE = 0.023	

Figure 3.7 The graphs of the actual and predicted illuminance values with the generated dataset

3.3.4 Hyperparameter Tuning

There are two types of parameters in ML models (Kuhn & Johnson, n.d.). One is model parameters, which may be set up and updated during learning from data. The second is referred to as ‘hyperparameters’. Hyperparameters are the parameters that adjust the model's performance and vary in different ML models. They should set before training an ML model to increase the model's learning capacity, leading to an increase in model performance. Therefore, hyperparameters are used in different configurations to design optimal model architecture, and this is called 'hyperparameter tuning' (L. Yang & Shami, 2020). Proper hyperparameter tuning should be adjusted to increase the model's performance and avoid problems, including underfitting or overfitting. The selected ML models and tuned hyperparameters are given in Appendix C.

Different hyperparameter configurations are used for different ML models. For MLP models, *batch_size*, *epoch_size, learning_rate,* and *number_of_neurons* are used as hyperparameters, and their different configurations are evaluated. The *batch_size* refers to how many samples are used in each iteration. The *epoch_size* defines how often the entire training set is seen from the network. The study defines epoch size as constant using the '*early stopping*' method. *Early stopping* is a regularization strategy to prevent overfitting that stops the model training by observing the accuracy gap between training and validation data. It is also useful to reduce the computational time of the training (Goodfellow et al., n.d.). The *learning_rate* modifies the step size at each iteration. A high learning rate can lead the loss function to behave in an undesired divergent way, while a low learning rate can lead the model to progress slowly. Therefore, different options for learning should be tuned properly. Except for these parameters, different hyperparameters are used in MLP models, including *activation functions* and *optimizers*. *Activation functions* are the functions that determine the output of a given node. Several activation functions are used for different ML models, such as *Linear, Sigmoid, Tanh, ReLU, Leaky ReLU, Parametrized ReLU, Exponential Linear Unit, Swish,* and *Softmax*. They can be implemented in input and hidden layers based on the aim of the ML problem and can be tuned to improve the model's performance. At the scope of the study, for the MLP for a regression model, ReLU is used in input and hidden layers. ReLU is rectified linear activation function that gives output as a positive if the input is positive, whereas the output is concluded as 0 if the input is negative (Sharma et al., 2020).

Therefore, the range of input values given to the model should be in [0,1] for ReLU usage. Accordingly, in the study, before being given to the ML model, inputs are scaled to range from [0,1]. *Optimizers* specify how the learning rate or weights of the neuron should alter to minimize losses. Several optimizers are applied in ML models, including *Gradient Descent* (Cauchy, 1847), Stochastic *Gradient Descent* (Robbins & Monro, 1951)*, AdaDelta* (Zeiler, 2012)*,* and *Adam* (Kingma & Ba, 2014). *Adam* is one of the most preferred optimizers as it achieves better results with high speed than other optimizers. At the scope of the study, Adam is chosen as an optimizer for MLP, and its learning rate is tuned to choose the best performance.

For RF, *number_of_estimators*, *max_features*, *max_depth*, *min_samples_split*, *min_samples_leaf*, *bootstrap* are tuned hyperparameters. *number_of_estimators* represents the number of trees in the forest. *max_features* defines the number of features for the best split and can be assigned as '*auto, sqrt, log2, None*'. In the studied RF model, *max_features* is defined as *'auto', representing* the situation where the number of max features equals the number of features. *max_depth* defines the depth of the trees. The parameters determining the minimum number of instances needed to form a split and a leaf node are *min _sample_split* and *min_samples_leaf*. *Bootstrap* determines whether the entire dataset is used as is or with specific bootstrap samples used in building trees.

For XGBoost , *n_estimators, max_depth, min_child_weight, gamma, learning_rate, subsample* and *colsample_bytree* are tuned hyperparameters. *n_estimators* is the number of trees in XGB. *max_depth* represents a tree's maximum depth, limiting how deep each tree can grow. The increase in the parameters leads the model to overfit more likely. *min_child_weight* is the minimum sum of instance weight that a child needs. *Gamma* shows the lowest loss reduction necessary to create a new partition on a tree leaf node. *learning_rate* is the regularization parameter. The *subsample* is the ratio of subsamples in a training sample to grow a tree. The parameter *colsample_bytree* specifies the proportion of features (columns) used to construct each tree.

There are two types of hyperparameter tuning: manual and automatic search (Wu et al., 2019). Manual search requires hyperparameters' manual setting. It can be used if the person who developed the model is experienced enough to predict which parameters will affect the model approximately and how much. On the other hand, automatic search is based on exhaustive searching, including grid and random search. In order to train the model, a grid search can explore all combinations of a given value of hyperparameters. Oppositely, random search combines given hyperparameters randomly, not all of them. Even a random search can not give the best results as it does not try all the given combinations, it has less computational cost than a grid search. The study applies the random search method with the cross-validation method.

Cross-validation is a technique that uses various subsets of the original data rather than just one on each iteration. If only a portion of the original dataset is used, it might not accurately represent the entire dataset, leading to biased results. Consequently, in a random search, 'nested-cross-validation' is used (Figure 3.8). In this method, the original dataset is split into five identical parts, and the number of iterations is set to 50. The methodology produces unbiased outcomes since it considers all data rather than just a specific subset. The models with the highest performance were selected as the final models due to hyperparameter tuning, and their performances were compared. Analysis of the architecture of the ML models and their performance as a result of hyperparameter tuning is given in Section 4.1.

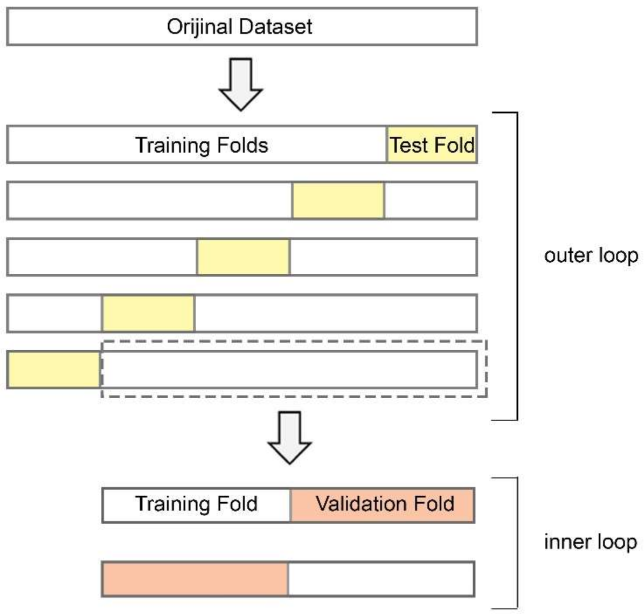


Figure 3.8. Nested cross-validation

3.4 Electric Lighting Energy Consumption Calculations

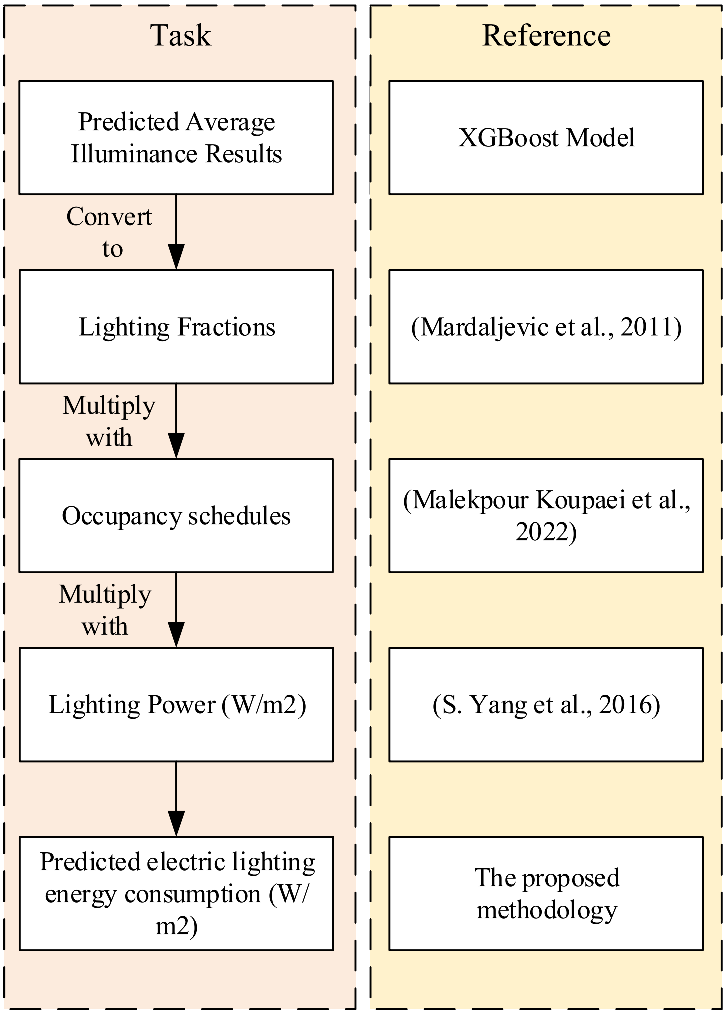


Figure 3.9 The flow of the electric lighting energy consumption calculation

Electric lighting use consumption requires three inputs: lighting fractions obtained from predicted daylighting illuminances, occupant presence schedules, and lighting power density (W/m^2). Figure 3.9 demonstrates the flow of the electric lighting energy consumption calculation.

Lighting fractions at hourly resolution represent the rate at which artificial lighting is turned on, depending on daylighting. The fact that the lighting fraction is equal to zero indicates that the space is sufficiently illuminated; therefore, artificial lighting is unnecessary. The situation where the lighting fraction is equal to one represents that the space is not illuminated at all, and the artificial lighting must be turned on completely. The study conducted a comprehensive literature search to determine how the predicted daylighting illuminance values obtained from the ML model can be converted into lighting fractions to calculate electrical lighting consumption. As a result, an equation has been obtained to create lighting fractions based on daylighting in residential buildings. This equation was obtained by analyzing the relationship between daylighting illuminance and lighting fraction values in the reference source (Mardaljevic et al., 2011). The lighting fraction values depending on the daylighting illuminance values of the reference study examined are given in Table 3.4 and Figure 3.10. The values given by this study are taken from the equation proposed to calculate the electricity consumption of the RT 2005 residential model in France.

Table 3.4 Parameterisation of the lighting fraction in the RT 2005 Model (Mardaljevic et al., 2011)

Daylighting Illuminance(lux)	*Lighting Fraction*
0	1
100	1
200	0.05
2800	0

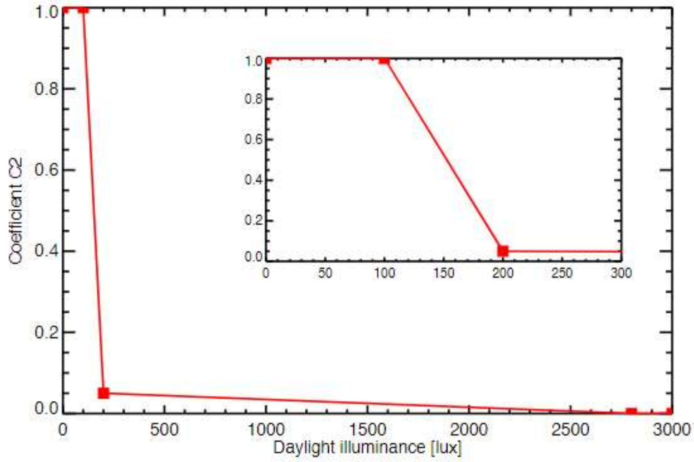


Figure 3.10. Electric light switch-on probability as a function of daylight illuminance (Mardaljevic et al., 2011)

Different equations were created depending on the reference for the four ranges (0-100, 100-200, 200-2800, 2800, and above) determined in the reference, and the equations were converted to lighting fraction values by applying the equations according to the ranges of predicted daylighting illuminance values.

Obtained lighting fractions based on daylighting illuminance are hourly values multiplied hourly by occupant presence. The aim here is to calculate the lighting fraction not only depending on the daylighting but also based on the occupant's presence. Electric lighting use correlates directly to the number of occupants living in the house. Accurately modeling the number of occupants living in the house and their situation of being at home correctly is important for electric lighting use calculations. The discrepancy between the simulated/predicted and actual energy consumption stems from inadequate consideration of occupant behavior in urban building energy models resulting in a performance gap (Happle et al., 2018). There are occupant presence schedules created by Energy Plus to be used in energy models (Figure 3.11). However, these created schedules do not change according to the number of households; they are the same for all households.

Figure 3.11. Standard midrise apartment occupant presence schedule obtained from Energy Plus (Department of Energy)

Creating an occupant presence schedule depending on the number of changing households will increase the diversity of occupants and allow to reflect the current diversity in the calculation of the electric lighting consumption to be calculated accordingly. A literature review was conducted to consider the different occupant profiles, which vary according to the number of households in calculating electric lighting energy use. For residential buildings, a reference (Malekpour Koupaei et al., 2022) was found with occupant presence schedules that change according to the number of households and days (weekdays, Saturdays, Sundays). The occupant presence schedule values were obtained from the graphics of this reference. The graphics produced by the reference are based on the 2019 American Time-Use Survey (ATUS) data. The produced values of the occupant schedules that vary according to the number of different households and days are shown in Figure 3.12. The data in the source used is based on regularly collected American occupancy research data. Due to privacy considerations, it is highly challenging to find occupant data, particularly in residential buildings. For the reader's information, the chosen reference (Malekpour Koupaei et al., 2022), which varies depending on the household sizes and days of the week, is the most varied reference for residential buildings. In future studies, if data representing Bahçelievler for residential buildings can be found, the data can be easily replaced with the data here.

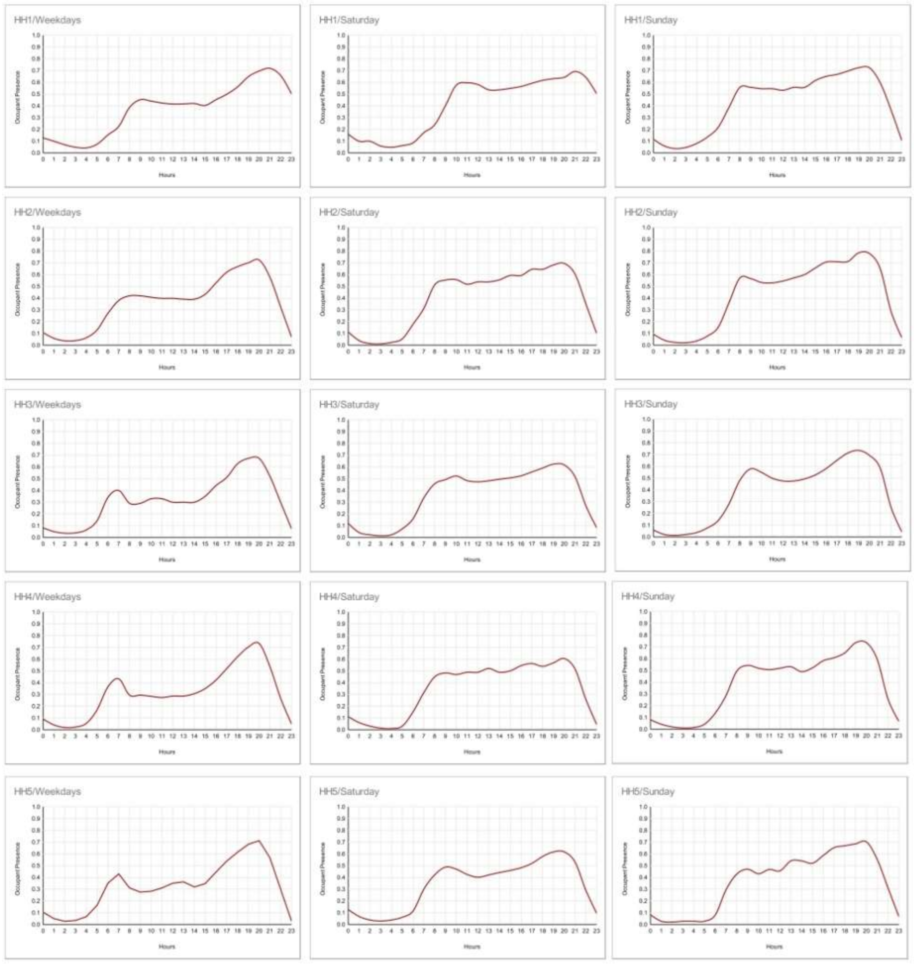


Figure 3.12. Generated occupant schedules based on (Malekpour Koupaei et al., 2022)

The generated occupant schedules are assigned according to the number of households in the units. The household information is taken from TUIK (Turkish Statistical Institute) data. Figure 3.13 shows Turkey's household numbers distribution according to TUIK data.

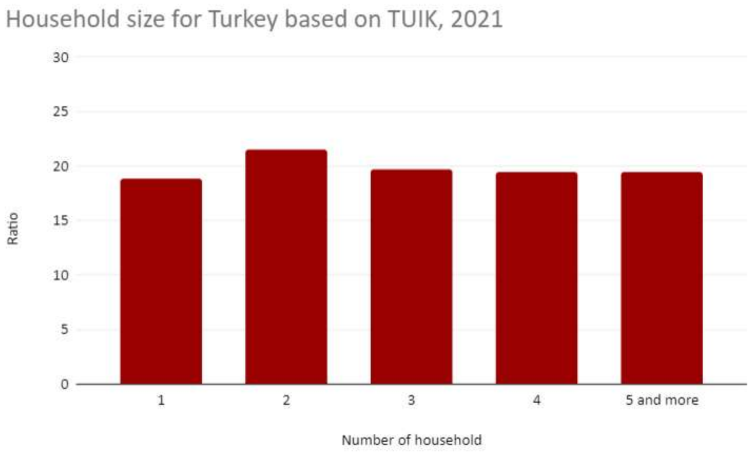


Figure 3.13. The household size for Turkey based on TUIK, 2021 data statistics

Household numbers are assigned to each unit according to their distribution. The occupant presence schedules created above according to the households are assigned to the units whose number of households is known.

Lighting fractions produced based on daylighting were multiplied by the occupant presence schedules. The obtained value should be multiplied by the energy (lighting power) consumed per square meter by the artificial lighting used in the space. These values are assigned as different values for each unit. Lighting power density values were produced uniformly in the range of 12-19 W/m^2, according to a study found in the literature (S. Yang et al., 2016). The scatter plot of the generated values is shown in Figure 3.14.

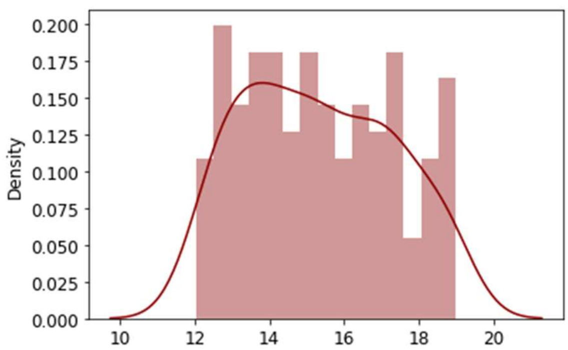


Figure 3.14. The scatter plot of generated values for lighting power density (W/m^2)

U.S. Green Building Council recommended 11.8 W/m^2 for rooms in residential buildings (U.S. Green Building Council). For the reader's information, since the lower limit of the reference selected in the literature (Yang et al., 2016) coincides with the recommended value, the ranges of values are referenced to form a distribution. The selected lighting power densities obtained within the scope of the study and the electric lighting consumption values calculated accordingly are consistent within themselves. However, different types of lamps (especially energy-efficient lamps) were not considered in the study. The lamps selected in the study are lamps with low performance and are not energy efficient. When it is desired to work with lamps with high energy efficiency, the lighting power density of 6.5 W/m^2 recommended by ASHRAE for residential buildings can be taken as a reference (ASHRAE, 2021). When the studies using energy-efficient lamps in the literature are examined, it is observed that the lowest limit for lighting power density value is 2.5 W/m^2 (Ahmed & Asif, 2020).

The equation calculates the electrical lighting per square meter, and multiplying this result by the total square meter of the area gives the total amount of lighting energy used. The results of each unit's electrical lighting energy consumption will be detailed in Chapter 4.

CHAPTER 4

RESULTS

The results of the proposed methodology in Chapter 3 are given in this chapter. Results are reported under two main sections. The first section presents the architecture and comparative performance results of the selected models as a result of hyperparameter tuning in Section 3.3.4. The impact of each input on predicting the daylighting illuminance of the models is also discussed here. The second section reports the electrical lighting consumption data calculated in the first section based on the method proposed in Section 3.4. The thesis aims to observe the calculated electrical lighting consumption by considering the effect of surrounding buildings (urban form parameters). Therefore, in this section, the proposed methodology was applied to the different datasets in which context buildings are excluded from the data, and electrical lighting energy use consumption was calculated accordingly. The electrical lighting energy use consumption data was calculated with two different scenarios, with and without the surrounding buildings, and analyzed comparatively.

4.1 Performances of Utilized ML Models for Daylight Illuminance Prediction

Three different ML models were developed to solve the regression problem defined in the study (daylighting illuminance prediction), and their parameters were tuned to increase their performance. Table 4.1 shows the final parameters of the three utilized ML models.

Table 4.1 The descriptions of the final predictive models

Model	Description
MLP	Neural Networks MLP Regressor Hidden_layer_sizes = [64,64] learning_rate = [0.002] batch_size = [1024]
RF	Ensemble Models Random Forest Regressor n_estimators = [50] max_depth = [54] min_samples_leaf = [1] min_samples_split = [20]
XGBoost	Ensemble Models XGBoost Regressor n_estimators = [300] max_depth = [5] min_child_weight = [1] gamma = [0.4], learning_rate = [0.8] subsample = [0.8] colsample_bytree: [0.8]

Different ML models were tuned with different hyperparameters. Figure 4.1 shows the actual and predicted illuminance values using tuned models.

The final MLP model contains two hidden layers with 64 neurons, a 'ReLu' activation function in each layer, and an 'Adam' optimizer. Compared to the initial model before tuning, the R^2 values increased from 0.923 to 0.9236, and the error values remained almost the same. MLP model shows higher accuracy and lower error at lower learning rates in the given hyperparameter range. However, the MLP model has shown the least increase in model performance after hyperparameter tuning.

The final RF model showed better performance compared with MLP. R^2 of the RF model increased from 0.916 to 0.928. While the MAE value remained the same, MSE and RMSE values decreased in the tuned model.

The highest increase in performance was observed in XGBoost due to hyperparameter tuning. The R^2 value for XGBoost increased from 0.914 to 0.942. Among the three models, XGB had the highest R^2. Therefore, XGB was used when calculating the electrical lighting consumption based on daylight in Section 4.2.

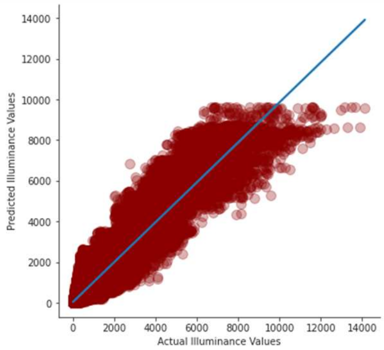
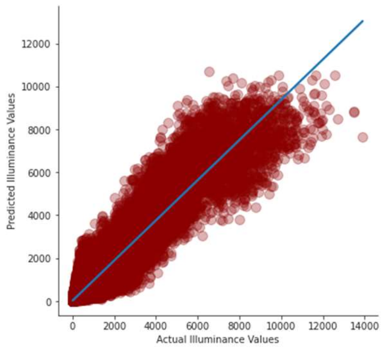
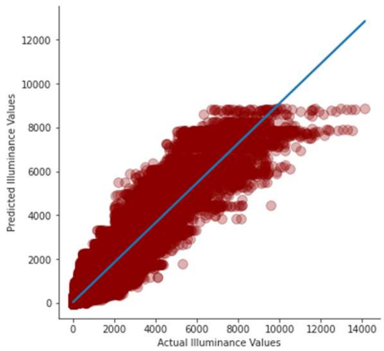
Model	
MLP $R^2 = 0.9236$ MAE = 0.009 MSE = 0.0004 RMSE = 0.022	
RF $R^2 = 0.928$ MAE = 0.008 MSE = 0.0004 RMSE = 0.021	
XGBoost $R^2 = 0.942$ MAE = 0.01 MSE = 0.001 RMSE = 0.022	

Figure 4.1 The performances of hyperparameter-tuned models

4.1.1 Discussion

In the literature, studies that developed an ML model that predicts the illuminance value were examined, and the results of the proposed method were compared with these studies. Using different datasets and models makes a one-to-one comparison of results impossible, but results based on studies can be interpreted. In the studies examined, the most successful study in estimating the illuminance value was the Deep Neural Network, and the R^2 value of the model was 0.99 (Ngarambe et al., 2020). This value indicates a very high forecasting performance. When the input parameters of the study were examined, it was observed that the external obstruction form parameters were not included, while the parameters of the climatic conditions were given in detail. In this case, the effect of surrounding buildings on daylighting is ignored. When the effect of the surrounding buildings is ignored, the model's performance is expected to be higher. When the surrounding buildings are not included, the illuminance values are expected to increase regularly as the upper floors are climbed, and the south façade receives more light directly. However, considering the surrounding buildings, each apartment in the building may illuminate differently, unexpectedly, due to reflections in the surrounding buildings. Therefore, considering surrounding buildings is a much more complex problem for the machine learning model to learn. In the study within the scope of the thesis, the $R^2 = 0.942$ value of the prediction model is within the scope of a high accuracy value, considering both external (climatic conditions, external obstruction parameters) and internal parameters.

In the study, the uneven distribution of the dataset results in poorer prediction performances of ML models. As mentioned in Section 3.3.4, the fact that the data concentrated in a certain range and became representative of the whole data set led the model to predict values that are not in this value range with lower performance. It has also been stated in studies in the literature that data augmentation improves the performances of ML models (Olu-Ajayi et al., 2022). Therefore, data augmentation was applied in this study. For data augmentation, new scenarios are proposed to vary the data distribution. The purpose of the suggested novel scenarios is to broaden the distribution of inputs in the data set and, by exposing the ML model to more data, to improve prediction accuracy. It has been observed that the performance of the models, which are retrained with the data, augmented with the proposed scenarios, has increased significantly compared to the initial situation. By obtaining the R^2 =0.94, no further data augmentation was made in the study. However, when actual and predicted values are compared, it is observed that ML models still predict data above a certain value (10000 lux) with lower performance even at high R^2. As 10000 lux does not define a threshold when calculating the electrical lighting consumption based on daylighting, this did not cause any issues with operation. The study's suggested data augmentation strategy can be implemented in future studies if the ML model is intended to perform better on this value. Data can be generated using various scenarios to get lux values of more than 10000 lux, and the ML model can be trained once more.

4.1.2 Correlations Between Input Features and Outputs

The correlation matrix analyzes the relationships between each feature, enabling the analysis of the relation between inputs and output features. The study analyzed the relationship between average illuminance and each input. Figure 4.2 shows the correlation matrix of the data.

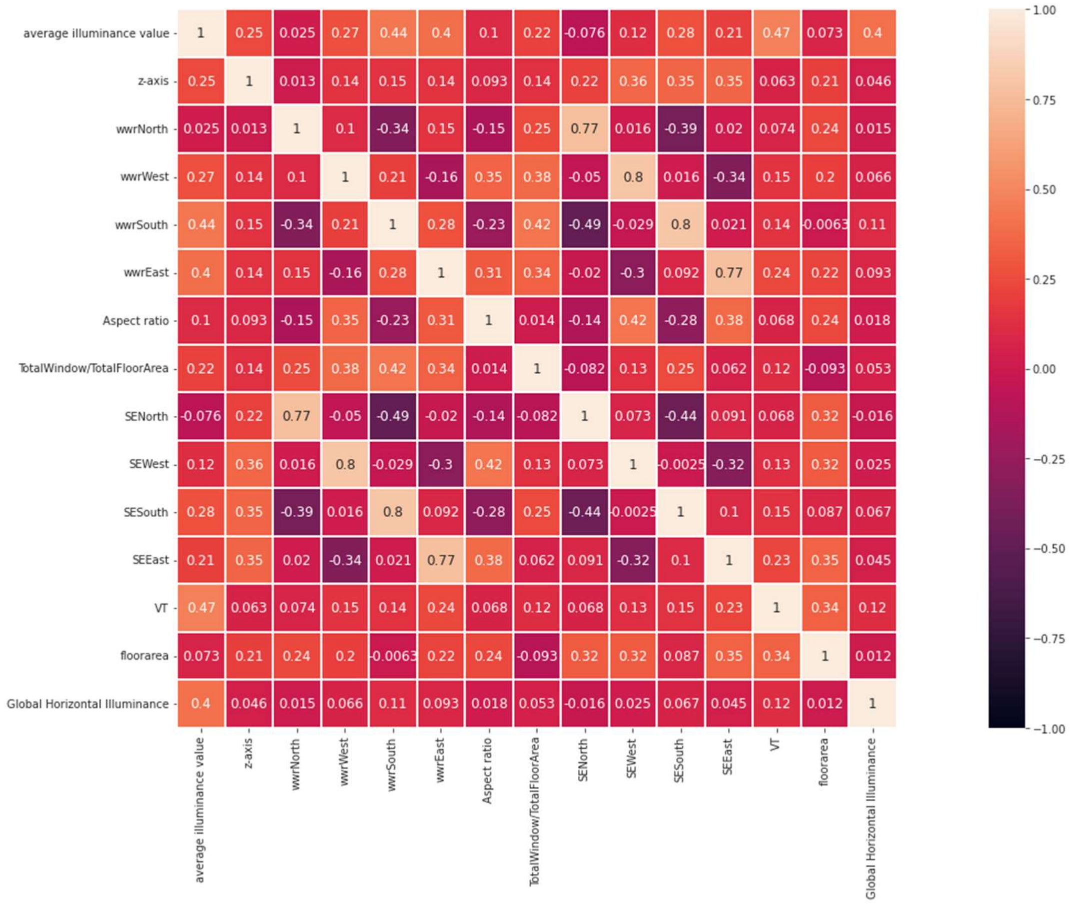


Figure 4.2 Correlation matrix to analyze relationships between inputs and output

It was observed that VT is the most efficient parameter in predicting daylighting illuminance. An increase in VT leads to an increase in the daylight penetrating the space, and a decrease results in the space being less illuminated by daylight. There is a positive correlation between VT and daylighting illuminance. The importance value is expected to be high in predicting the daylighting illuminance value.

The wwrSouth and wwrEast were the parameters most correlated with illuminance after VT. Considering that there are a high amount of sunlight coming from the south, east and west façades throughout the year, it is expected that the increase in the ratio of windows on these façades will directly and effectively affect the illuminance.

Internal parameters were the three most influential parameters in estimating illuminance, as classified in Section 2.4. After these parameters, the most effective parameter in estimating the average illuminance is the global horizontal illuminance, chosen as one of the external parameters. GHI is the amount of light that falls on the horizontal plane, parallel to the ground, hourly throughout the year, so it is expected to be highly correlated with the average illuminance.

SESouth was one of the external parameters positively correlated with illuminance. SESouth calculates the amount of sky visible through windows, considering the influence of surrounding buildings on the south façade. As the SE decreases, the building remains under the shading effect of the surrounding buildings more, and accordingly, the average illuminance value will decrease. The increase in SE indicates that the visible sky will increase, and the average illuminance value will increase accordingly. The fact that the SE on the south side is more effective than the SE on the other directions can be explained by the high amount of daylight coming from the south throughout the year.

Although the z-axis was positively correlated with illuminance, it was not correlated as highly as expected. Section 2.8 also covers this situation. The amount of illumination does not necessarily increase on the upper floors, according to an analysis of the units' daylighting illuminance values at the urban scale. A downstairs unit may receive more illumination from nearby buildings than an above flat. So, when analyzing units at the urban level, it is important to consider the surrounding environment (such as the effect of the surrounding buildings and the floor on which it is located).

Apart from these, the relationship of all input features with output is given in the correlation matrix. The fact that the wwr and SE values on the north façade have the least relationship with the average illuminance value is significant since there is no sufficient daylight from the north side throughout the year.

4.2 Predicted Electric Lighting Energy Consumptions

Electric lighting energy consumption for each zone was calculated following the model development. As stated in Section 3.4, the process consists of three steps: converting the predicted daylighting illuminances to lighting fractions, multiplying the obtained lighting fractions with the created occupancy presence values, and multiplying the resulting value with the lighting power density. The value obtained from these processes gives the hourly electrical lighting consumption in Wh/m^2. The annual electrical lighting energy consumption per square meter is calculated by summing the hourly values throughout the year and converting them to kWh. In order to calculate the electrical lighting energy used for the total illumination of the space, the consumption data obtained in kWh/m^2 is multiplied by the total square meter of the space. Electric lighting energy consumption was calculated with the proposed method for all zones in the original dataset used in the ML model. Figure 4.3 demonstrates the predicted electrical lighting consumption distribution, and Figure 4.4 shows the predicted annual electrical lighting consumption calculated for each zone in kWh/m^2 and kWh.

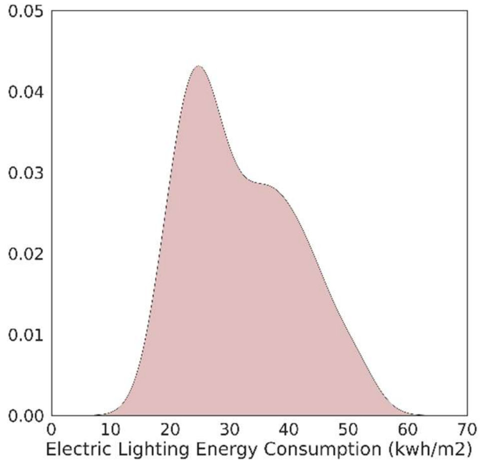


Figure 4.3 Predicted annual electric lighting energy consumption (kWh/m^2)

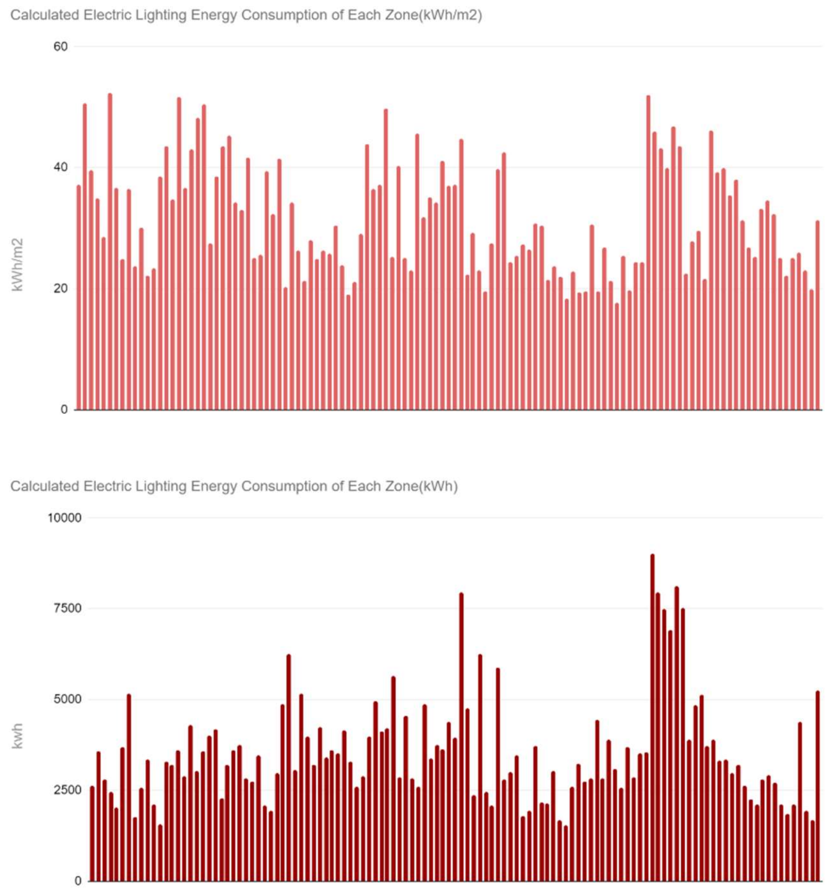


Figure 4.4 The predicted annual electrical lighting consumption for each zone in kWh/m^2 and kWh

Electric lighting consumption differed according to the units' illumination amount, lighting power density, and occupancy types. Table 4.2 shows the distribution statistics of predicted electric lighting energy consumption. Among the units analyzed, the average electrical lighting energy usage per square meter was 31.82 kWh. The lowest electrical lighting consumption per m2 per year is 17.7 kWh/m^2, and the highest electrical lighting consumption is 52.28 kWh/m^2, according to the examination of the predicted lighting energy consumption results based on daylighting, occupancy, and lighting power densities. When the units with the least electrical lighting consumption were examined, it was analyzed that they received light from the south façade and generally had windows on three façades. The units with the highest electrical lighting consumption were those located on the ground and first floors and those that did not receive light from the south façade. Sufficient daylight coming from the south façade throughout the year has increased the illuminance values, and accordingly, the artificial lighting consumption of the units on the south façade due to daylighting has decreased.

Table 4.2 Distribution statistics of predicted annual electric lighting energy consumption (kWh/m^2)

mean	31.820124
std	9.153912
min	17.702580
25%	24.468645
50%	30.353210
75%	38.928980
max	52.281470

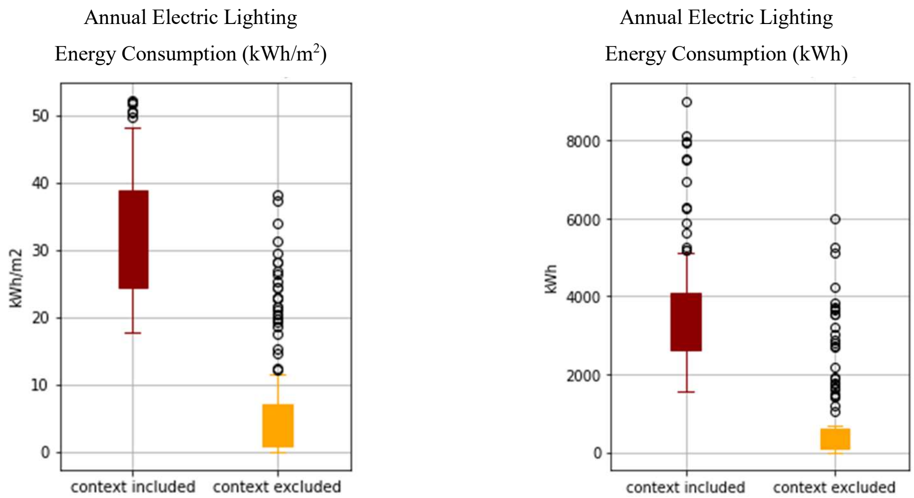
4.3 Comparative Analysis of Impact of Context on Predicted Electric Lighting Consumptions

One of the aims of the method proposed in this thesis emphasizes that it is essential to include urban form parameters in the ML models that predict daylighting performance metrics on an urban scale. Accordingly, a comparative analysis of the proposed method was made for two different cases: electric lighting energy consumption data was calculated when the impact of the surrounding buildings was included in the calculations (represented in Section 4.2) and when the units were considered independent of the urban scale. 'Context included' represents the electrical lighting energy consumption data generated based on the ML model that was trained considering the impact of the surrounding buildings and the estimated daylighting illuminance values associated with it (the results are also given in Section 4.2). 'Context excluded', on the other hand, represents that any obstruction does not shade each unit, and accordingly, the electrical lighting energy consumption is calculated with the analyzed daylighting illuminance values. Table 4.3 shows the different parameters used in the context included and excluded and how each value is calculated.

Table 4.3 Descriptions of context included and context excluded

	Context included	Context excluded
Urban shading	with context	without context (sky exposure = 100)
Daylight illuminance calculation	ML (XGBoost)	ML (XGBoost)
Lighting schedule generation	(Mardaljevic et al., 2011)	(Mardaljevic et al., 2011)
Occupancy schedule	(Malekpour Koupaei et al., 2022)	(Malekpour Koupaei et al., 2022)

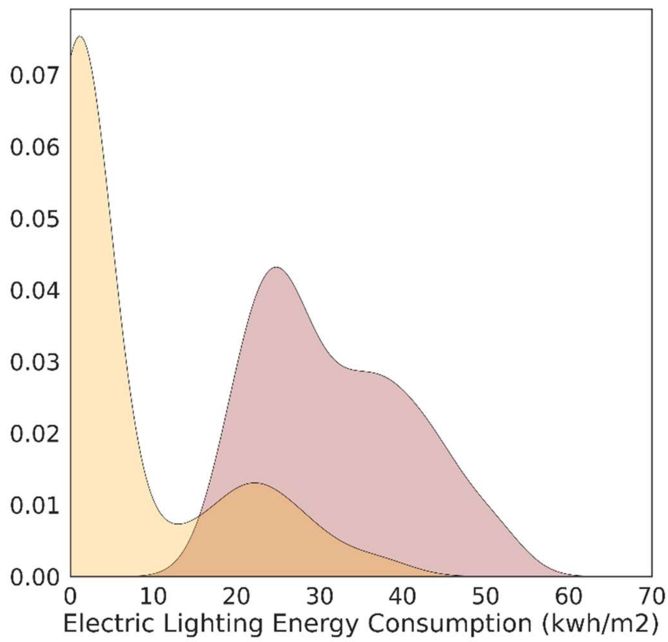
Figure 4.5 shows the comparison of annual electric lighting energy consumption in cases of context buildings are included and excluded, and Figure 4.6 shows the density plots of annual electric lighting energy consumption in cases of context buildings included and excluded.



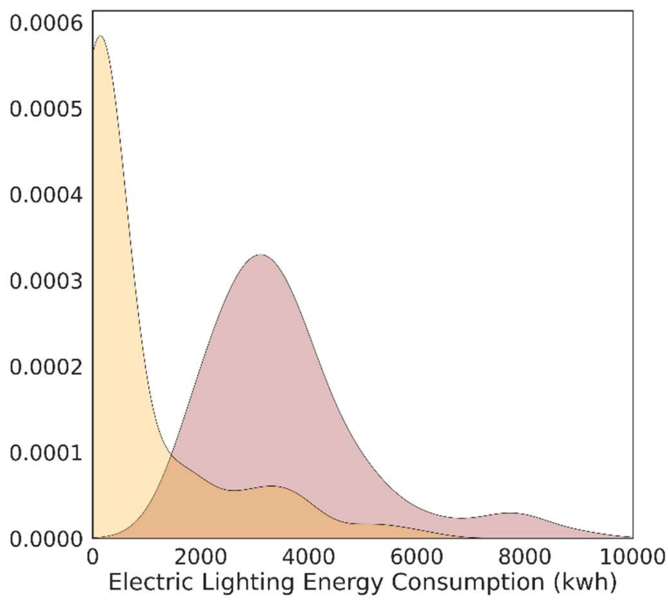
(a) : Annual electric lighting energy consumption per square meter

(b) : Total annual electric lighting energy consumption

Figure 4.5 Comparison of annual electric lighting energy consumption



(a) : Density plots of annual electric lighting energy consumption per square meter in case of context included and excluded (red: context included, orange: context excluded)



(b) : Density plots of total annual electric lighting energy consumption in case of context included and excluded (red: context included, orange: context excluded)

Figure 4.6 Density plots of annual electric lighting energy consumption in cases of context buildings are included and excluded

Table 4.4 shows the statistical descriptions of electric lighting consumption in cases of context included and excluded. When the context is included, the units' average annual electrical lighting energy consumption is 31.8 kWh/m^2, while this value decreased to 6.55 kWh/m^2 when the contexts are excluded. Additionally, when the context buildings are considered, the calculated electrical lighting consumption reaches a maximum value of 52.28 kWh/m^2, whereas when the context is excluded, the highest value drops to 38.15 kWh/m^2. The fact that the units are handled independently of the surrounding buildings on an urban scale has led to the overestimation of the daylighting illuminance values, resulting in the underestimation of the electrical lighting energy consumption calculated based on daylight.

Table 4.4 Statistical descriptions of electric lighting consumption in cases of context included and excluded

Electric lighting consumption results (kWh/m^2)	**mean**	**std**	**min**	**0.25**	**0.5**	**0.75**	**max**
Context included	31.82	9.15	17.7	24.46	30.35	38.92	52.28
Context excluded	6.55	10.04	0	0.79	1.29	7.08	38.15

4.4 Comparative Analysis of UBEM Simulation Results and Predicted UBEM Results

Standard lighting schedules that include hourly lighting fractions throughout the year are used when calculating electrical lighting consumption in Urban Building Energy Modeling (UBEM). These fractions are multiplied by the lighting power density in UBEM simulations to determine the energy used for electric lighting. These schedules, however, are constant throughout all units, independently from daylight illuminances. At the same time, common simulation models portray occupants as following identical schedules and engaging in the same behaviors leading to the calculation of inaccurate hourly demand peaks and, ultimately, an inaccurate representation of urban energy demands (el Kontar & Rakha, 2018). The typical lighting and occupancy schedules used to estimate the energy consumption of electric lighting in UBEMs are shown in Appendix D.

Table 4.5 Descriptions of UBEM simulated and UBEM predicted

	UBEM simulated	UBEM predicted
Urban shading	with context	with context
Daylight illuminance calculation	Independent from daylight illuminance	ML (XGBoost)
Lighting schedule generation	Standard lighting schedule	(Mardaljevic et al., 2011)
Occupancy schedule	Standard occupancy schedule	(Malekpour Koupaei et al., 2022)

The estimated electrical lighting consumption produced by the suggested method is compared with the electrical lighting consumption calculated by standard lighting and occupancy schedules. For this comparison, the electrical lighting consumption results calculated using the approach described in the thesis for the same units (predicted UBEM) were compared to the simulation results of the project with the code 120M997 (simulated UBEM). Table 4.5 shows the different parameters in the methods used in ‘UBEM simulated’, and ‘UBEM predicted’, and how each value is calculated. Figure 4.7 demonstrates the annual electric lighting energy consumption (kWh/m^2) in cases of predicted UBEM and simulated UBEM, and Figure 4.8 shows the density plots of annual electric lighting energy consumption in cases of predicted UBEM and simulated UBEM.

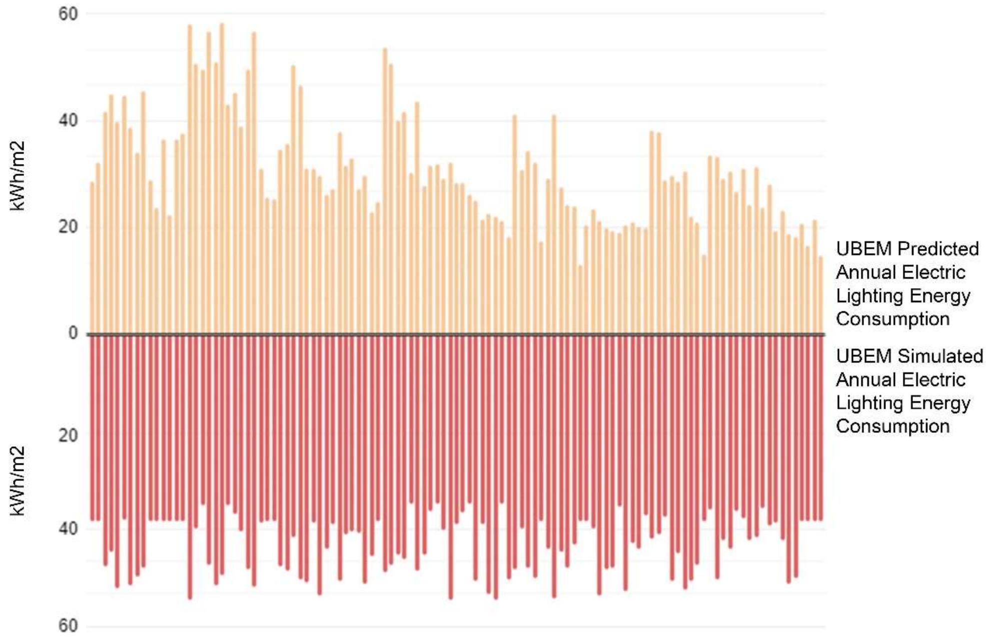


Figure 4.7 Comparison of annual electric lighting consumption in cases of UBEM predicted and simulated

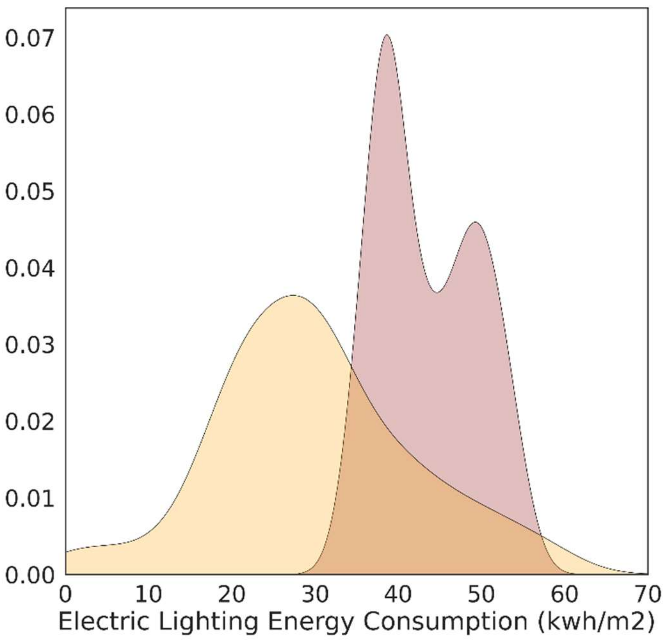


Figure 4.8 Density plots of annual electric lighting energy consumption in cases of UBEM predicted and UBEM simulated (orange: UBEM predicted, red: UBEM simulated)

Table 4.6 Statistical descriptions of electric lighting consumption in cases of UBEM predicted and UBEM simulated

Electric lighting consumption results (kWh/m2)	**mean**	**std**	**min**	**0.25**	**0.5**	**0.75**	**max**
UBEM predicted	31.22	10.53	12.97	23.01	29.5	37.75	58.34
UBEM simulated	43.56	5.95	34.74	38.38	42.13	48.6	54.58

Table 4.6 shows the statistical descriptions of electric lighting consumption in UBEM-predicted and UBEM-simulated cases. The average electrical lighting energy usage was 43.56 kWh/m^2 when the UBEM simulated electrical lighting energy consumption statistics were considered. In the UBEM simulation results, the unit

with the lowest electric lighting energy consumption value of 34.74 kWh/m^2 and the highest value of 54.58 kWh/m^2 used electric lighting energy.

Predicted UBEM's electric lighting energy consumption statistics indicate that the average value was 31.22 kWh/m^2. While the minimum electrical lighting energy consumption value observed in the UBEM predicted results was 12.97 kWh/m^2, the maximum value was 58.34 Kwh/m^2.

The average lighting energy consumption is greater in the simulation findings when UBEM simulation and predicted electric lighting energy consumption results are compared. This demonstrates that when calculations are performed using standard schedules on an urban scale in the research area, the calculated results overestimate the energy used for electric lighting. Overestimating the energy use for electrical lighting also results in overestimating the internal load of the building. It was found that, within the scope of the studied area, the units required less artificial light than the standard lighting and occupancy schedule indicated, resulting in reduced electrical lighting energy consumption.

When the distribution graphs are examined, it is seen that the electrical lighting energy consumption data of the units calculated with UBEM simulations range between 34.74 and 54.5, while this range widens to the range of 12.97 and 58.34 in UBEM predicted. This results from the varying electrical lighting consumption needs of the units depending on daylighting on an urban scale. On an urban scale, 'UBEM predicted' more accurately captured the variation in electrical lighting energy requirements that alter based on the daylighting illuminance of the units and variable occupancy profiles. The only factor affecting the values of the units in the UBEM simulation results that varied from one another was the lighting power density.

CHAPTER 5

CONCLUSION

The study's contributions were discussed in this chapter by going over the research objectives again. The proposed method's contributions and limitations were examined.

5.1 Contributions

The developed methodology was applied to residential units in Bahçelievler, Ankara. The methodology was based on a data-driven approach with four main steps. The findings demonstrated that various ML models can be utilized to forecast indoor daylighting illuminances in urban settings. With the suggested methodology, indoor daylighting illuminances in an urban setting and, consequently, electric lighting energy consumption can be predicted. For predicting interior daylighting illuminances, the ML model in the presented method works well with an R^2 value of 0.94 and an MAE value of 0.01. The following paragraphs will explore the proposed methodology's contributions based on dataset generation, ML model development, and electric lighting energy consumption steps.

Dataset generation consists of two steps: 3D modeling and simulation-based data generation. Simulation models were utilized to generate the hourly average daylighting illuminance data that would be used to train machine learning algorithms. The simulation models required the preparation of three distinct types of information geometric, semantic, and climate data. The use of geometric data from publicly accessible sources demonstrates the applicability of the developed methodology in urban settings with comparable open sources, such as EPC and NVI. In the absence of publicly accessible data such as VT, data can also be generated using common distribution types and value interval determination. Generating data based on distribution provides an advantage in considering the diversity of units on the urban scale. The created dataset is available if third parties request the generated dataset to calculate hourly daylighting illuminance on an urban scale. The success of the generated input data varies according to the user's modeling experience and model level. For the readers' information, even if the same methodology proposed in the thesis is used, the results calculated by another model or modeling the same model by another person can differ. The differences between the generated simulation-based data will always remain a challenging issue in energy modeling.

The ML model development aims to develop an ML model to estimate indoor daylighting illuminances in an urban context accurately. The suggested methodology's ability to forecast indoor daylighting illuminance with hourly resolution is one of its novelties. Predicting hourly indoor illuminances enables hourly electricity lighting consumption to be calculated. In order to analyze whether the produced electrical energy is consumed hourly or not, it is important to calculate data with hourly resolution. Hourly average illuminance information is obtained from the simulation model prepared with input data. Although illuminance values were calculated at grid resolution in the study, the values calculated at each grid resolution were given to ML models by taking the average value to represent the entire space. The main reason for choosing the average value representing the space instead of calculating each grid is as follows: In the study, it was desired to analyze how the consideration of variable internal and external parameters at the urban scale affects the ML models and indoor daylighting illuminance estimation performance and the effect on the electrical lighting energy consumption calculated based on daylight. Each place is thought to be illuminated with a single artificial lighting, and calculations are made according to this assumption. The fact that each unit was different from the others in square meters would uncontrollably increase the number of grid points, and the electrical lighting consumption was calculated accordingly, considering that there was a light source at each grid point.

Several ML methods were compared to observe which is the most successful in predicting indoor daylighting illuminances in an urban context. High-performance ML models were achieved with the data augmentation method proposed in the study.

Data augmentation: The distribution of the data in the dataset and the dataset's size significantly affect ML models' performance. This study observed that ML models' performance was very low when they were trained with the initial data (V1-base scenario). The distribution of the available data was examined, and it was observed that the dataset was unevenly distributed. A substantial part of the data has been collected in a certain value range and has become representative of the whole dataset. In order to expand the distribution range of the data and to augment the data that is scarce in the dataset, new scenarios were produced, and new input parameter ranges were determined. New simulations were made with the input features whose intervals were expanded. The solution is to widen the space set so that ML models can recognize as wide a range of data as possible and make predictions with better performance. The prediction performance of the retrained ML models with the proposed method has increased. Estimating daylighting illuminance values at an urban scale is a complex problem, and it has been seen that increasing the dataset size and diversity affects ML models' performance. In this study, no further data augmentation was made to improve the model performance when the R^2 ratio of 0.94 and an MAE value of 0.01 was obtained. By using the data augmentation method proposed in this study, data diversity in the ML model, where higher performance is expected, can be increased by generating new data using different ranges of different input parameters.

Electric lighting energy consumption calculation includes three basic steps: Creating the lighting fraction values according to the illuminance values from the ML models, multiplying these values with the created occupancy schedules, and multiplying the obtained value with the lighting power density of the units. The equation used to create the lighting fractions is based on graphs from a publication regarding a code used for residential buildings. In order to reflect the diversity of the occupancy profile, different occupancy schedules have been created according to household numbers and days of the week. Lighting power densities were obtained by creating a distribution according to the value range obtained from the literature. The developed method, with references from different studies, shows a novel approach that provides electricity consumption for lighting depending on variable lighting and occupant profiles on an urban scale. Validation of the calculated values is challenging. No hourly resolution consumption data can be validated for electrical lighting consumption in cities. This makes it impossible to validate the results obtained with real data. Studies in the literature can be analyzed with the results calculated with the proposed method comparatively. However, the studies in the literature are very different from each other compared to the data used. For example, most of the studies in the literature do not consider a lighting energy consumption calculation based on the illuminance value. Also, the occupancy profiles that studies use, if any, differ greatly from each other. The lighting power density values of the examined units also diverge from each other. Many of the studies in the literature use a fixed lighting power density without creating dispersion for each unit. Therefore, it is impossible to compare the results calculated with the obtained method with the studies in the literature. The most realistic comparison scenario is the project coded 120M997, where the same units are analyzed with urban-scale building energy simulations. Here, electrical lighting consumption is calculated with standard lighting schedules independent of the lighting performance of the units. However, considering different occupancy profiles, having identical units, and creating the lighting power densities by distribution was the closest result that could be validated with the method proposed in the thesis. It has been observed that the electrical lighting consumption data calculated with the method proposed in the thesis is much less than the results obtained with the simulations. Accordingly, the electrical lighting and the internal load handled independently from daylighting at the urban scale are overestimated.

5.2 Limitations and Future Work

The results show that several machine learning models can be used to predict interior daylighting illuminances in urban contexts. Indoor daylighting illuminances in an urban environment and electric lighting energy consumption can be calculated using the proposed methodology. The developed method enables indoor daylighting illuminance estimation by considering the shading effect of the surrounding buildings on an urban scale and calculating the electrical lighting energy consumption in hourly resolution depending on the predicted daylighting. At the same time, the methodology considers variable occupancy profiles at the urban scale with references from different sources. The study can represent the diversity of units on an urban scale in terms of daylighting illuminance and occupancy. However, the study can be strengthened and challenged in several ways. In the following paragraphs, several limitations of the proposed methodology and how future works will increase the applicability of the developed methodology will be discussed.

Orientation of the buildings: In the study, each unit is positioned at an angle of 90 so that the y-axis points north-south and the x-axis points east-west. This enabled the input parameters to be handled easily in four directions (north, south, east, and west). Window ratio and SE information are entered as separate input features for each direction. However, if the units were standing at a different angle than 90, the intermediate directions (such as northwest and southeast) would have made it impossible to represent the parameters as only four directions. In such a case, either alternative direction information should be provided as parameters to the ML model, or the orientation of the buildings should be specified using both direction and angle. Bahçelievler has a gridiron morphology, so the buildings are located parallel to each other. This ensures that the directions represented for each building are the same. However, if the buildings were not parallel, the same directions could not be mentioned for each building.

Form of the units: The units are modeled to each hand to fit into a rectangular layout, simplifying their layout. Layout information was obtained from the 2D file from the ministry. However, while modeling, each building area has been transformed into a rectangular form without deterioration to prevent complex problems arising from geometries. This approach is widely used in the literature, especially in energy modeling studies.

Data imbalance: In the study, an imbalance was observed in the distribution of the data obtained by simulation, requiring the need to produce data with new scenarios. When the data is more evenly distributed, data augmentation may not be needed, and ML models can be trained with high performance with fewer datasets.

This thesis dealt with a gridiron urban morphology; as a result, the trained model's predictive capacity is highest in similar morphologies. In future works, other morphologies (organic, radial, no pattern) can be studied, and corresponding ML models can be developed with a higher level of generality.

This study's objective is to estimate the values of interior daylighting illuminance at the urban scale. The regression issue was resolved using three different ML models: MLP, RF, and XGBoost. The study compared the success of different ML models in estimating the indoor daylighting illuminance values instead of estimating with a single ML model. These models performed well in terms of prediction. Different hyperparameters were used for different ML models examined in the study. Hyperparameters combined with the random search are tuned to improve model performance. However, the parameters and values specified here can always be expanded and differentiated to improve the performance of ML models. The estimation performance can be improved by employing various models and models with various architectures in future works.

REFERENCES

Acosta, I., Campano, M. Á., Domínguez, S., & Fernández-Agüera, J. (2019). Minimum Daylight Autonomy: A New Concept to Link Daylight Dynamic Metrics with Daylight Factors. *LEUKOS - Journal of Illuminating Engineering Society of North America*, *15*(4), 251–269. https://doi.org/10.1080/15502724.2018.1564673

Acosta, I., Navarro, J., & Sendra, J. J. (2013). Predictive method of the sky component in a courtyard under overcast sky conditions. *Solar Energy*, *89*, 89–99. https://doi.org/10.1016/j.solener.2012.12.009

Adres Sorgulama. (n.d.). Retrieved December 1, 2022, from https://adres.nvi.gov.tr/VatandasIslemleri/AdresSorgu

Ali, U., Shamsi, M. H., Hoare, C., Mangina, E., & O'Donnell, J. (2021). Review of urban building energy modeling (UBEM) approaches, methods and tools using qualitative and quantitative analysis. *Energy and Buildings*, *246*, 111073. https://doi.org/10.1016/j.enbuild.2021.111073

Alshaibani, K. (1999). A Daylight Factor for Clear Sky Conditions. *Architectural Science Review*, *42*(4), 293–296. https://doi.org/10.1080/00038628.1999.9697431

Amasyali, K., & El-Gohary, N. (2016). Building Lighting Energy Consumption Prediction for Supporting Energy Data Analytics. *Procedia Engineering*, *145*, 511–517. https://doi.org/10.1016/j.proeng.2016.04.036

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. In *Renewable and Sustainable Energy Reviews* (Vol. 81, pp. 1192–1205). Elsevier Ltd. https://doi.org/10.1016/j.rser.2017.04.095

Ang, Y. Q., Berzolla, Z. M., Letellier-Duchesne, S., Jusiega, V., & Reinhart, C. (2022). UBEM.io: A web-based framework to rapidly generate urban building energy models for carbon reduction technology pathways. *Sustainable Cities and Society*, *77*(November 2021). https://doi.org/10.1016/j.scs.2021.103534

Anil K., J., & Jianchang, M. (1996). Artificial Neural Networks: A Tutorial. *Computer*, *29*(3), 31–44.

Argonne, Z., Capanna, S., Ma, O., Spitsen, P., Hamos, I., Rench-, D., Gilman, P., Mengers, J., Hume, D., Johnson, A., Battey, H., Joseck, F., Nealer, R., Croteau, H., Mcmahon, K., & Haq, Z. (2017). *2017 Renewable Energy Data Book*. https://www.nrel.gov/docs/fy19osti/72170.pdf

Ashmore, J., & Richens, P. (2001). Computer Simulation in Daylight Design: A Comparison. *Architectural Science Review*, *44*(1), 33–44. https://doi.org/10.1080/00038628.2001.9697451

Association Suisse des Electriciens. 1989. Éclairage intérieur par la lumière du jour. Zurich (Switzerland): Association Suisse Des Electriciens. Swiss Norm SN 418911.

Autodesk. (2000). www.lightscape.com

Aydinalp, M., Ugursal, V. I., & Fung, A. S. (n.d.). *Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks*. www.elsevier.com/locate/apenergy

Ayoub, M. (2019a). 100 Years of daylighting: A chronological review of daylight prediction and calculation methods. *Solar Energy*, *194*(August), 360–390. https://doi.org/10.1016/j.solener.2019.10.072

Ayoub, M. (2019b). A multivariate regression to predict daylighting and energy consumption of residential buildings within hybrid settlements in hot-desert climates. *Indoor and Built Environment*, *28*(6), 848–866. https://doi.org/10.1177/1420326X18798164

Ayoub, M. (2020). A review on machine learning algorithms to predict daylighting inside buildings. *Solar Energy*, *202*(January), 249–275. https://doi.org/10.1016/j.solener.2020.03.104

Bainbridge, D. A., & Haggard, K. L. (2011). *Passive solar architecture : heating, cooling, ventilation, daylighting, and more using natural flows*. Chelsea Green Pub.

Becker M and Golay P. Rhino NURBS 3D modeling. Indianapolis: New Riders, 1999.

Beckers, B., & Rodríguez, D. (2009). Helping architects to design their personal daylight. *WSEAS Transactions on Environment and Development*, *5*(7), 467–477.

Bellia, L., Fragliasso, F., & Stefanizzi, E. (2017). Daylit offices: A comparison between measured parameters assessing light quality and users' opinions. *Building and Environment*, *113*, 92–106. https://doi.org/10.1016/j.buildenv.2016.08.014

Bian, Y., & Ma, Y. (2017). Analysis of daylight metrics of side-lit room in Canton, south China: A comparison between daylight autonomy and daylight factor. *Energy and Buildings*, *138*, 347–354. https://doi.org/10.1016/j.enbuild.2016.12.059

Bichiou, Y., & Krarti, M. (2011). Optimization of envelope and HVAC systems selection for residential buildings. *Energy and Buildings*, *43*(12), 3373–3382. https://doi.org/10.1016/j.enbuild.2011.08.031

Boubekri, M. (2014). Daylighting Design. In *Daylighting Design*. https://doi.org/10.1515/9783038214786

Brandão De Vasconcelos, A., Pinheiro, M. D., Manso, A., & Cabaço, A. (2016). EPBD cost-optimal methodology: Application to the thermal rehabilitation of the building envelope of a Portuguese residential reference building. *Energy and Buildings*, *111*, 12–25. https://doi.org/10.1016/j.enbuild.2015.11.006

Breiman, L. (2001). *Random Forests* (Vol. 45).

Bryan, H. J., & Carlberg, D. B. (1985). Development of protractors for calculating the effects of daylight from clear skies. *Journal of the Illuminating Engineering Society*, *14*(2), 649–657. https://doi.org/10.1080/00994480.1985.10748804

Bugeat, A., Fernández, E., Beckers, B., & Aguerre, J. (2019). A multi-scale consideration of daylight in a real urban context. *Building Simulation Conference Proceedings*, *5*, 3460–3467. https://doi.org/10.26868/25222708.2019.210650

Carol Willis, Form Follows Finance: Skyscrapers and Skylines in New York and Chicago, Princeton Architectural Press, New York, 1995

Cauchy, A. (1847). M´ethode g´en´erale pour la r´esolution des syst`emes d'´equations simultan´ees. *Comptes Rendus Hebd. Seances Acad. Sci.*, *25*, 536–538.

Chartered Institution of Building Services Engineers [CIBSE]. (1999). Lighting Guide LG10: Daylighting and window design. *Building*, 1–99.

Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System*. https://doi.org/10.1145/2939672.2939785

CIBSE. (2013). *LG9: Lighting Guide 9 Lighting for Communal Residential Buildings*.

CIE 117-1995 Discomfort Glare in Interior Lighting. CIE. 1995. ISBN 978-3-900734-70-1.

Corrado, V., Ballarini, I., & Paduos, S. (2014). Assessment of cost-optimal energy performance requirements for the Italian residential building stock.

Energy Procedia, 45, 443–452. https://doi.org/10.1016/j.egypro.2014.01.048

Council, U. G. B. (2013). LEED reference guide for building design and construction. US Green Building Council.

Crawley, D. B., Lawrie, L. K., Winkelmann, F. C., Buhl, W. F., Huang, Y. J., Pedersen, C. O., Strand, R. K., Liesen, R. J., Fisher, D. E., Witte, M. J., & Glazer, J. (2001). EnergyPlus: Creating a new-generation building energy simulation program. *Energy and Buildings*, *33*(4), 319–331. https://doi.org/10.1016/S0378-7788(00)00114-6

Das, H. S., & Roy, P. (2019). A deep dive into deep learning techniques for solving spoken language identification problems. In *Intelligent Speech Signal Processing* (pp. 81–100). Elsevier. https://doi.org/10.1016/B978-0-12-818130-0.00005-2

Defoe, P. S. (2016). *The Assessment of Daylight in Rights to Light Cases The Assessment of Daylight in Rights of Light Cases*. *January 2013*.

Department of Energy, U. (2015). *AN ASSESSMENT OF ENERGY TECHNOLOGIES AND RESEARCH OPPORTUNITIES Chapter 5: Increasing Efficiency of Building Systems and Technologies*.

Dogan, T., & Park, Y. C. (2019). A critical review of daylighting metrics for residential architecture and a new metric for cold and temperate climates. *Lighting Research and Technology*, *51*(2), 206–230. https://doi.org/10.1177/1477153518755561

Dogan, T., Reinhart, C., & Michalatos, P. (2012). Urban daylight simulation calculating the daylit area of urban designs. *Proceedings of SimBuild*, *5*(1), 613–620.

Dresler, A. (1954). The “Reflected Component” in Daylighting Design. *Lighting Research and Technology*, *19*(2 IEStrans), 50–60. https://doi.org/10.1177/147715355401900203

D. Ripley, B. (2008). Pattern Recognition and Neural Networks. *Cambridge University Press*.

Drury, B., Curtis, O., Linda, K., & Frederick, C. (2000). EnergyPlus : Energy simulation program. *ASHRAE Journal*, *42*(April), 49–56.

Dufton, A.F., 1946. Protractors for the computation of daylight factors. D.S.I.R. building research technical paper no. 28.

el Kontar, R., & Rakha, T. (2018). Profiling Occupancy Patterns to Calibrate Urban Building Energy Models (UBEMs) Using Measured Data Clustering. *Technology Architecture and Design*, *2*(2), 206–217. https://doi.org/10.1080/24751448.2018.1497369

Enkvist, P.-A., Dinkel, J., & Lin, C. (2010). *Impact of the financial crisis on carbon economics Version 2.1 of the Global Greenhouse Gas Abatement Cost Curve Impact of the financial crisis on carbon economics: Version 2.1 of the global greenhouse gas abatement cost curve*. www.mckinsey.com/globalGHGcostcurve

Evins, R. (2013). A review of computational optimisation methods applied to sustainable building design. *Renewable and Sustainable Energy Reviews*, *22*, 230–245. https://doi.org/10.1016/j.rser.2013.02.004

Fakra, A. H., Miranville, F., Boyer, H., & Guichard, S. (2011). Development of a new model to predict indoor daylighting: Integration in CODYRUN software and validation. *Energy Conversion and Management*, *52*(7), 2724–2734. https://doi.org/10.1016/j.enconman.2011.01.019

Feuermann, D., & Novoplansky, A. (1998). Reversible low solar heat gain windows for energy savings. *Solar Energy*, *62*(3), 169–175. https://doi.org/10.1016/S0038-092X(98)00015-2

Frühling, H.G., 1928. Tagesbeleuchtung von Innenräumen: ihre Messung und ihre Berechnung nach der Wirkungsgradmethode [Daylighting of Indoors, Their Measurement and Calculation by the Efficiency Method]. Technical University of Berlin, Berlin, Germany.

Fuller, M. 1985. Concepts and Practice of Architectural Daylighting. New York: Van Nostrand Reinhold Company.

Fung, Y. W., & Lee, W. L. (2012). Developing a simplified parameter for assessing view obstruction in high-rise high-density urban environment. *Habitat International*, *36*(3), 414–422. https://doi.org/10.1016/J.HABITATINT.2012.01.001

Gadelhak, M. I. (2015). *M . Sc . Thesis : High Performance Facades : Designing Office Building Facades to Enhance Indoor Daylighting Performance High Performance Facades : Designing Office Building Facades to Enhance Indoor Daylighting Performance . A Thesis Presented in Partial*. *July*. https://doi.org/10.13140/RG.2.1.4301.1684

Gherri, B. (2015). *Assessment of Daylight Performance in Buildings*. WIT Press. http://library.witpress.com

Goodfellow, I., Bengio, Y., & Courville, A. (n.d.). *Deep learning*.

Halonen, L., Tetri, E., & Bhusal, P. (n.d.). *Energy Efficient Electric Lighting for Buildings ECBCS Annex 45*.

Hamdy, M., Hasan, A., & Siren, K. (2013). A multi-stage optimization method for cost-optimal and nearly-zero-energy building solutions in line with the EPBD-recast 2010. *Energy and Buildings*, *56*, 189–203. https://doi.org/10.1016/j.enbuild.2012.08.023

Han, Y., Shen, L., & Sun, C. (2021). Developing a parametric morphable annual daylight prediction model with improved generalization capability for the early stages of office building design. *Building and Environment*, *200*(April), 107932. https://doi.org/10.1016/j.buildenv.2021.107932

Happle, G., Fonseca, J. A., & Schlueter, A. (2018). A review on occupant behavior in urban building energy models. In *Energy and Buildings* (Vol. 174, pp. 276–292). Elsevier Ltd. https://doi.org/10.1016/j.enbuild.2018.06.030

HK-BEAM Society. (2004). *An environmental assesment for new buildings*. http://www.

Hopkinson, R. G., Longmore, J., & Petherbridge, P. (1954). An Empirical Formula for the Computation of the Indirect Component of Daylight Factor. *Lighting Research and Technology*, *19*(7 IEStrans), 201–219. https://doi.org/10.1177/147715355401900701

Hopkinson, R.G., Petherbridge, P., Longmore, J. 1966. Daylighting. London: Heinemann.

Howard Davis, The future of ancient lights, J. Archit. Plan. Res. 6 (1989) 132–153

IESNA (Ed.). (2000). *IESNA 9th Handbook*. ISBN 0-87995-150-8

IESNA. (2007). *Recommended practice for lighting and the visual environment for senior living*.

IESNA. (2012). IES Spatial Daylight Autonomy (sDA) and Annual Sunlight Exposure (ASE), Daylight Metrics Committee. Approved Method IES LM-83-12. *Illuminating Engineering Society of North America*.

International WELL Building Institute. (2020). WELL certification.

ISICAM sistemleri. Isıcam Cam Yalıtımı Sistemleri. (n.d.). Retrieved December 1, 2022, from https://www.isicam.com.tr/tr

Iuliano, G., Bellia, L., Isoardi, G., Coyne, S., Hirning, M., Garcia-Hansen, V., de Rosa, F., & Pagliano, L. (n.d.). *DAYLIGHT GLARE: A REVIEW OF DISCOMFORT INDEXES Per-Pixel Light ing Dat a Analysis Mehlika Inanici A review of indices for assessing visual comfort wit h a view t o t heir use in opt imizat ion processes t o …*.

Jakubiec, J. A., & Reinhart, C. F. (2011). DIVA 2.0: Integrating daylight and thermal simulations using rhinoceros 3D, DAYSIM and EnergyPlus. *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, 2202–2209.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. https://doi.org/10.1126/science.aac4520

Jordan, M., Kleinberg, J., & Schölkopf, B. (n.d.). *Information Science and Statistics*.

Kaminska, A., & O˙ Zadowicz, A. (2018). *Lighting Control Including Daylight and Energy Efficiency Improvements Analysis*. https://doi.org/10.3390/en11082166

Kazanasmaz, T., Grobe, L. O., Bauer, C., Krehel, M., & Wittkopf, S. (2016). Three approaches to optimize optical properties and size of a South-facing window for spatial Daylight Autonomy. *Building and Environment*, *102*, 243–256. https://doi.org/10.1016/j.buildenv.2016.03.018

Kazanasmaz, T., Günaydin, M., & Binol, S. (2009). Artificial neural networks to predict daylight illuminance in office buildings. *Building and Environment*, *44*(8), 1751–1757. https://doi.org/10.1016/j.buildenv.2008.11.012

Kingma, D. P., & Ba, J. (2014). *Adam: A Method for Stochastic Optimization*. http://arxiv.org/abs/1412.6980

Kirimtat, A., Koyunbaba, B. K., Chatzikonstantinou, I., & Sariyildiz, S. (2016). Review of simulation modeling for shading devices in buildings. *Renewable and Sustainable Energy Reviews*, *53*, 23–49. https://doi.org/10.1016/J.RSER.2015.08.020

Knoop, M., Stefani, O., Bueno, B., Matusiak, B., Hobday, R., Wirz-Justice, A., Martiny, K., Kantermann, T., Aarts, M. P. J., Zemmouri, N., Appelt, S., & Norton, B. (2020). Daylight: What makes the difference? *Lighting Research and Technology*, *52*(3), 423–442. https://doi.org/10.1177/1477153519869758

Kota, S., & Haberl, J. S. (2009). Historical Survey of Daylighting Calculations Methods and Their Use in Energy Performance Simulations. *Proceedings of the Ninth International Conference for Enhanced Building Operations*, 9.

Kuhn, M., & Johnson, K. (n.d.). *Applied Predictive Modeling*.

Kurian, C. P., George, V. I., Bhat, J., & Aithal, R. S. (2006). ANFIS model for the time series prediction of interior daylight illuminance. *ANfIS MODEL FOR THE TIME SERIES PREDICTION OF INTERIOR DAYLIGHT ILLUMINANCE*, *6*, 35–40. www.icgst.comwww.icgst.com

Le-Thanh, L., Nguyen-Thi-Viet, H., Lee, J., & Nguyen-Xuan, H. (2022). Machine learning-based real-time daylight analysis in buildings. *Journal of Building Engineering*, *52*(November 2021), 104374. https://doi.org/10.1016/j.jobe.2022.104374

Li, D. H. W., Lam, J. C., & Li, D. H. W. (n.d.). *Measurements of solar radiation and illuminance on vertical surfaces and daylighting implications*. www.elsevier.com/locate/renene

Littlefair, P. (2001). Daylight, sunlight and solar gain in the urban environment. *Solar Energy*, *70*(3), 177–185. https://doi.org/10.1016/S0038-092X(00)00099-2

Littlefair, P. (2002). Daylight prediction in atrium buildings. *Solar Energy*, *73*(2), 105–109. https://doi.org/10.1016/S0038-092X(02)00038-5

Li, W., Zhou, Y., Cetin, K., Eom, J., Wang, Y., Chen, G., & Zhang, X. (2017). Modeling urban building energy use: A review of modeling approaches and procedures. *Energy*, *141*, 2445–2457. https://doi.org/10.1016/j.energy.2017.11.071

Lockley, S. W. (2009). Circadian Rhythms: Influence of Light in Humans. *Encyclopedia of Neuroscience*, 971–988. https://doi.org/10.1016/B978-008045046-9.01619-3

Lorenz, C.-L., & Jabi, W. (2017). *Predicting Daylight Autonomy Metrics Using Machine Learning*. https://www.researchgate.net/publication/322049892

Lynes, J.A., 1979. A sequence for daylighting design. Light. Res. Technol. 11, 102–106. https://doi.org/10.1177/14771535790110020101.

Malekpour Koupaei, D., Cetin, K. S., & Passe, U. (2022). Stochastic residential occupancy schedules based on the American Time-Use Survey. *Science and Technology for the Built Environment*, *28*(6), 776–790. https://doi.org/10.1080/23744731.2022.2087536

Mangan, S. D., & Oral, G. K. (2016). Assessment of residential building performances for the different climate zones of Turkey in terms of life cycle energy and cost efficiency. *Energy and Buildings*, *110*, 362–376. https://doi.org/10.1016/j.enbuild.2015.11.002

Mardaljevic, J. (2006). Examples of climate-based daylight modelling. In CIBSE National Conference. *CIBSE National Conference 2006:*

Engineering the Future 21-22 March 2006, Oval Cricket Ground, London, UK, *67*.

Mardaljevic, J., Andersen, M., Roy, N., Christoffersen, J., Gateway, T., Enac, E., & Lipid, I. A. (2012). DAYLIGHTING METRICS : IS THERE A RELATION BETWEEN USEFUL DAYLIGHT ILLUMINANCE AND DAYLIGHT GLARE PROBABILITY ? VELUX A / S , Adalsvej. *Proceedings of the Building Simulation and Optimization Conference BSO12*. https://infoscience.epfl.ch/record/179939?ln=en

Mardaljevic, J., Andersen, M., Roy, N., Christoffersen, J., Mardaljevic, J., A., Marilyne, R., & Nicolas, C. J. (2011). Daylighting metrics for residential buildings. *Proceedings of the 27th Session of the CIE (2011)*, *January*, 18. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.372.1079&rep=rep1&type=pdf

Millet, M S, Adams, C, & Bedrick, J R. Graphic daylighting design method: including clear sky conditions. United States.

Mohelnikova, J., & Hirs, J. (2016). Effect of externally and internally reflective components on interior daylighting. *Journal of Building Engineering*, *7*, 31–37. https://doi.org/10.1016/j.jobe.2016.04.009

Muneer, T., & Beliveau, Y. (n.d.). *ECBCS Annex 29 / SHC Task 21 Daylight in Buildings Energy Conservation in Buildings and Community Systems Programme Daylight in Buildings Energy Conservation in Buildings & Community Systems*.

Munoz, C. M., Esquivias, P. M., Moreno, D., Acosta, I., & Navarro, J. (2014). Climate-based daylighting analysis for the effects of location, orientation and obstruction. *Lighting Research and Technology*, *46*(3), 268–280. https://doi.org/10.1177/1477153513487005

Nabil, A., & Mardaljevic, J. (2005). Useful daylight illuminance: A new paradigm for assessing daylight in buildings. *Lighting Research and Technology*, *37*(1), 41–59. https://doi.org/10.1191/1365782805li128oa

Nasrollahi, N., & Shokri, E. (2016). Daylight illuminance in urban environments for visual comfort and energy performance. *Renewable and Sustainable Energy Reviews*, *66*, 861–874. https://doi.org/10.1016/j.rser.2016.08.052

Nault, E., Moonen, P., Rey, E., & Andersen, M. (2017). Predictive models for assessing the passive solar and daylight potential of neighborhood designs: A comparative proof-of-concept study. *Building and Environment*, *116*, 1–16. https://doi.org/10.1016/j.buildenv.2017.01.018

Nazzal, A. A. (2005). A new evaluation method for daylight discomfort glare. *International Journal of Industrial Ergonomics*, *35*(4), 295–306. https://doi.org/10.1016/J.ERGON.2004.08.010

Ne'eman, E., Light, W., & Hopkinson, R. G. (1976). Recommendations for the admission and control of sunlight in buildings. *Building and Environment*, *11*(2), 91–101. https://doi.org/10.1016/0360-1323(76)90025-1

Ngarambe, J., Adilkhanova, I., Uwiragiye, B., & Yun, G. Y. (2022). A review on the current usage of machine learning tools for daylighting design and control. *Building and Environment*, *223*(August), 109507. https://doi.org/10.1016/j.buildenv.2022.109507

Ngarambe, J., Irakoze, A., Yun, G. Y., & Kim, G. (2020). Comparative performance of machine learning algorithms in the prediction of indoor daylight illuminances. *Sustainability (Switzerland)*, *12*(11), 1–23. https://doi.org/10.3390/su12114471

O'Connor, J. (1997). *Tips for Daylighting With Windows*. 1–107. http://windows.lbl.gov/daylighting/designguide/dlg.pdf

Olabi, A. G., & Abdelkareem, M. A. (2022). Renewable energy and climate change. *Renewable and Sustainable Energy Reviews*, *158*(January), 112111. https://doi.org/10.1016/j.rser.2022.112111

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., & Ajayi, S. (2022). Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering*, *45*(July 2021), 103406. https://doi.org/10.1016/j.jobe.2021.103406

Papinutto, M., Boghetti, R., Colombo, M., Basurto, C., Reutter, K., Lalanne, D., Kämpf, J. H., & Nembrini, J. (2022). Saving energy by maximising daylight and minimising the impact on occupants: An automatic lighting system approach. *Energy and Buildings*, *268*, 112176. https://doi.org/10.1016/J.ENBUILD.2022.112176

Phillips, D., & Gardner, C. (2012). Daylighting: Natural light in architecture. In *Daylighting: Natural Light in Architecture*. https://doi.org/10.4324/9780080477053

Rackes, A., & Waring, M. S. (2017). Alternative ventilation strategies in U.S. offices: Comprehensive assessment and sensitivity analysis of energy saving potential. *Building and Environment*, *116*, 30–44. https://doi.org/10.1016/j.buildenv.2017.01.027

Reinhart, C. F., & Herkel, S. (2000). The Simulation of annual daylight illuminance distributions-a state-of-the-art comparison of six RADIANCE-based methods. *Energy and Buildings*, *32*(2), 167–187. https://doi.org/10.1016/S0378-7788(00)00042-6

Reinhart, C. F., Mardaljevic, J., & Rogers, Z. (2006). Dynamic daylight performance metrics for sustainable building design. *LEUKOS - Journal of Illuminating Engineering Society of North America*, *3*(1), 7–31. https://doi.org/10.1582/LEUKOS.2006.03.01.001

Reinhart, C.F., Rakha, T., Weissman, D., 2014. Predicting the daylit area – A comparison of students assessments and simulations at eleven schools of architecture. Leukos 10, 193–206.

Reinhart, C. F., & Walkenhorst, O. (2001). Validation of dynamic RADIANCE-based daylight simulations for a test office with external blinds. *Energy and Buildings*, *33*(7), 683–697. https://doi.org/10.1016/S0378-7788(01)00058-5

Reinhart, C. F., & Wienold, J. (2011). The daylighting dashboard - A simulation-based design analysis for daylit spaces. *Building and Environment*, *46*(2), 386–396. https://doi.org/10.1016/j.buildenv.2010.08.001

Rogers, Z., & Goldman, D. (2006). Daylighting metric development using daylight autonomy calculations in the sensor placement optimization tool. Boulder, Colorado, USA: Architectural Energy Corporation: http://www. archenergy. com/SPOT/SPOT_Daylight% 20Autonomy% 20Report. pdf.

Rostamizadeh, A., & Talwalkar, A. (n.d.). *Foundations of Machine Learning second edition*.

Roy, N. (2014). *Daylight, Energy and Indoor Climate-Basic Book*. https://www.researchgate.net/publication/303667224

Saratsis, E., Dogan, T., & Reinhart, C. F. (2017). Simulation-based daylighting analysis procedure for developing urban zoning rules. *Building Research and Information*, *45*(5), 478–491. https://doi.org/10.1080/09613218.2016.1159850

Segal, M. R. (2003). *UCSF Recent Work Title Machine Learning Benchmarks and Random Forest Regression Publication Date Machine Learning Benchmarks and Random Forest Regression*.

Serra, R. (1998). Daylighting. *Renewable & Sustainable Energy Reviews*, *2*(1–2), 115–155. https://doi.org/10.1016/s1364-0321(98)00014-8

Sharma, S., Sharma, S., & Athaiya, A. (2020). ACTIVATION FUNCTIONS IN NEURAL NETWORKS. In *International Journal of Engineering Applied Sciences and Technology* (Vol. 4). http://www.ijeast.com

Singh, R., & Rawal, R. (2011). Effect of surface reflectance on lighting efficiency in interiors. *Proceedings of Building Simulation 2011: 12th Conference of International Building Performance Simulation Association*, 2301–2308.

Sophia Douglass Pfeiffer, Ancient lights: legal protection of access to solar energy, Am. Bar Assoc. J. 68 (1982) 288–291.

Szokolay, S. v., & Brisbin, Christopher. (2004). *Introduction to architectural science : the basis of sustainable design*. Elsevier, Architectural Press.

Tabadkani, A., Roetzel, A., Li, H. X., & Tsangrassoulis, A. (2021). Daylight in buildings and visual comfort evaluation: The advantages and limitations. In *Journal of Daylighting* (Vol. 8, Issue 2, pp. 181–203). Solarlits. https://doi.org/10.15627/jd.2021.16

Thanachareonkit, A., Scartezzini, J. L., & Andersen, M. (2005). Comparing daylighting performance assessment of buildings in scale models and test modules. *Solar Energy*, *79*(2), 168–182. https://doi.org/10.1016/j.solener.2005.01.011

The IES code for interior lighting. The CIBSE, London. CIBSE, 1994, Code for interior lighting

Tregenza, P. R. (1980). The daylight factor and actual illuminance ratios. *Lighting Research & Technology*, *12*(2), 64–68. https://doi.org/10.1177/096032718001200202

Tregenza, P. R. (1989). Modification of the split-flux formulae for mean daylight factor and internal reflected component with large external obstructions. *Lighting Research & Technology*, *21*(3), 125–128. https://doi.org/10.1177/096032718902100305

Tregenza, P. R., & Waters, I. M. (1983). Daylight coefficients. *Lighting Research & Technology*, *15*(2), 65–71. https://doi.org/10.1177/096032718301500201

Turan, I., Chegut, A., Fink, D., & Reinhart, C. (2020). The value of daylight in office spaces. *Building and Environment*, *168*(October 2019), 106503. https://doi.org/10.1016/j.buildenv.2019.106503

TUIK, "Turkish Statistical institute, Data Portal for Statistics," 2022.

Tzempelikos, A., & Athienitis, A. K. (2007). The impact of shading design and control on building cooling and lighting demand. *Solar Energy*, *81*(3), 369–382. https://doi.org/10.1016/j.solener.2006.06.015

United, N. (2018). World Urbanization Prospects. In *Demographic Research*. file:///C:/Users/rocey/Downloads/WUP2018-Report.pdf

U.S. Green Building Council. (2008). *LEED ® for Homes Rating System*.

U.S. Green Building Council. (n.d.). Lighting. U.S. Green Building Council. Retrieved December 27, 2022, from https://www.usgbc.org/credits/homes/v4-draft/eac13?view=language

Veitch, J. A., & Galasiu, A. D. (2012). The Physiological and Psychological Effects of Windows, Daylight, and View at Home: Review and Research Agenda. *Research Report (National Research Council of Canada. Institute for Research in Construction)*. https://doi.org/10.4224/20375039

Waide, P., & Tanishima, S. (2006). Light ' S Labour ' S Lost. *The International Energy Agency (IEA)*, *Policies f*(In support of the G8 Plan of Action).

Waldram, P.J., Waldram, J.M., 1923. Window design and the measurement and pre- determination of daylight illumination. Illum. Eng. XVI 90–122.

Walsh, J. W. T. (1951). The Early Years of Illuminating Engineering in Great Britain. *Lighting Research and Technology*, *16*(3 IEStrans), 49–60. https://doi.org/10.1177/147715355101600301

Ward, G. J. (1994). The RADIANCE lighting simulation and rendering system. *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1994*, 459–472. https://doi.org/10.1145/192161.192286

Ward, G. J., & Rubinstein, F. M. (1988). A New Technique for Computer Simulation of Illuminated Spaces. *Journal of the Illuminating Engineering Society*, *17*(1), 80–91. https://doi.org/10.1080/00994480.1988.10748710

Webb, A. R. (2006). Considerations for lighting in the built environment: Non-visual effects of light. *Energy and Buildings*, *38*(7), 721–727. https://doi.org/10.1016/j.enbuild.2006.03.004

Wienold, J., & Christoffersen, J. (2006). Evaluation methods and development of a new glare prediction model for daylight environments with the use of CCD cameras. *Energy and Buildings*, *38*(7), 743–757. https://doi.org/10.1016/j.enbuild.2006.03.017

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

Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26–40. https://doi.org/10.11989/JEST.1674-862X.80904120

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

Yang, S., Tian, W., Cubi, E., Meng, Q., Liu, Y., & Wei, L. (2016). Comparison of Sensitivity Analysis Methods in Building Energy Assessment. *Procedia Engineering*, *146*, 174–181. https://doi.org/10.1016/j.proeng.2016.06.369

Ye, Y., Xu, P., Mao, J., & Ji, Y. (2016). Experimental study on the effectiveness of internal shading devices. *Energy and Buildings*, *111*, 154–163. https://doi.org/10.1016/J.ENBUILD.2015.11.040

Yu, X., & Su, Y. (2015). Daylight availability assessment and its potential energy saving estimation -A literature review. *Renewable and Sustainable Energy Reviews*, *52*, 494–503. https://doi.org/10.1016/j.rser.2015.07.142

Zeiler, M. D. (2012). *ADADELTA: An Adaptive Learning Rate Method*. http://arxiv.org/abs/1212.5701

Zhang, J., Heng, C. K., Malone-Lee, L. C., Hii, D. J. C., Janssen, P., Leung, K. S., & Tan, B. K. (2012). Evaluating environmental implications of density: A comparative case study on the relationship between density, urban block typology and sky exposure. *Automation in Construction*, *22*, 90–101. https://doi.org/10.1016/j.autcon.2011.06.011

Zhang, Z. (2018). Multivariate Time Series Analysis in Climate and Environmental Research. *Multivariate Time Series Analysis in Climate and Environmental Research*, *November 2017*. https://doi.org/10.1007/978-3-319-67340-0

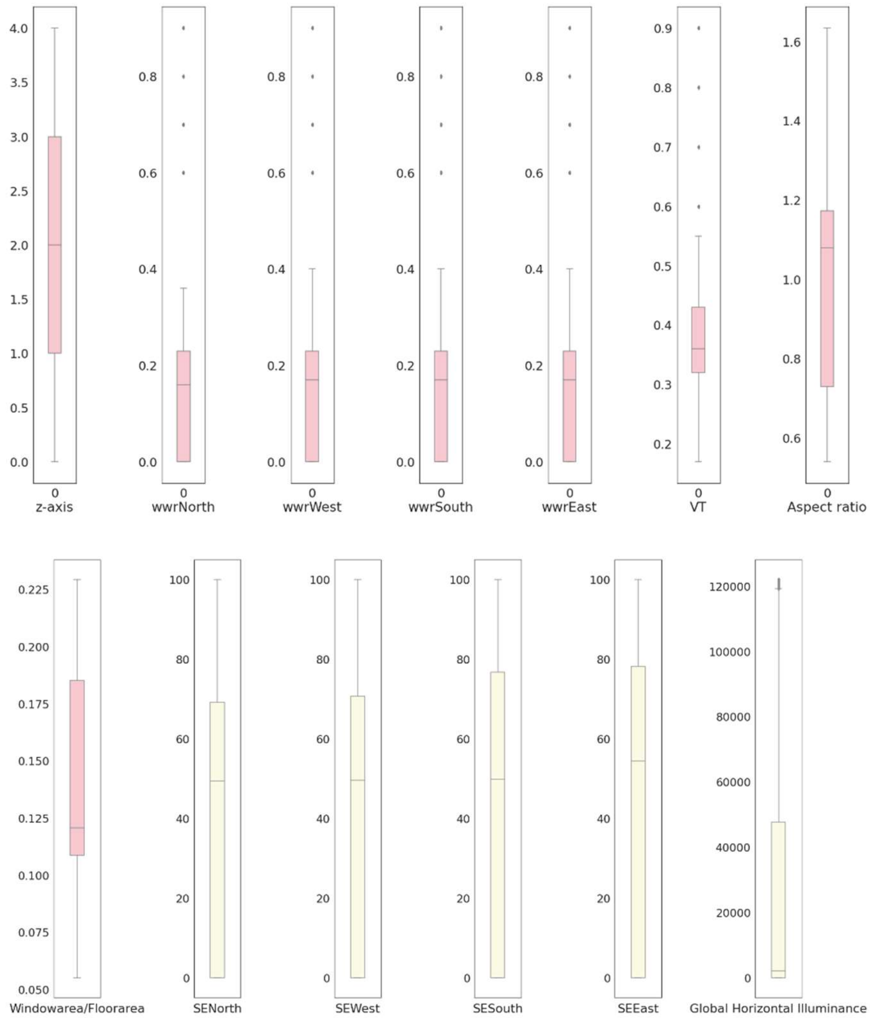
Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. In *Renewable and Sustainable Energy Reviews* (Vol. 16, Issue 6, pp. 3586–3592). https://doi.org/10.1016/j.rser.2012.02.049

Zhou, S., & Liu, D. (2015). Prediction of Daylighting and Energy Performance Using Artificial Neural Network and Support Vector Machine. *American Journal of Civil Engineering and Architecture*, *3*(3A), 1–8. https://doi.org/10.12691/ajcea-3-3A-1

Zumthor, P. (2006). *Atmospheres*.

APPENDICES

A. Distribution of Input Features



B. Statistical Descriptions of Input Data

	mean	std	min	0.25	0.50	0.75	max
average illuminance value	436.94	1188.39	0.00	0.00	6.70	210.58	15032.40
z-axis	1.70	1.28	0.00	1.00	2.00	3.00	4.00
wwrNorth	0.13	0.13	0.00	0.00	0.16	0.23	0.90
wwrWest	0.14	0.14	0.00	0.00	0.17	0.23	0.90
wwrSouth	0.15	0.15	0.00	0.00	0.17	0.23	0.90
wwrEast	0.14	0.14	0.00	0.00	0.17	0.23	0.90
Aspect ratio	1.02	0.36	0.54	0.73	1.08	1.17	1.63
TotalWindow/Total FloorArea	0.14	0.05	0.06	0.11	0.12	0.19	0.23
SENorth	41.42	36.72	0.00	0.00	49.50	69.27	100.00
SEWest	42.10	37.15	0.00	0.00	49.61	70.73	100.00
SESouth	43.13	37.81	0.00	0.00	49.95	76.76	100.00
SEEast	44.09	37.76	0.00	0.00	54.38	78.25	100.00
VT	0.38	0.13	0.17	0.32	0.36	0.43	0.90
floorarea	115.10	37.20	70.50	82.90	106.60	145.00	213.20
Global Horizontal Illuminance	25801.85	34490.33	0.00	0.00	2097	47670	121992

C. Tuned Hyperparameters for Prediction Models

Model	Description
MLP	Neural Networks MLP Regressor Hidden_layer_sizes = [32,32], [32,64], [64,32], [64,64] learning_rate = [0.001, 0.002, 0.004, 0.008] batch_size = [256, 512, 1024, 2048, 4096]
RF	Ensemble Models Random Forest Regressor n_estimators = [10, 20, 30, 40, 50] max_depth = [10, 21, 32, 43, 54, 65, 76, 87, 98, 110, None] min_samples_leaf = [1, 3, 5, 7], min_samples_split = [2, 5, 10, 20]
XGBoost	Ensemble Models XGBoost Regressor n_estimators = [100, 200, 300] max_depth = [3, 5, 7, 9, 11] min_child_weight = [1, 3, 5] gamma = [0.0,0.1,0.2, 0.3, 0.4, 0.5], learning_rate = [0.1, 0.2, 0.4, 0.8] subsample = [0.5, 0.7, 0.8, 0.9] colsample_bytree: [0.6, 0.7, 0.8, 0.9]

D. Standard Midrise Apartment Occupancy Schedule