

Machine learning based prediction of long-term energy consumption and overheating under climate change impacts using urban building energy modeling

Ilkim Canli Akyol ^{a,d,1}, Eren Gökberk Halacli ^{b,1}, Sevval Ucar ^{b,1}, Orcun Koral Iseri ^a, Feyza Yavuz ^b, Dilara Guney ^a, Fatma Ece Gursoy ^a, Ayca Duran ^g, Cagla Meral Akgul ^{c,e}, Sinan Kalkan ^{b,e,2}, Ipek Gursel Dino ^{a,e,f,2,*}

^a Dept. of Architecture, Middle East Technical University, Universiteler Mah, Dumlupinar Bulvari, No:1, 06800 Ankara, Turkiye

^b Dept. of Computer Engineering, Middle East Technical University, Universiteler Mah, Dumlupinar Bulvari, No:1, 06800 Ankara, Turkiye

^c Dept. of Civil Engineering, Middle East Technical University, Universiteler Mah, Dumlupinar Bulvari, No:1, 06800 Ankara, Turkiye

^d ODTÜ-GÜNAM (Center for Solar Energy Research and Applications), Universiteler Mah, Dumlupinar Bulvari, No:1, 06800 Ankara, Turkiye

^e METU-ROMER (Center for Robotics and Artificial Intelligence), Universiteler Mah, Dumlupinar Bulvari, No:1, 06800 Ankara, Turkiye

^f METU-CLIMATE (Climate Change and Sustainable Development Application and Research Center), Universiteler Mah, Dumlupinar Bulvari, No:1, 06800 Ankara, Turkiye

^g Architecture and Building Systems, ETH Zurich, Stefano-Francini-Platz, 18093 Zurich, Switzerland

ARTICLE INFO

ABSTRACT

Keywords:

Urban building energy modeling (UBEM)
Machine learning
Energy use prediction
Indoor overheating
Simulation
Facade retrofit
Climate change impacts

In cities, well-informed decisions targeting improved building energy performance under climate change impacts require tools that can make long-term projections and devise effective strategies. Physics-based Urban Building Energy Models (UBEM) can calculate building performance for future years; however, this process is challenging as (i) future weather files are generated only for discrete years, and (ii) physics-based simulations are computationally demanding, hindering the evaluation of a high number of buildings for all future years. Alternatively, machine learning (ML) approaches can offer high-precision estimations at a lower computational cost. In this paper, a UBEM-assisted ML-based approach that predicts residential buildings' heating energy use and indoor overheating for the current and future years is proposed. A UBEM of a residential district is developed, and simulations are performed using weather files of the current year, 2050, and 2080 to develop training/testing datasets. Multi-layer perceptrons are trained to a very high predictive performance (with an R^2 score of 0.98 and 0.96 for the two output features), with a remarkable speed advantage (~430 times faster than simulations). Finally, the results of the long-term analysis of three urban-scale retrofit scenarios are presented, which offers insights into the potential use of the proposed ML models.

Abbreviations

ANN	Artificial Neural Network	GAN	Generative Adversarial Network
CDD _{23.3}	Cooling Degree Days with 18.3 °C base temperature	GBM	Gradient Boosting Machine
DNN	Deep Neural Network	GHR	Annual Average Global Horizontal Radiation
DY _{years}	Dataset comprising of the Y Data ($Q_{Heating}$ or IOD) from Years	HDD _{18.3}	Heating Degree Days with 18.3 °C base temperature
EPC	Energy Performance Certificate	IOD	Indoor Overheating Degree
EPD	Equipment Power Density	KDE	Kernel Density Estimator
EUI	Energy Use Intensity	LPD	Lighting Power Density
		MAE	Mean Absolute Error
		ML	Machine Learning

* Corresponding author.

E-mail address: ipekg@metu.edu.tr (I.G. Dino).

¹ These authors contributed equally.

² Equal contribution for senior authorship

MLP	Multi-Layer Perceptron
MSE	Mean Squared Error
M_{years}^Y	An MLP Model Trained on D_{years}^Y
NPDE	Normalised Prediction Distribution Errors
PDE	Parametric Probability Density Estimation
PPL	People Density
Q _{heating}	Heating energy consumption
R ²	Coefficient of determination
RBFNN	Radial Basis Function Neural Network
ReLU	Rectified Linear Unit
RF	Random Forest
RMSE	Root Mean Squared Error
SE _{Dir}	Sky Exposure of a window facing direction Dir
SHGC	Solar Heat Gain Coefficient
SVM	Support Vector Machine
T _{ave}	Monthly mean outdoor air temperature
T _{comf}	Upper comfort threshold temperature
T _{comf,i,z}	Upper comfort temperature limit for zone z during the occupied hour i
T _{heating}	Heating setpoint
T _{indoor}	Indoor air temperature
TMY	Typical Meteorological Year
T _n	Comfort neutrality temperature
T _{oper,i,z}	Operative temperature at the time step i in the zone z
T _{outside}	Annual average dry bulb temperature
UBEM	Urban Building Energy Modeling
U _{ground}	Thermal transmittance of the ground
U _{roof}	Thermal transmittance of the roof
U _{wall}	Thermal transmittance of the wall
U _{window}	Thermal transmittance of the window
WWR	Window-to-wall ratio
η_{boiler}	Boiler efficiency

1. Introduction

Cities play a significant role in contributing to climate change, as they are responsible for >70 % of annual global carbon emissions and >60 % of global energy consumption (International Energy Agency, 2021). Buildings are responsible for 30 % of global final energy consumption and 26 % of global energy-related emissions, 18 % of which are attributable to the generation of heat and electricity in buildings (International Energy Agency, 2023). Cities are also expected to become increasingly more vulnerable to extreme climate events such as heat waves, droughts, storms, and floods as a result of climate change (Bektaş & Sakarya, 2023). Studies show that more pronounced overheating will be experienced in the future, which will have a strong effect on building energy use and occupant comfort (Peacock et al., 2010). Residential buildings are significant due to the large number of residential buildings that do not adhere to the current building codes (Dino & Meral Akgül, 2019). To mitigate climate change effects, it is necessary to reduce the energy consumption and associated CO₂ emissions of buildings, especially in highly urbanized areas. This makes building energy performance analysis at an urban scale crucial for stakeholders to anticipate and prepare for climate change impacts, identify vulnerabilities and risks associated with climate change, and evaluate potential opportunities for energy retrofit (Ali et al., 2024; European Environment Agency). However, the decision-making processes are still challenging for stakeholders due to the lack of tools facilitating in-depth, urban-scale analyses of the building stock (Ali et al., 2021; Reinhart & Davila, 2016).

1.1. Simulation-based calculation of climate change impacts

Building energy modeling and simulations allow the analysis of building energy use, occupant comfort, and possible reduction of energy use and greenhouse gas (GHG) emissions (Hong et al., 2020). During

simulations, climate change impacts can be taken into account with the use of future weather files. Available weather file generation tools, *CCWorldWeatherGen* (Jentsch et al., 2013), *WeatherShift* (WeatherShift et al., 2013), and '*Future weather generator*' (Rodrigues et al., 2023) can generate new weather files using the morphing method that transforms present-day time series data to estimate future weather data belonging to various climate change scenarios (Rodrigues et al., 2023). Morphing-based generators, also known as statistical downscaling methods, can capture local weather conditions with low computational power requirements (Moazami et al., 2019). Using such future weather files, many studies analyzed the impact of climate change by assessing energy consumption and adaptation measures (Tootkaboni et al., 2021; Berardi & Jafarpur, 2020; Zhai & Helman, 2019; Shen et al., 2019; Rey-Hernández et al., 2018), optimizing design parameters and strategies (Gercek & Durmuş Arsan, 2019; Baba et al., 2022; Flores-Larsen et al., 2019; Imam et al., 2023; Nguyen et al., 2021; D'Agostino et al., 2022), and adaptation strategies in building systems (Triana et al., 2018; Jafarpur & Berardi, 2021; Troup et al., 2019). However, these studies use weather files that are generated for discrete future years. For instance, *CCWorldWeatherGen* and *Future Weather Generator* generate files for the years 2020, 2050, and 2080, while *WeatherShift* generates files for 2035, 2065, and 2095. Therefore, it is not possible to estimate building performance for the in-between future years or perform a continuous assessment into the future.

Another challenge for evaluating climate change impact in cities is the broad spatial scale and the high number of buildings that performance analyses must target (Reinhart & Davila, 2016). Urban Building Energy Modeling (UBEM) is an approach that allows the modeling and analysis at the neighborhood or urban level and can provide computational insight into the energy demand of buildings (Hong et al., 2020). UBEM development can involve top-down or bottom-up approaches. In top-down UBEM, the analysis starts at the higher-level energy consumption, gradually examining individual buildings while considering historical data (Reinhart & Davila, 2016). As they are based on statistical and aggregated data, top-down UBEM falls short of predicting future trends or incorporating scenario analyses in any spatial or temporal detail (Ferrando et al., 2020).

Bottom-up UBEM, in contrast, can discretely model individual buildings, calculate their energy consumption, and aggregate bottom-level data upwards to the city level to be able to understand the energy demand in a neighborhood (Reinhart & Davila, 2016; Swan & Ugursal, 2009). Bottom-up UBEM approaches are categorized into physics-based (white box) models, reduced-order dynamic models (gray box), and data-driven (black box) models (Hong et al., 2020). Physics-based models can support various tasks, including design optimization, retrofit planning, and code compliance in high spatial and temporal resolution through simulation-based calculations (Ferrando et al., 2020). Nevertheless, these models require vast amounts of detailed data to characterize every building. Therefore, bottom-up UBEM development is cumbersome, and simulations are computationally demanding when applied at the urban scale (Ali et al., 2024; Hong et al., 2020; Pan et al., 2023; Koral Iseri et al., 2025). Data-driven models can alleviate many disadvantages of physics-based models if training data on building characteristics and the associated energy demand is provided (Ali et al., 2024; Hong et al., 2020). In various studies concerning energy demand prediction, different data-driven techniques have been used (Ali et al., 2024), including Support Vector Machine (SVM) (Guo et al., 2018; Paudel et al., 2017; Liu et al., 2020); Radial Basis Function Neural Network (RBFNN) (Tran et al., 2020); Random Forest (RF) (Pham et al., 2020; Wang et al., 2018; Cáceres et al., 2021; Chen et al., 2019); Artificial Neural Network (ANN) (Li et al., 2019), (Li et al., 2015; Alobaidi et al., 2018; Lin et al., 2022); Gradient Boosting Machine (GBM) (Touzani et al., 2018); Light Gradient Boosting Machine (Wang et al., 2021); Decision Trees (Ramos et al., 2022; Shcherbakov et al., 2013); Deep Neural Network (DNN) (Olu-Ajai et al., 2022). These machine learning algorithms have been applied to various use cases or

requirements, including calculating building energy consumption, heating energy consumption, and indoor overheating degree (Cáceres et al., 2021; Alabdai et al., 2018; Na & Wang, 2022; Todeschi et al., 2021; Kamel et al., 2020; Kontokosta & Tull, 2017; Afafia et al., 2021; Lei et al., 2021; Somu et al., 2021; Fayaz & Kim, 2018; Barbaresi et al., 2022).

Despite their widespread use, the majority of the studies utilizing data-driven techniques focus on individual buildings rather than conducting an urban-scale analysis (Chen et al., 2022). A limited number of studies combined urban-scale analyses with ML models. Ali et al. proposed an approach to predict the energy performance of residential buildings using both ensemble-based ML and end-use demand segregation methods (Ali et al., 2024). Tian et al. proposed a method that combines a physics-based model and a generative adversarial network (GAN) to predict the daily power demand for buildings on a large scale (Tian et al., 2022). Rahman et al. developed deep recurrent neural network models to predict hourly electricity consumption and used these models to forecast aggregate electricity consumption in residential buildings (Rahman et al., 2018). Jiang et al. proposed a semi-supervised deep neural network to predict energy use intensity (EUI) from unlabeled samples and tested it in Manhattan (Jiang et al., 2022). Ye et al. tested five GANs that predict building power demand at a large scale and evaluated their performance for power grid operation (Ye et al., 2022). Yu et al. predicted future annual electricity demand using a Radial Basis Function neural network based on possible future conditions of socio-economic development (Yu et al., 2015). To predict the life cycle building performance, Zou et al. generated future weather files for every year from 2020 to 2099 and performed simulations for each building on a parametric simulation tool (Zou et al., 2021). On the building scale, Tamer et al. proposed a multivariate regression-based prediction approach to quantify the climate change impacts on a typical office building across 81 cities (Tamer et al., 2022).

1.2. Gaps in the literature

While ML models have been widely applied at the building scale, only a few studies have integrated data-driven approaches into urban-scale UBEMs. Moreover, there has been limited consideration of climate change and long-term forecasting to support energy renovations using ML models with fast and precise predictive capacity. Furthermore, most previous research has purely focused on heating and cooling energy use, while naturally ventilated buildings and the thermal discomfort experienced in these buildings have been overlooked. However, there is much potential in the calculation of long-term future predictions on a yearly resolution that can allow cumulative calculations of performance metrics (i.e., the total heating or cooling energy use throughout successive years). ML can perform analyses with significantly less computation time and negligible compromise in precision. Predictive modeling towards future climatic conditions can allow decision-makers to understand how climate change affects buildings and which measures they can take.

This study presents a novel machine learning-based approach built on a zone-level Urban Building Energy Model (UBEM) of a residential neighborhood, aimed at fast and accurate prediction of heating energy consumption ($Q_{heating}$) and indoor overheating degree (IOD) under future climate change scenarios. The rest of this paper is structured as follows. The second section presents the methodology of UBEM development and ML-model training. The third section presents the results of the reliability of ML-based predictions, the performance of the ML models on unseen years, and the ML-based prediction and analysis of three retrofit packages between the years 2020 and 2080. Finally, Section 4 concludes this work by describing research limitations and pointing to future research directions. The dataset, code, and trained models presented in this paper are also released in a public GitHub repository for further research and to assist in reproducing the modeling and validation results (<https://github.com/metu-energy/energy-journ>

al).

2. Methodology

In this study, the ML-based method for predicting long-term building energy use and thermal comfort is introduced, which facilitates evaluating the (i) impact of climate change on existing buildings and (ii) building retrofit scenarios under long-term impacts. The ML models are trained and tested using the existing buildings' data in a residential neighborhood in Ankara, Turkey, as well as their energy simulation results. A zone-level UBEM of the neighborhood is developed to perform simulations. The method is developed based on three main steps (Fig. 1): (i) UBEM-based dataset development, (ii) ML model development, and (iii) Evaluation.

2.1. Study area

The studied neighborhood is selected as it primarily consists of residential buildings and represents a good mix of diverse building thermal characteristics due to different years of construction (years 1950–2022). The total area of the neighborhood is 0.57 km² and has 642 buildings, of which 593 are residential (that were modelled) and the remaining are commercial/public buildings. On average, the residential buildings feature 3.9 floors and 11.8 apartment units per building. These buildings also represent a well-established typology in Ankara, as the land sizes and proportions are very similar, and the zoning regulations mandating floor area ratio and building coverage ratios are the same.

2.2. UBEM-based dataset development

Dataset development consists of three steps: (i) Urban Building Energy Modeling, (ii) Future weather file generation, and (iii) Simulation-based data generation.

2.2.1. Urban building energy modeling

During UBEM development, our critical considerations are as follows:

High spatial resolution in UBEM leads to a more precise representation of the characteristics of model components. As a result, it is possible to enhance predictive accuracy from simulation results that are closer to the actual building behavior. At an urban scale, shoebox models are prevalent, as they are easier to model and computationally efficient (Dogan & Reinhart, 2017). However, shoebox models have limited accuracy as they neglect the zone-level variations, thereby inadequately representing the thermal loads. This is especially the case in building with many zones that have multiple zones that have different zonal occupancy patterns and user preferences. Therefore, each building, to the level of its zones with different characteristics, is modelled.

High data resolution enables the capture of various model details that can increase both the accuracy and precision of results. In the literature, the archetype-based modeling approach is commonly used, in which representative building typologies are identified with a distinct set of characteristics (Sokol et al., 2017). This simplified approach is advantageous when dealing with large building stocks as it eliminates the need for individualized and detailed models. However, a greater level of model fidelity is necessary for accurate simulation results, especially when dealing with zone-level energy models. Therefore, a non-archetype approach to modeling is pursued, where both zone-level and building-level data are identified, gathered, and used. The main challenge to be overcome in this situation is the reduced model reliability due to missing data. To this end, data from existing resources is used when available, and data augmentation and data generation methods are employed for missing data.

In the study, 593 residential buildings, containing 6452 zones, are modeled and simulated in the selected neighborhood (Fig. 2). The 3D building geometries and their programs are acquired from the city GIS

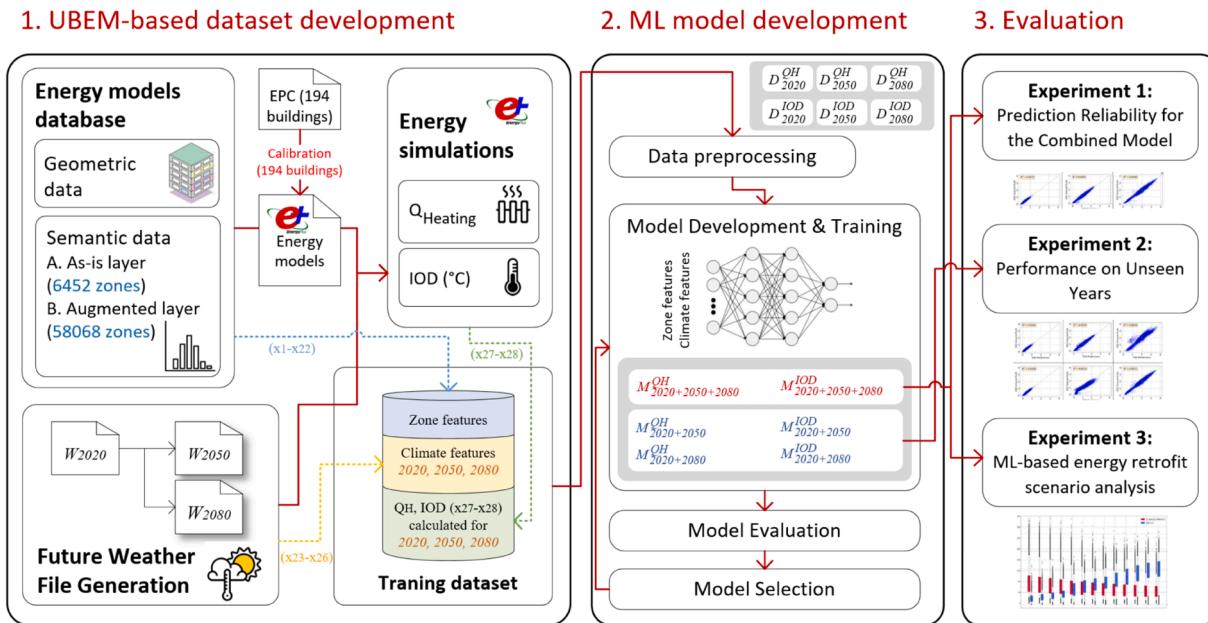


Fig. 1. An overview of the proposed approach.



Fig. 2. A satellite image of the neighborhood (left) and street view images of some example buildings (right) are considered in the paper.

database and the National Address Inquiry System, respectively. In each building, an apartment unit is modeled as a single thermal zone.

After 3D modeling, semantic data needs to be associated with the buildings and their zones (Fig. 3, Table 1). Semantic data is necessary because it provides important details and the contextual conditions that are necessary to build accurate energy models.

Building-level data includes parameters that apply uniformly to the building. These include envelope materials (U_{Roof} , U_{Wall} , $U_{Windows}$, U_{Ground}), boiler efficiency (η_{boiler}), window-wall ratio (WWR) in four directions, Solar Heat Gain Coefficient (SHGC), and facade infiltration rate. Zone-level data, on the other hand, varies from zone to zone. This data pertains to zone internal loads (lighting power density (LPD), equipment power density (EPD), people density (PPD)) and heating setpoint ($T_{Heating}$).

The semantic data, both building-level and zone-level, comprises two layers:

The as-is data layer. The as-is data layer aims to represent the buildings' as-is conditions. Therefore, the main data source is the existing documentation of the buildings in the area. The Energy Performance Certificates (EPC) of 194 buildings out of 593 (corresponding to 66 % missingness) are acquired, from which their envelope U-values and WWR (x_1-x_8 in Table 1) are extracted and assigned to the corresponding buildings. For the remaining 399 buildings, new data are generated using data imputation. Kernel density estimation (KDE), a Normalised Prediction Distribution Errors (NPDE) technique, is employed, which estimates a density distribution by averaging over kernel homogeneous functions that are centered at each sample point (Kamalov, 2020). In KDE, the probability density for the distribution is

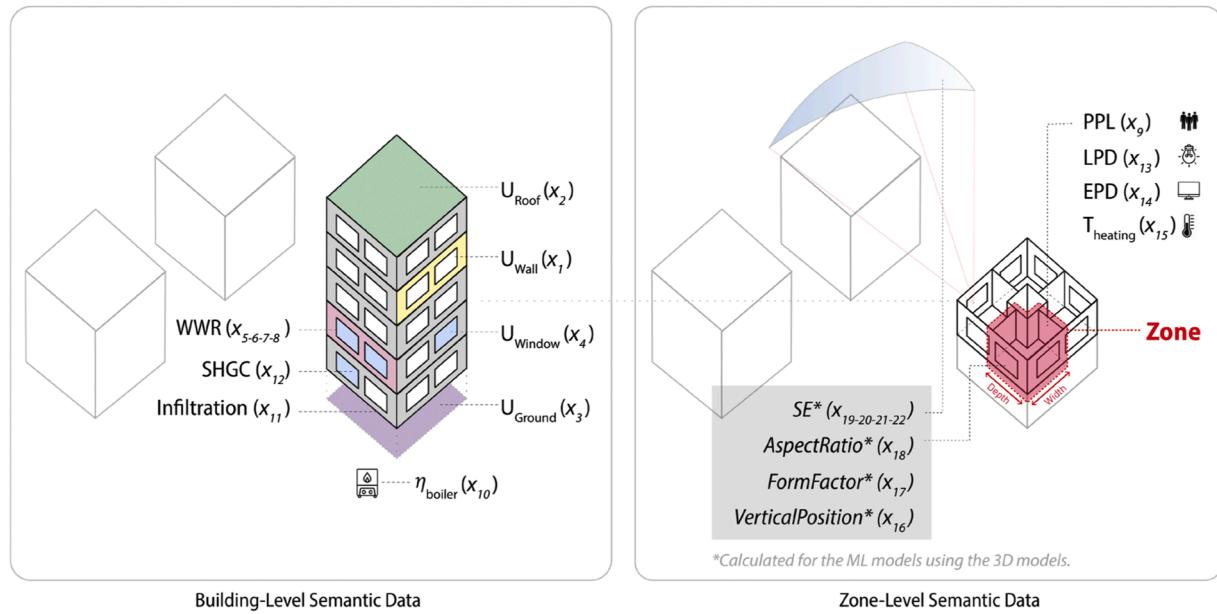


Fig. 3. Building-level (left) and zone-level (right) semantic data.

Table 1
Semantic data and parameter distributions.

Name	Unit	Assigned to	Data source	AS-IS DATA LAYER (N = 6452)			AUGMENTED DATA LAYER (N = 64,520)			
				Sample size	Distrib. type	Generated instances	Data source	Distrib. type	Generated instances	
x ₁	U _{wall}	W/m ² K	B	EPC	194	KDE	399	U(0.15, 2.81)	5337	
x ₂	U _{roof}	W/m ² K	B	EPC	194	KDE	399	U(0.15, 3.66)	5337	
x ₃	U _{ground}	W/m ² K	B	EPC	194	KDE	399	for the minimum value	U(0.15, 3.55)	5337
x ₄	U _{window}	W/m ² K	B	EPC	194	KDE	399	U(0.85, 3.65)	5337	
x ₅	WWR _N									
x ₆	WWR _S									
x ₇	WWR _E									
x ₈	WWR _W									
x ₉	PPL	ppl//m ²	Z	(TURKSTAT, 2021)	NA	KDE	6452	U(0.1, 0.42)	5337	
x ₁₀	η _{boiler}	%	B	(Boait et al., 2012; Ala et al., 2019)	NA	U(0.8, 0.95)	593	(Boait et al., 2012; Ala et al., 2019)	U(0.8, 0.95)	5337
x ₁₁	Infiltration	m ³ /s-m ²	B	(ANSI/ASHRAE/IES Standard 90.1, 2023)	NA	U(0.000285, 0.0005)	593		U(0.000285, 0.0005)	5337
x ₁₂	SHGC		B	(Energy Performance Data, 2023)	NA	U(0.3, 0.85)	593	(Energy Performance Data, 2023)	U(0.3, 0.85)	58,068
x ₁₃	LPD	W/m ²	Z	(Pohoryles et al., 2020)	NA	U(8, 12)	6452	(Ahmed & Asif, 2020; Ang et al., 2022)	U(2.5, 12)	58,068
x ₁₄	EPD	W/m ²	Z	(Brohus et al., 2009)	NA	U(1.21, 2.93)	6452	(Ahmed & Asif, 2020; Nagpal & Reinhart, 2018)	U(1.75, 8)	58,068
x ₁₅	T _{heating}	°C	Z	(Akgüç & Yilmaz, 2022)	NA	U(20,24)	6452	(Akgüç & Yilmaz, 2022)	U(20,24)	58,068

NA: Not available, KDE: Kernel density estimator; U: Uniform distribution; EPC: Energy Performance Certificate; B: Building; Z: Zone.

assumed, which would be decisive for further predictions if the data were constrained to fall within a particular parametric family. The kernel-based estimator is defined as follows:

$$f(x) = \frac{1}{Nh} \sum_{i=1}^N K\left(\frac{x - x_i}{h}\right), \quad (1)$$

where $K(\cdot)$ is e.g. the Gaussian kernel with h as the bandwidth. In other words, the kernel-based estimator is a sum of the distributions placed at the observations with the kernel function $f(x)$ for the shape of the dis-

tributions and the bandwidth for their width.

For x_9 , National statistics for household densities are acquired from the Turkish Statistical Institute (TURKSTAT), and KDE is applied to generate people density values for all zones. Table 1 shows the distribution types of all parameters, and Fig. 4 shows the KDE distributions. The other parameters ($x_{10}-x_{15}$) are fully unavailable (missingness=100 %); therefore, the existing literature (Table 1) is used that is based on parametric probability density estimation (PDE) to generate new data. Uniform distributions are used, where the distribution of a random variable x is described as:

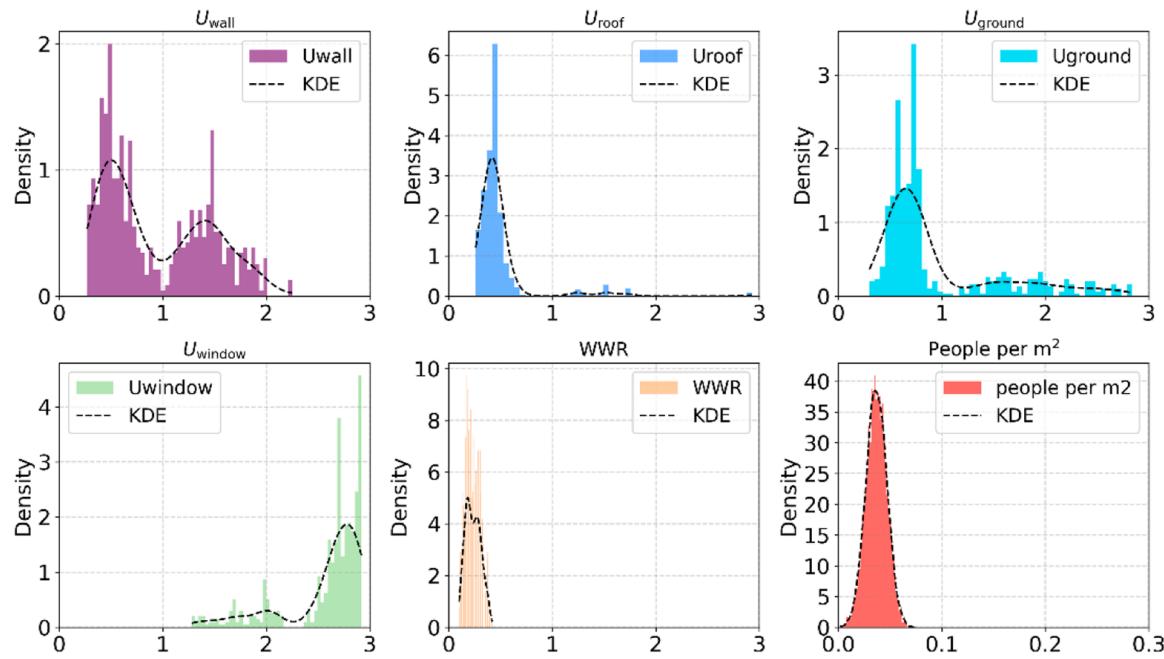


Fig. 4. KDE Distributions of features x_1 - x_9 , generated using the data from the existing buildings' Energy Performance Certificates.

$$P(a < X < b) = \int_a^b f(x) dx, \text{ for all } a < b \quad (2)$$

Augmented data layer. The as-is layer represents the existing building and zone parameters; however, it does not yet contain all possible values in the neighborhood, as its missingness = 66 %. It is also important to note that the as-is layer exhibits an uneven distribution, with sharp peaks and large regions of the dataset remaining sparse or blank. This type of distribution is problematic, as model training becomes biased towards the regions where data is dense, which can lead to overfitting in these areas. As a result, the model might learn to make highly accurate predictions for these densely populated regions (the majority class), but fail to generalize to the areas where data is sparse. This becomes a critical problem for decision-making, as the predictions for the underrepresented cases may be inaccurate or even completely missed. This situation is also problematic in the estimation of high-performance building interventions during retrofit decision-making, as the as-is data layer model does not contain the parameter values of such systems. Therefore, a second semantic data layer is created using data augmentation, where wider value ranges are used to enhance the diversity of the dataset. New data is generated with uniform distributions to ensure that the dataset is consistently random without any bias. For the envelope U-values (x_1 - x_4), the Passivhaus standard is used (McLeod et al., 2012) as the minimum values, while the maximum values remain the same as the as-is layer. For x_{12} - x_{15} , the existing literature is used that focuses on residential building retrofit (Energy Performance Data, 2023; Pohoryles et al., 2020; Ahmed & Asif, 2020; Ang et al., 2022; Brohus et al., 2009; Nagpal & Reinhart, 2018; Akgül & Yilmaz, 2022). The remaining parameter ranges remained the same, as they already represent a wide enough range. This augmented layer also significantly expands the scale of the dataset, as nine data instances for each zone and building are generated. As such, the size of the dataset increased ten-fold (5930 buildings and 64,520 zones).

The resulting dataset, which included both the as-is and augmented data layers, which are used both for energy modeling and ML model training.

2.2.1.1. Simulation setup. All buildings are heated using natural gas boilers with different efficiencies, which is representative of buildings in

Turkiye, where natural gas is the most commonly used fuel (IEA 2017). Cooling energy consumption is not accounted for, as mechanical cooling is rarely used in the neighborhood. Additionally, mechanical ventilation is provided following the ANSI/ASHRAE Standard 62.1–2019 with People Outdoor Air Rate = 2.5 L/s-person plus Area Outdoor Air Rate = 0.3 L/s·m² (ASHRAE 2015). All zones are naturally ventilated through operable windows with an opening fraction of 40 % when $T_{indoor} > 24.5^\circ\text{C}$ and $T_{indoor} > T_{outdoors}$, following (Dino & Meral Akgül, 2019). Due to data privacy concerns, it was not possible to access detailed occupant-related schedules for each zone at the urban scale. Therefore, the heating, occupancy, equipment, and lighting schedules of the prototype mid-rise apartment building model defined in the ANSI/ASHRAE/IES Standard 90.1 are used (ANSI/ASHRAE/IES Standard 90 2023). This decision is consistent with other literature focusing on building energy consumption studies in Turkiye (Noaman & El-Ghaour, 2024; Tomrukcu & Ashrafiyan, 2024). Context shading is also critical for UBEM, especially in densely built environments, as it influences the calculation of solar heat gain through the envelope and the associated heating/cooling loads. However, shading is rarely considered in existing UBEMs, reducing the reliability of simulation results. In the studied UBEM, each building is modeled as a shading element to other buildings if it falls within a predefined range. Context shading is considered an input feature in our ML models (See Section 2.3.1).

2.2.2. Future weather file generation

The simulations for the current weather conditions are performed using a typical meteorological year (TMY) file of Ankara obtained during 2014–2018. The data in this TMY file is collected within the city, which already considers the microclimate specific to that local urban context. For the sake of convenience, this weather file is denoted with W_{2020} . Two new weather files (W_{2050} and W_{2080}) are generated for the 2050 and 2080 climatic conditions using the 'Future Weather Generator'. 'Future Weather Generator' is an existing tool that forecasts 8760-h weather parameters by transforming current local weather data in the form of EPW files into future scenarios using the output of nine general circulation models (GCM) environmental variables that contributed to the 6th IPCC Assessment Report (2022) (Rodrigues et al., 2023). The tool implements four Shared Socioeconomic Pathways (SSP), including SSP1–2.6, SSP2–4.5, SSP3–7.0, and SSP5–8.5. 'Future Weather

Generator' modifies dry bulb temperature, dew point temperature, relative humidity, atmospheric pressure, global horizontal radiation, wind speed, and variables with precipitation. The SSP5–8.5 scenario (the worst-case scenario) was considered during the generation of these files to ensure that buildings are prepared for the most extreme weather conditions during decision-making. These three weather files are used during simulations.

Using the hourly data in these files, four weather metrics are calculated: annual average dry bulb temperature ($T_{outside}$), Heating Degree Days ($HDD_{18,3}$), Cooling Degree Days ($CDD_{23,3}$), and annual average global horizontal radiation (GHR) (Table 2). The results indicate a consistent warming trend, with $T_{outside}$ increasing from 13.37 to 18.20 °C from 2020 to 2080 (Fig. 5). The calculated HDD and CDD values also indicate a decrease of 32.1 % and an increase of 260.5 % from 2020 to 2080 respectively. Global horizontal radiation (GHR) values remained relatively steady during the same period. These weather metrics are used as input features in the training of ML models, where each year (2020, 2050 and 2080) is characterized by these four weather metrics. These values are also used in our training and testing datasets (see Section 3.1).

2.2.3. Simulation-based data generation

Finally, the developed energy models that are built upon the as-is and augmented data layers are used to calculate performance objectives through energy simulations. As the simulation engine, EnergyPlus V9.2 is used through the Grasshopper Ladybug Tools interface (Roudsari & Pak, 2013). Heating energy use ($Q_{heating}$) and indoor overheating degrees (IOD) are calculated in each zone using W_{2020} , W_{2050} , and W_{2080} . The calculated values will be the output layer of our ML models (Section 2.3.1):

$Q_{heating}$: The annual heating energy consumption in kWh/m² for each residential zone is calculated. The boiler efficiency (η_{boiler}) values presented previously in the dataset are used to account for the heating system in a building.

IOD : As the buildings in the selected neighborhood are naturally ventilated, the indoor overheating degree (IOD) is used to quantify the degree of indoor thermal stress that a residential zone will experience. An existing method, which considers both the intensity and frequency of overheating, is used for residential buildings (Hamdy et al., 2017). In this method, intensity is considered as the difference between the indoor temperature and an upper threshold for comfort (T_{comf}). First, the comfort neutrality temperature (T_n) is calculated as in Eq. (3), where T_{ave} is the monthly mean outdoor air temperature, calculated according to ASHRAE 55–2013 (Ashrae, 2013). The upper-temperature limit ($T_{comf,i,z}$) of thermal comfort for zone z during the occupied hour i is then calculated using Eq. (4). Finally, IOD is calculated as in Eq. (5), where N : total hours in a given calculation period (1 May- 30 September), t : time step (1 h), $T_{oper,i,z}$: operative temperature at the time step i in the zone z .

$$T_n = 0.31T_{ave} + 17.8, \quad (3)$$

$$T_{comf,i,z} = T_n + 2.5, \quad (4)$$

$$IOD = \frac{\sum_{i=1}^{N_{occ(z)}} [\max\{T_{Oper,i,z} - T_{comf,i,z}, 0\} \times t_{i,z}]}{\sum_{i=1}^{N_{occ(z)}} t_{i,z}}. \quad (5)$$

Table 2

Weather input features calculated from W_{2020} , W_{2050} , and W_{2080} .

	Year (y)				
	2020	2050	2080	Regression equation	R ²
HDD (x_{23})	2499.4	2017.4	1697.7	$HDD_y = 13.362y + 29463$	0.9865
CDD (x_{24})	259.6	596.9	934.6	$CDD_y = 11.25y - 22466$	1
$T_{outside}$ (x_{25})	13.4	16.1	18.2	$T_{out,y} = 0.0805y - 149.15$	0.9954
GHR (x_{26})	197.9	206.3	213.3	$GHR_y = 0.2572y - 321.4$	0.9976

2.2.3.1. Model validation. To validate the energy models, the actual heating energy use ($Q_{heating,EPC}$) values documented in the 194 buildings with EPCs are benchmarked against the EnergyPlus simulation results ($Q_{heating}$) in the as-is layer (Fig. 6). As mentioned before, there might exist different levels of uncertainties in energy models due to model parameters that have an influence on energy use but are not documented anywhere. These parameters correspond to parameters x_9-x_{15} in the proposed framework. To minimize these parameters' influence on simulation results, the following steps are followed: First, using a correlation matrix, $T_{heating}$ (x_{15}) and $Infiltration$ are identified as the parameters with the most direct influence on $Q_{heating}$. For each remaining parameter (x_9-x_{14}), the mean value of the value range shown in Table 1 is calculated, and this value is used in the building. Finally, $T_{heating}$ and $Infiltration$ are varied within the value range specified in the same table to minimize the absolute error between actual and simulated values.

The calibration error, which is defined as the percentage error (PE) between the $Q_{heating,EPC}$ and $Q_{heating}$, is used to evaluate model performance to represent the aggregate and individual building demands of the neighborhood (See Eq (6)):

$$PE = \frac{Q_{heating,EPC} - Q_{heating}}{Q_{heating,EPC}} \times 100 \% \quad (6)$$

PE is calculated as 1.65 %, indicating that the urban level heating demand can be calculated quite accurately. The mean absolute percentage error (MAPE) of 3.17 % has been observed, with a minimum and maximum absolute percentage error of 0.0002 % and 39.23 %, respectively. Focusing on the latter value, high error values on the individual building level can be expected due to the high level of uncertainty resulting from missing data related mostly to occupant preferences, infiltration rates, and heat system efficiency. Similar discrepancies were also widely reported in other urban-scale energy models. Lower levels of data missingness can reduce errors and improve the reliability of UBEMs.

Results

Using EnergyPlus V9.2, 5930 buildings (593 buildings from the as-is and 5337 buildings from the augmented layers) for each year (2020, 2050, and 2080) are simulated. As a result, 193,560 zone's $Q_{heating}$ and IOD values are calculated. The entire simulation process took 12,346 min (205.7 h), corresponding to approximately 0.064 min (3.82 s) per zone. These simulations were performed with a single CPU belonging to TRUBA (Turkish National High Performance Computing Center). The server used for this task is a Dell R640 machine, with two Intel Xeon 6148 2.40 GHz processors and 384 GB RAM, running CentOS Linux v7.

Figs. 7 and 8 show the 3D view and the data distribution of the results in 2020, 2050, and 2080. The results point to a steady increase in IOD from 0.35 °C to 1.94 °C due to the increasing outdoor temperatures, while $Q_{heating}$ decreased from 79.02 to 47.0 kWh/m². Expectedly, the highest indoor air temperatures are observed in the apartment unit with the highest IOD , with 34.75 °C, 39.22 °C, and 42.68 °C in 2020, 2050, and 2080, respectively.

Moreover, while the spreads of $Q_{heating}$ in 2020, 2050, and 2080 are similar and overlap to a large degree, non-overlapping distributions between the three years in IOD are observed. The implications of these distributions on the performance of ML-based predictions for the years lying between 2020, 2050 and 2080 will be discussed in Section 3.2.

2.3. Development of ML models

2.3.1. Dataset description and preprocessing

In the study, ML models aim to predict zones' $Q_{heating}$ and IOD for any year between 2020 and 2080. In the previous section, the development of training and testing datasets through energy modeling and simulations using W_{2020} , W_{2050} , and W_{2080} is described. To train the ML models, 22 features that describe the zones/buildings (x_1-x_{22}), and four climate features ($x_{23}-x_{26}$) of the years 2020, 2050, and 2080 extracted

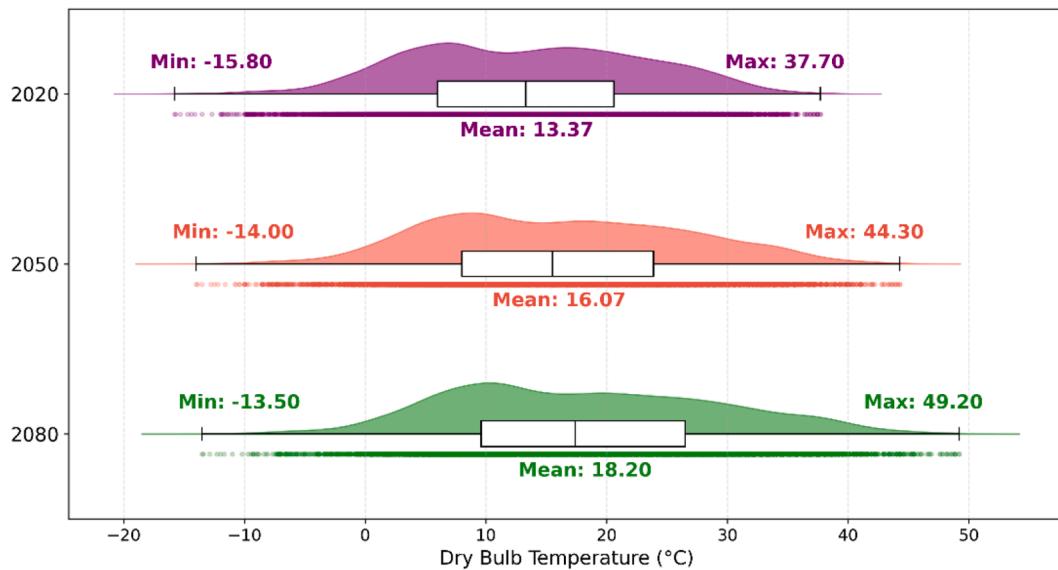


Fig. 5. The distribution of hourly $T_{outside}$ values in W_{2020} , W_{2050} , and W_{2080} .

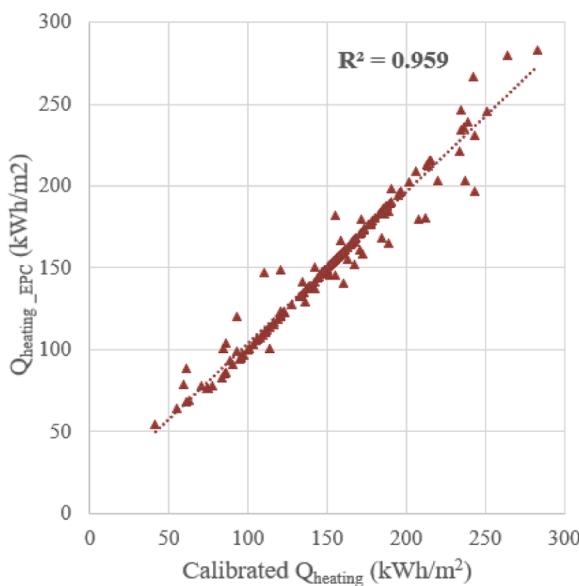


Fig. 6. The comparison of EPC-based and calibrated heating energy use values ($n = 194$).

from W_{2020} , W_{2050} , and W_{2080} are used. Amongst the zone/building inputs, the first 15 features were already explicitly used in UBEM development (see the previous section). Additional features that are used as input to ML models are x_{16} - x_{18} , which describe zone geometries, and x_{19} - x_{22} , which describe the level of context shading that the zone is exposed to. These parameters are calculated separately using the 3D model:

- verticalPosition (x_{16}) is a categorical feature that indicates where the zone is vertically located in a building (0=ground, 2=roof, 1=otherwise). This feature is important as verticalPosition influences both the level of context shading (lower levels are subject to more shading), envelope materials, and the different boundary conditions, resulting in different levels of conductive heat loss from surfaces.
- formFactor (x_{17}) indicates the ratio between the external surface area and total volume. The form factor directly influences conductive heat

gain/loss, as the higher surface area results in increased heat flow through the envelope.

- aspectRatio (x_{18}) is the ratio of a zone's width (x -direction) to its length (y -direction). The aspect ratio has an impact on solar exposure, as different facade orientations are subject to varying levels of incoming solar radiation.
- sky exposure (SE) on all sides (x_{19} - x_{22}) is the fraction of the sky hemisphere that is visible from the center point of the window, as calculated by Ladybug Tools. This feature is influential on both $Q_{heating}$ and IOD , as lower SE results in reduced solar heat gains through windows.

The climate input features consist of HDD , CDD , $T_{outside}$, and GHR (x_{23} - x_{26}) for W_{2020} , W_{2050} , and W_{2080} . Each weather file is used in the simulation of 64,520 zones. For each input sample belonging to a zone simulated for the weather file W_{year} , the four climate features of that year (HDD_{year} , CDD_{year} , $T_{outside_year}$, and GHR_{year}) are appended to that sample. This means that the same four climate input feature values repeat in the dataset for the simulations made for the same year. Although this brings about redundancies in our dataset, it is still necessary to associate the output features with the specific year's climatic features with which they are calculated.

Lastly, the calculated $Q_{heating}$ and IOD (y_1 - y_2) values are used as the output layer of ML models. The input and output features are shown in Table 3. The input feature distributions are shown in Fig. 9. A correlation matrix is presented in Fig. 10 which shows the effects of input parameters (in Table 3) on output parameters ($Q_{heating}$ and IOD). This matrix is constructed in Seaborn (Waskom, 2024) by calculating the pairwise correlation coefficients between variables in the dataset, resulting in a square matrix where each cell represents the correlation between two variables. The matrix uses the Spearman correlation coefficient (ρ), which measures the monotonic relationship between two variables. The correlation matrix indicates that *formFactor* and *verticalPosition* yielded the highest correlation with $Q_{heating}$ ($\rho=0.63$ and 0.38). This is due to the high conductive heat loss from zones with large external surfaces (walls and roofs), especially with high *uwall* and *uwindow* values. Expectedly, $Q_{heating}$ is highly correlated to $T_{heating}$ ($\rho=0.46$). In IOD , high positive correlations are observed with the same two parameters (*formFactor* and *verticalPosition*) with $\rho=0.25$ and $\rho=0.62$ respectively. *SHGC* and *SESouth* showed high correlations with IOD due to high solar heat gain through the windows ($\rho=0.22$ and $\rho=0.28$, respectively). The results also suggest that a strong positive

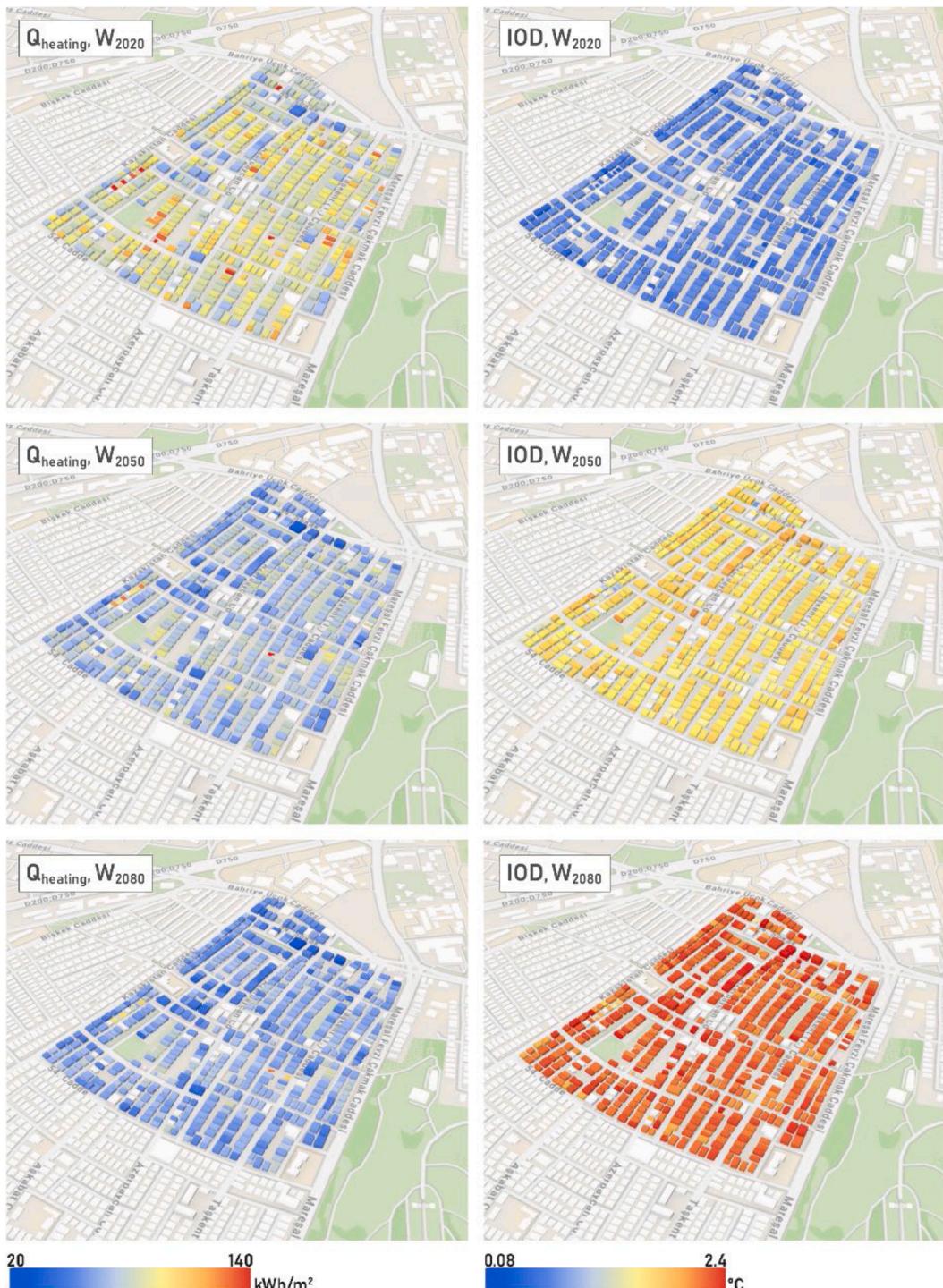


Fig. 7. The change in all residential buildings' Q_{heating} and IOD for W₂₀₂₀, W₂₀₅₀, and W₂₀₈₀. Although Q_{heating} and IOD are calculated per zone, these were recalculated for each building while aggregating and averaging the zone-level results.

correlation with u_{roof} and IOD due to the roof surfaces' high solar exposure. Internal loads due to equipment, lighting, and people showed weak positive and weak negative correlation with IOD and Q_{heating} , respectively.

The input data for ML model training is pre-processed following the common practices: The *verticalPosition* parameter is encoded using one-hot-encoding, which increases the number of input features to 28. Except for *verticalPosition*, the rest of the input dataset is scaled using a standard scaler as follows:

$$\hat{x} = \frac{(x - \mu)}{\sigma}, \quad (7)$$

where \hat{x} is the normalized value of the feature value x ; μ and σ are the mean and standard deviation of the feature over the training set.

The final dataset consisting of 193,560 samples is generated using the simulation of each building for each weather file W₂₀₂₀, W₂₀₅₀ and W₂₀₈₀. This dataset is split into three sets for training (70 %), validation (15 %), and testing (15 %), following common practice.

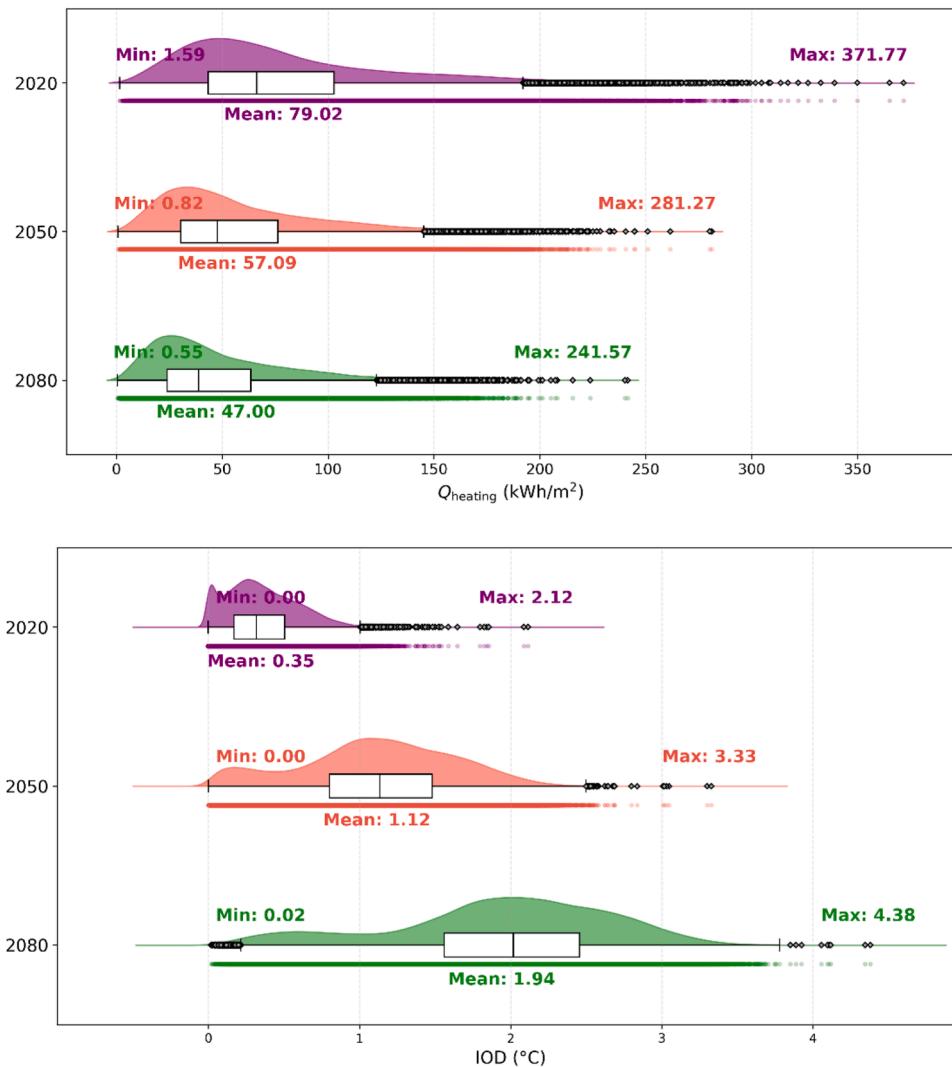


Fig. 8. Zone Q_{heating} (top) and IOD (bottom) in 2020, 2050, and 2080 ($n = 64,520$ for each year).

2.3.2. Model development

Multi-layer Perceptrons (MLPs), which are artificial neural networks commonly used in practice (IEA, 2017), are chosen as the ML model. MLPs are known to be universal approximators, i.e., having the capacity to represent any arbitrary function with a sufficient number of hidden layers and neurons (Hornik et al., 1989). Due to their strong capacity and ease of customization and training, the literature has dominantly used MLPs for data-driven energy prediction (Seyedzadeh et al., 2019; Amasyali & El-Gohary, 2018): MLPs are used in 47 % of studies, whereas SVM with 25 %, statistical learning algorithms with 24 % and decision trees with 4 %. Therefore, MLPs are selected for this study.

In our problem, a dataset $D := (x_i, y_i)$ consisting of input ($x_i \in \mathbb{R}^D$) and target ($y_i \in \mathbb{R}^C$) values are given and the mapping $y_i = f(x_i; \theta)$ is to be learned. The function $f(\cdot; \theta)$ is a parametric one, and its parameters θ are optimized through a training procedure.

2.3.3. Multi-Layer perceptrons (MLPs)

An MLP is essentially a stack of layers wherein each layer consists of a set of small processing units, called (artificial) neurons, which can represent “simple” functions. Denoting a variable in the l^{th} layer by $(\cdot)^{[l]}$, the j^{th} neuron in the l^{th} layer performs the following transformation:

$$a_j^{[l]} = \sigma(W_j^{[l]} \cdot a^{[l-1]} + b_j^{[l]}) \quad (8)$$

where $\sigma(\cdot)$ is a non-linearity (in our case, the Rectified Linear Unit, a.k.a., ReLU (Nair & Hinton, 2010)), and $W_j^{[l]}$ and $b_j^{[l]}$ denote the parameters of the neuron. The 0^{th} layer is the input (the i^{th} sample x_i), and the last layer the estimated output (\hat{y}_i):

$$\begin{aligned} a^{[0]} &= x_i, \\ a^{[L]} &= \hat{y}_i, \end{aligned} \quad (9)$$

with L denoting the number of layers in the MLP model.

The set of parameters in MLP, i.e., $\theta = \cup_{i,l} \{W_j^{[l]}, b_j^{[l]}\}$ are optimized to minimize the following objective (loss) function:

$$\mathcal{L}_i = \frac{1}{2} (\hat{y}_i - y_i)^2, \quad (10)$$

where \hat{y}_i is the predicted output variable for the i^{th} sample x_i and y_i is the correct value. The parameters θ are randomly initialized to small values ($\theta_0 \leftarrow$ small random values) at the beginning and updated iteratively using gradient descent:

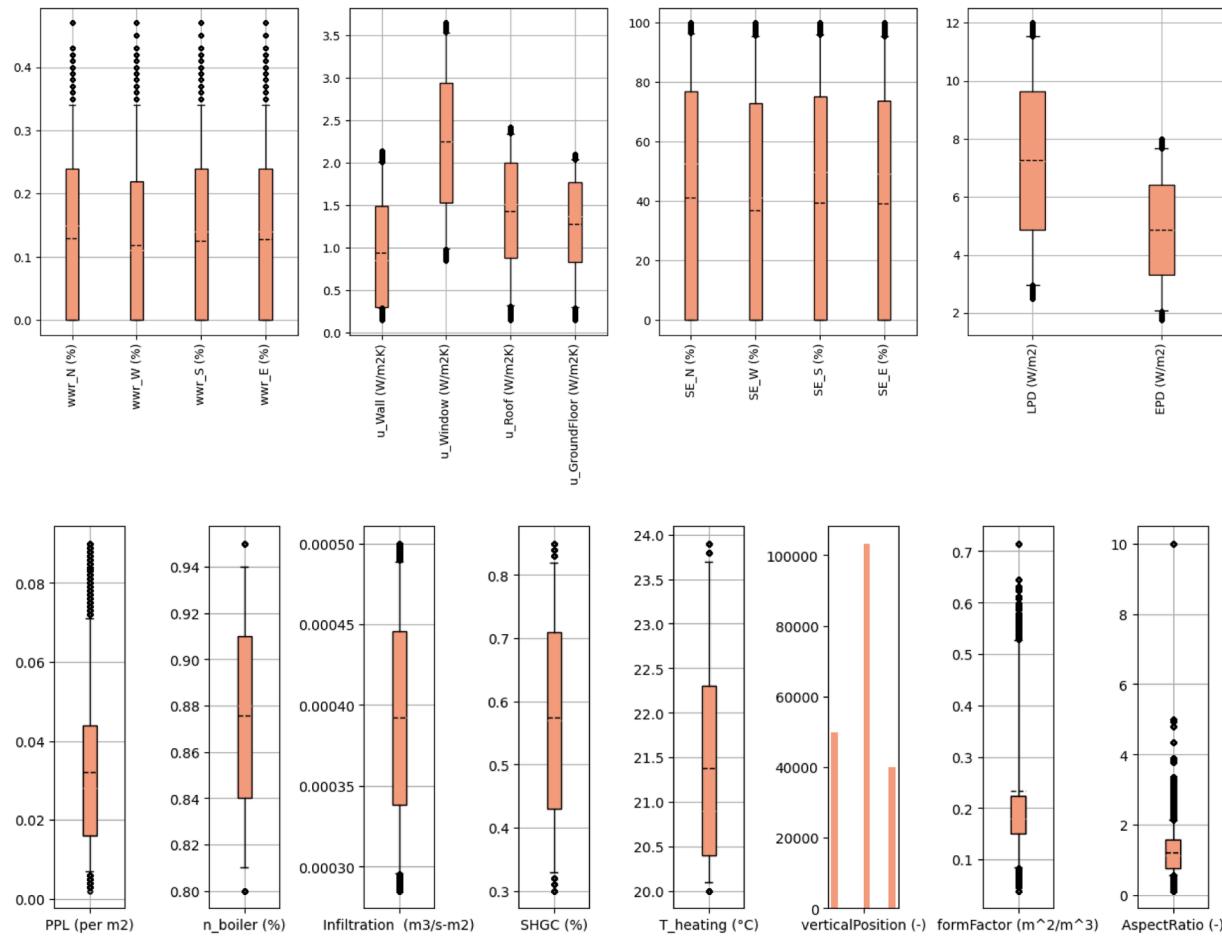
$$\theta_{t+1} \leftarrow \theta_t - \eta \frac{\partial \mathcal{L}}{\partial \theta_t}. \quad (11)$$

For the two regression tasks for Q_{heating} and IOD estimation, two

Table 3

The input and output features used in the ML models.

	Abbreviation	Unit	Source	Type	Min	Max	
Zone and building level input features	x ₁	U _{wall}	W/m ² K	EPC	Continuous	0.15	1.9
	x ₂	U _{roof}	W/m ² K	EPC	Continuous	0.147	2.42
	x ₃	U _{ground}	W/m ² K	EPC	Continuous	0.145	2.1
	x ₄	U _{window}	W/m ² K	EPC	Continuous	0.85	3.65
	x ₅	WWR _N		EPC	Continuous	0	0.47
	x ₆	WWR _S		EPC	Continuous	0	0.47
	x ₇	WWR _E		EPC	Continuous	0	0.47
	x ₈	WWR _W		EPC	Continuous	0	0.47
	x ₉	PPL	people/m ²	Distributions	Continuous	0.002	0.09
	x ₁₀	η _{boiler}		Distributions	Continuous	0.8	0.95
	x ₁₁	Infiltration	m ³ /s·m ²	Distributions	Continuous	0.000286	0.0005
	x ₁₂	SHGC		Distributions	Continuous	0.301	0.849
	x ₁₃	LPD	W/m ²	Distributions	Continuous	2.5	12
	x ₁₄	EPD	W/m ²	Distributions	Continuous	1.75	8
	x ₁₅	T _{heating}	°C	Distributions	Continuous	20	24
	x ₁₆	verticalPosition		3D Model	Categorical	0	2
	x ₁₇	formFactor	m ² /m ³	3D Model	Continuous	0.039	0.715
	x ₁₈	aspectRatio		3D Model	Continuous	0.115	10
Climate input features	x ₁₉	SE _N	%	3D Model	Continuous	0	100
	x ₂₀	SE _W	%	3D Model	Continuous	0	100
	x ₂₁	SE _S	%	3D Model	Continuous	0	100
	x ₂₂	SE _E	%	3D Model	Continuous	0	100
	x ₂₃	T _{outside}	°C	EPW file	Continuous	13.37	18.2
	x ₂₄	HDD	°C	EPW file	Continuous	1697.7	2499.4
	x ₂₅	CDD	°C	EPW file	Continuous	259.57	934.6
Output features	x ₂₆	GHR	W/m ²	EPW file	Continuous	197.88	213.31
	y ₁	Q _{heating}	kWh/m ²	Simulation	Continuous	1.591	371.769
	y ₂	IOD	°C	Simulation	Continuous	0	2.115

**Fig. 9.** Distributions of the input dataset (x₁ - x₁₈). See Table 3 for details.

	-U _{wall}	-U _{roof}	-U _{groundfloor}	-U _{window}	-WWr _N	-WWr _S	-WWr _E	-WWr _W	-PPL	-nboiler	-Infiltration	-SHGC	-LPD	-EPD	-T _{heating}	-verticalPosition	-formFactor	-SE _N	-SE _W	-SE _S	-SE _E
Q _{heating}	-0.26	0.53	-0.05	0.12	0.18	0.05	0.05	0.02	-0.11	-0.07	0.08	-0.10	-0.09	-0.14	0.46	0.38	0.63	0.23	0.06	0.10	0.10
IOD	-0.02	0.30	-0.57	-0.02	-0.08	0.13	0.02	0.03	0.04	0.01	0.00	0.22	0.06	0.07	0.03	0.62	0.25	0.09	0.18	0.28	0.18

Fig. 10. Correlation matrix between the UBEM parameters (horizontal axis) and performance objectives (vertical axis).

different MLP models (Seyedzadeh et al., 2019) are used and trained separately to optimize mean square error for their target values ($Q_{heating}$ or IOD).

2.4. Model training, evaluation and selection

2.4.1. Model training

As highlighted in the previous section, the networks are trained to minimize the mean-squared error of the predictions for both regression tasks as defined in Eq. (10). To this end, we use the ADAM optimizer (Kingma & Ba, 2014) with L2 regularization (with regularization coefficient set to 6e-6).

2.4.2. Performance metrics

For comparing the models, the following common regression metrics are used:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}, \quad (12)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}, \quad (13)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2, \quad (14)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (15)$$

where \hat{y}_i is a prediction, y_i is its correct target value and \bar{y} is the mean of correct values. Note that RMSE is a lower-better whereas R^2 is a higher-better measure.

2.4.3. Model selection

The ML model described in Section 2.3 has certain hyperparameters that require careful tuning:

- The number of layers: [3, 4, 5, 6],
- The number of neurons: [16, 32, 48, 64],
- The learning rate: Uniform(10^{-6} , 10^{-2}),
- Batch size: [32, 64, 128, 256].

For each model, these hyperparameters are tuned with the Bayesian Hyperparameter Optimization (BHO) tool provided by the Weights & Biases framework. For each model, 20 experiments (trials) turned out to be sufficient to find a good combination of hyperparameters for our tasks. The best model among these 20 experiments is selected using the R^2 measure on the validation set (recall from Section 2.3.1 that the dataset is split into 70 % training, 15 % validation, and 15 % testing sets). As a result of this selection process, we have obtained the hyperparameter values as listed in Table 4.

We have also evaluated the commonly-used activation functions

Table 4
Hyperparameter configuration of ML models.

ML Model	Number of Layers	Number of Neurons	Learning Rate	Batch Size
$M_{2020+2050+2080}^{QH}$	4	64	0.0027359	256
$M_{2020+2050}^{QH}$	5	64	0.0044480	256
$M_{2020+2080}^{QH}$	3	48	0.0016671	32
$M_{2020+2050+2080}^{IOD}$	6	64	0.0052804	128
$M_{2020+2050}^{IOD}$	3	32	0.0042791	32
$M_{2020+2080}^{IOD}$	3	16	0.0064574	256

(Rectified Linear Unit – ReLU (Nair & Hinton, 2010) – and Leaky Rectified Linear Unit – Leaky ReLU (Maas et al., 2013)) in hidden and output layers of MLP. ReLU provided slightly better results overall and is used in all models.

3. Results

In this section, the performance of the ML models trained and tested on data from different years will be analyzed. To make it easier to follow, D_{years}^Y to denote a dataset comprising of the Y data ($Q_{heating}$ or IOD) from years and M_{years}^Y to denote an ML model trained on D_{years}^Y will be used. For example, $D_{2020+2080}^{IOD}$ is the IOD dataset with samples from 2020 to 2080, and $M_{2020+2080}^{IOD}$ is the MLP model trained on $D_{2020+2080}^{IOD}$. To assess the effectiveness of M_{years}^Y , various evaluation metrics (Eqs. (11)–(14)), which provide a holistic view of the model performance, are used. Each result is obtained by averaging the performances of 10 randomly initialized versions of the models. Whenever possible, Welch's t -test is used to assess significance.

In the experiments, the following research objectives will be addressed (ROs):

RO1: What is the prediction reliability for the combined model $M_{2020+2050+2080}$ in forecasting $Q_{heating}$ and IOD ? (Section 4.1)

RO2: What are the performances for the models $M_{2020+2050}$ and $M_{2020+2080}$ in predicting $Q_{heating}$ and IOD on unseen years? (Section 4.2)

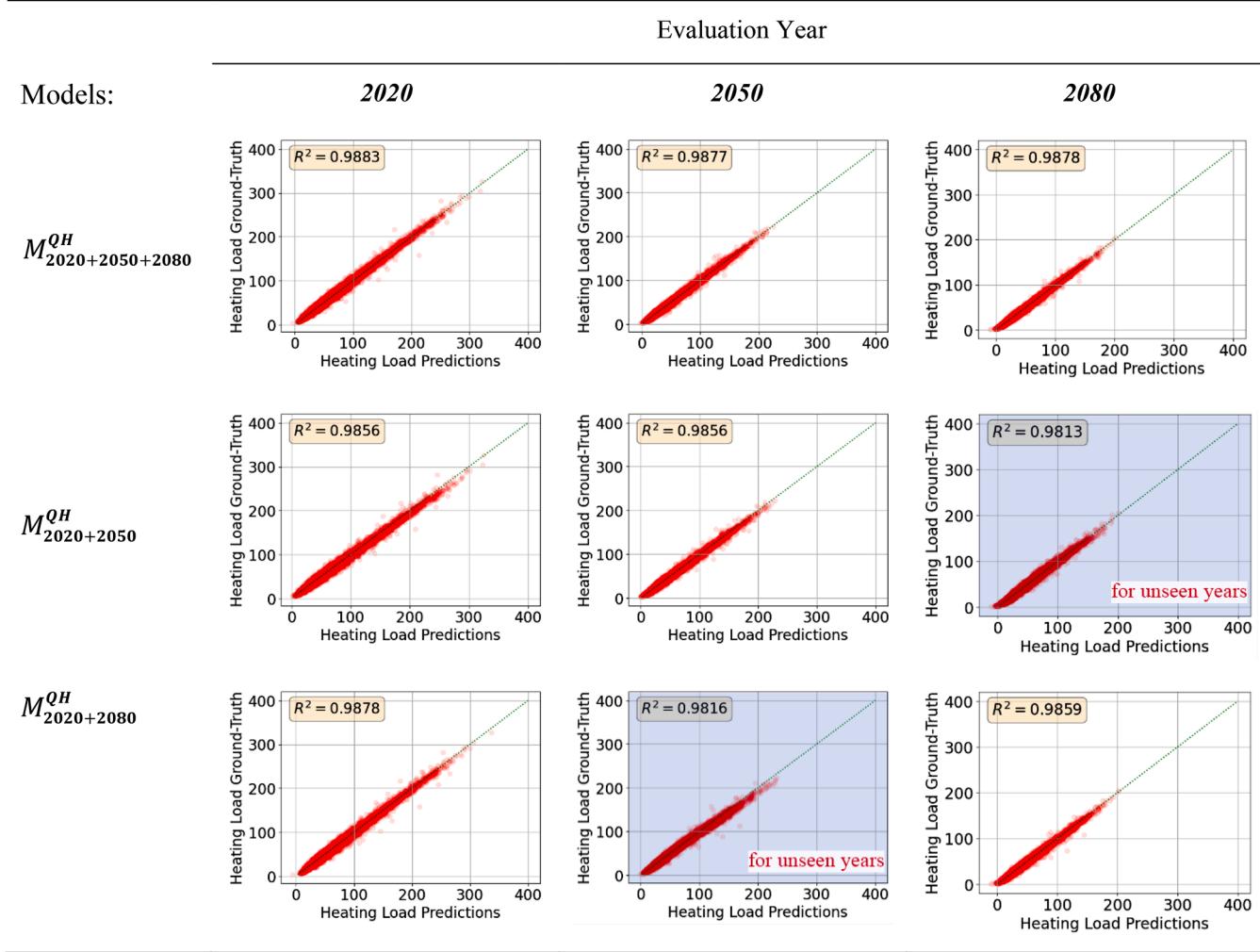
RO3: What are the long-term effects of different retrofit scenarios on $Q_{heating}$ and IOD (calculated using the $M_{2020+2050+2080}$ model)? (Section 4.3)

3.1. Experiment 1: Prediction reliability for the combined model

In this experiment, the first research question (RQ1) is investigated. To this end, two ML models ($M_{2020+2050+2080}^{IOD}$ and $M_{2020+2050+2080}^{QH}$) are trained with training data from simulations using W_{2020} , W_{2050} , and W_{2080} and the performances of these models for the same years are evaluated. In Table 5 and Table 6, the predictions of the models are compared against the correct values and R^2 values of the predictions. Table 7 shows the RMSE, MSE, and MAE of the trained models, with the scores averaged over 10 different seeds and their standard deviations computed. Table 4 shows the hyperparameter configuration of ML models with the highest validation score on their test sets. The results

Table 5

Q_{Heating} prediction performance of the ML models trained and tested on different years. The diagrams with gray shading show the unseen years' predictions.



suggest that the combined model $M_{2020+2050+2080}^{QH}$ can estimate very well with high R^2 values ($R^2 = \sim 0.99$) for all three prediction years. However, the quality of IOD predictions is slightly lower with $R^2 = \sim 0.96$ (p-value=0). The reason for this drop in performance can be explained by the fact that (i) adaptive comfort calculations take into account operative temperature, which is the mean of the indoor dry bulb temperature and mean radiant temperature; (ii) mean radiant temperature is the weighted average temperature of all zone surfaces that exchange radiant heat; (iii) the zones in the developed UBEM have the outdoor air, ground floor and/or the neighboring zone as boundary conditions; (iv) the boundary condition largely influences the surface temperature and therefore the indoor operative temperature and (v) the proposed MLP model's inputs cannot directly distinguish between these different boundary conditions. As a result, the MLP cannot make predictions as precisely as heating energy use.

Moreover, it is observed that the percentage errors between predicted and actual values (PE) are higher for *verticalBoundary* and *form-factor*. The Spearman correlation is calculated between *verticalBoundary* and the PE values of IOD and Q_{heating} as -0.57 and -0.20 respectively. Moreover, the Spearman correlation is calculated between *formFactor* and PE values of IOD and Q_{heating} as -0.30 and -0.37 respectively. These two features aim to describe the 3D geometry of a zone and the boundary conditions of its outside surfaces. The high correlations indicate that the boundary conditions (that are outside air and ground floor)

cannot be distinguished by the ML model, and the corresponding heat gain and heat losses are not completely taken into account. Moreover, it is observed certain edge cases where the prediction performances are observed to be less accurate. Specifically, it is observed that the model tends to perform worse for ground floor zones (*verticalPos=0*) when predicting IOD . This is mainly because, in the underlying energy simulations, the ground floor thermal boundary is modeled with a constant ground temperature of 18.0 °C throughout the year. Consequently, ground floors experience less temperature variability compared to upper floors, which may lead to differences in thermal behavior that are not fully captured by the ML model.

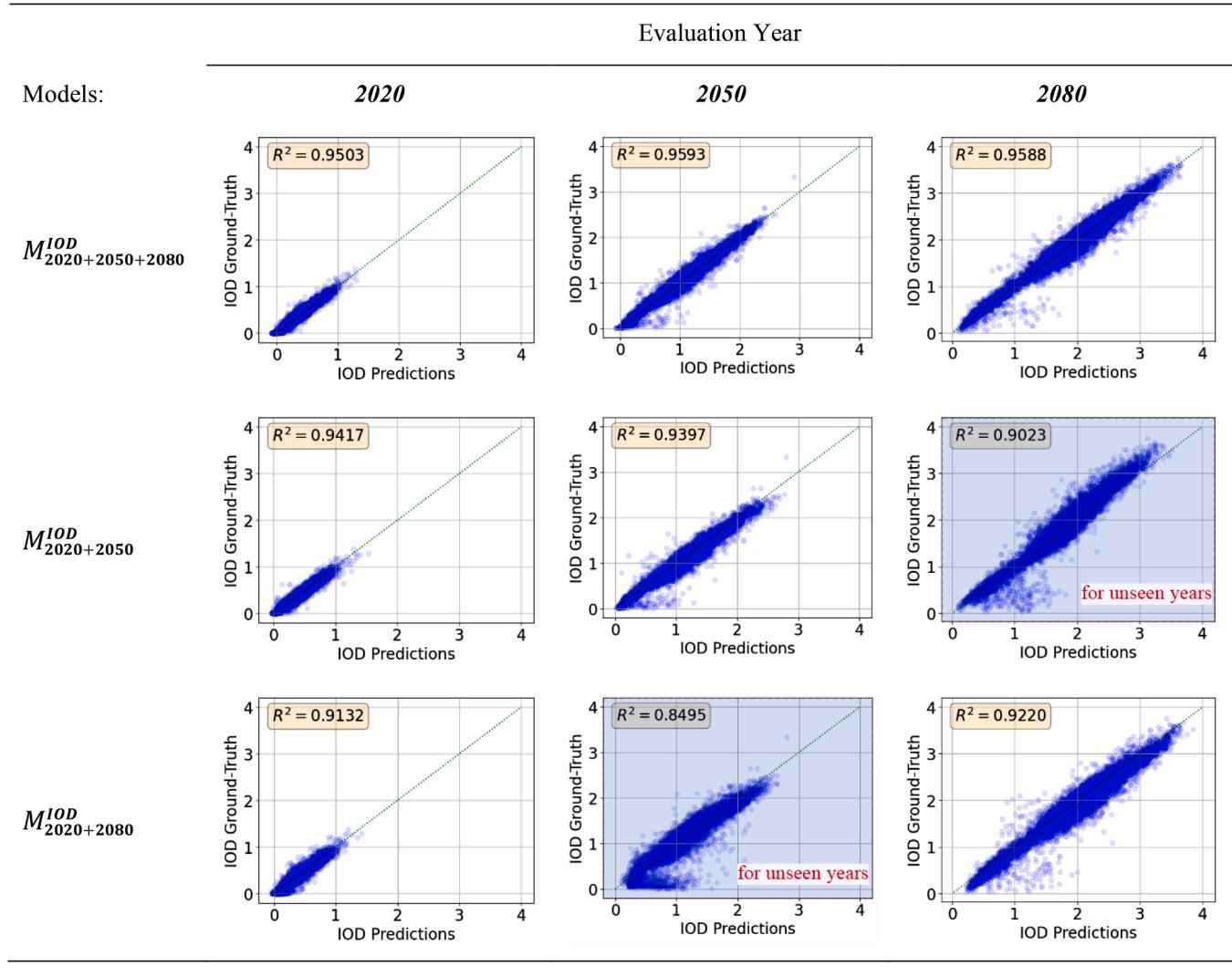
The MLP could be improved by integrating additional input features that capture the boundary conditions of each outside surface separately and even their orientation. However, extracting such new features from the 3D model would bring an extra burden, additional computational cost, and overcomplicate the MLP model. Nevertheless, this remains a research direction that can be explored in the future.

3.2. Experiment 2: Performance on unseen years

To investigate RQ2, the ML models are trained without providing data from certain years. For example, $M_{2020+2080}^{IOD}$ is trained and evaluated on IOD data for 2020, 2050, and 2080. In Tables 5 and 6, the qualitative and quantitative evaluations of the predictions are provided.

Table 6

IOD prediction performance of the ML models trained and tested on different years. The diagrams with gray shading show the unseen years' predictions.

**Table 7**

RMSE, MSE and MAE of the trained models.

ML model:	RMSE			MSE			MAE		
	2020	2050	2080	2020	2050	2080	2020	2050	2080
$M_{2020+2050+2080}^{QH}$	0.1291 ± 0.0023	0.0981 ± 0.0015	0.0850 ± 0.0013	0.0167 ± 0.0006	0.0096 ± 0.0003	0.0072 ± 0.0002	0.0985 ± 0.0016	0.0748 ± 0.0010	0.0646 ± 0.0009
$M_{2020+2050}^{QH}$	0.1345 ± 0.0043	0.1026 ± 0.0022	0.1462 ± 0.0410	0.0181 ± 0.0012	0.0105 ± 0.0005	0.0230 ± 0.0132	0.1026 ± 0.0035	0.0781 ± 0.0017	0.1175 ± 0.0375
$M_{2020+2080}^{QH}$	0.1366 ± 0.0058	0.1558 ± 0.0199	0.0913 ± 0.0028	0.0187 ± 0.0016	0.0247 ± 0.0063	0.0083 ± 0.0005	0.1044 ± 0.0050	0.1142 ± 0.0148	0.0691 ± 0.0024
$M_{2020+2050+2080}^{IOD}$	0.0641 ± 0.0027	0.1304 ± 0.0040	0.1889 ± 0.0091	0.0041 ± 0.0003	0.0170 ± 0.0011	0.0357 ± 0.0034	0.0474 ± 0.0028	0.0910 ± 0.0018	0.1317 ± 0.0046
$M_{2020+2050}^{IOD}$	0.0664 ± 0.0023	0.1381 ± 0.0053	0.4931 ± 0.1886	0.0044 ± 0.0003	0.0191 ± 0.0015	0.2787 ± 0.2270	0.0487 ± 0.0020	0.0981 ± 0.0038	0.4054 ± 0.1892
$M_{2020+2080}^{IOD}$	0.0762 ± 0.0046	0.5409 ± 0.3174	0.2061 ± 0.0138	0.0058 ± 0.0007	0.3933 ± 0.4772	0.0427 ± 0.0059	0.0562 ± 0.0046	0.3866 ± 0.2569	0.1511 ± 0.0108

The results show that $M_{2020+2050}^{QH}$ performs slightly worse (~0.006 lower R^2) on predicting $Q_{Heating}$ in 2080 (p -value < 0.01) and $M_{2020+2080}^{QH}$ on predicting 2050 (p -value < 0.001). The results in Table 6 suggest that ML models have more difficulty predicting IOD on unseen data. $M_{2020+2050}^{IOD}$ provides ~0.05 lower R^2 for 2080 (p -value < 0.01) and $M_{2020+2080}^{IOD}$ ~0.11 lower R^2 for 2050 (p -value < 0.03).

The higher performance loss in IOD prediction on unseen years (~0.05 and ~0.11 drop in R^2 score) compared to the loss in $Q_{Heating}$ prediction (~0.006 and ~0.006 drop in R^2 score) can be attributed to how the distributions of IOD and $Q_{Heating}$ change over the years. As shown in Fig. 8, the spread of $Q_{Heating}$ shrinks over time while the distributions still are overlapping. This suggests that a model that has seen

the 2020 $Q_{Heating}$ data will be able to generalize well to the data of 2050 and 2080. However, the distribution of IOD values shifts over the years without much overlap, which makes it difficult for an ML model to generalize since unseen years correspond to unseen IOD samples. This means that in IOD , training with partial data (2020+2050 or 2020+2080) results in a biased model that has not learned the patterns in the non-overlapping regions in our dataset.

3.3. Experiment 3: Energy retrofit packages analysis using ML-based prediction

The retrofitting of building envelopes is key to enhancing energy efficiency in the built environment. In this section, the trained ML models are used to make a long-term assessment of the following retrofit packages:

- A. improving the U-values of envelope materials (x_1 - x_4)
- B. reducing solar heat gain through window SHGC (x_{12})
- C. a combination of A and B (x_1 - x_4 and x_{12})

First, target values for each retrofit parameter are selected (Table 8). For U-values (x_1 - x_4), the target value is set to comply with the existing national standard that mandates thermal insulation levels for new and retrofitted buildings (TSE 825, 2013). For SHGC, the retrofit value is set to 0.3. To apply retrofit, only those buildings in the neighborhood that are below the target value are selected, so that well-performing buildings do not undergo any alterations. Fig. 11 shows the distributions of retrofit parameters before and after retrofit.

The climate features (x_{23} - x_{26}) for the years 2020, 2050 and 2080 and their regression equations are provided previously in Table 2. Using these equations, additionally the same climate features for the ten in-between years are calculated using the regression equations in Table 2. These features, together with the zone/building-related features, are used to predict $Q_{Heating}$ and IOD on the in-between years.

To prepare the input for the model, data from a CSV file containing all zones' input features is initially read. Next, one-hot encoding is used to convert categorical features into a numerical format. Then, the processed data is converted into a tensor. The tensor is a data structure used in PyTorch framework corresponds to a matrix representation of our input with the number of zones and features, and it encapsulates the essential information for our predictive model. Subsequently, the two ML models, $M_{2020+2050+2080}^{QH}$ and $M_{2020+2050+2080}^{IOD}$, are used to predict $Q_{Heating}$ and IOD for every 5 years from 2020 to 2080. These predictions are performed for the baseline model and the three retrofit packages.

When all three retrofit packages are considered, 238,993 zones need to be modified with retrofit actions. Moreover, as a long-term assessment is made for all years from 2020 to 2080, the number of predictions also multiplied by 13, amounting to 3106,909 in total. On the server (described in Section 2.1.3), energy simulation takes 3.82 s on average

Table 8
Existing and target mean values of different retrofit categories.

Retrofit Package	Parameter changed	Original mean value (see Fig. 9 for the distribution)	Target value	# of zones to be applied	Mean Value after retrofit
(A) Increase Thermal Insulation	Uwall (x_1)	0.944	0.5	40,583	0.430
	Uground (x_2)	1.283	0.45	14,937	0.435
	Uwindow (x_3)	2.249	1.8	42,212	1.637
	Uroof (x_4)	1.430	0.3	12,771	0.296
(B) Reduce Solar Heat Gain from windows	SHGC (x_{12})	0.547	0.3	64,004	0.3
(C) Combined	all above	same as above	same as above	64,486	same as above

per zone. On the same server, loading the ML model & the inputs, and then making a prediction take 0.00887 s for a single zone. Using ML provides a significant improvement (~430 faster) in computation time, especially considering that computational speed is a crucial factor in retrofit decision-making processes. The speed advantage of ML-based prediction makes it a compelling alternative to simulation-based performance analyses, especially considering the high number of alternative solutions to be explored by decision-makers.

The results of the predictions are as follows.

- The baseline predictions for $Q_{Heating}$ and IOD (Fig. 12-a) show that from 2020 to 2080, the increase in outdoor air temperature resulted in a marked decrease in $Q_{Heating}$, while concurrently increasing IOD . Mean $Q_{Heating}$ exhibited a notable decline, from 78.01 to 47.00 kWh/m², while the standard deviation declined from 48.88 to 31.14 over the same period. The average IOD exhibited an upward trend, increasing from 0.33 to 1.94 °C, with the standard deviation increasing from 0.22 to 0.71.
- Retrofit category A: The aim is the reduction of U-values of envelope constructions, thereby improving thermal insulation. Zones that already met the current building standards were not included in retrofitting. Fig. 12-b shows the $Q_{Heating}$ and IOD prediction. The improvement in U-values demonstrated a consistent decrease in the average energy expended for heating across each year of the calculation. In 2020, the improved U-values resulted in a reduction of average $Q_{Heating}$ from 78.01 to 51.31 kWh/m² as compared to the baseline. Similarly, in 2080, $Q_{Heating}$ decreased from 47.00 to 31.32 kWh/m². When the average IOD values are considered, a decline from 0.33 to 0.32 °C in 2020 and an increase from 1.94 to 2.00 °C in 2080 were observed.
- Retrofit category B: The SHGC values are reduced to mitigate solar heat gain (Fig. 12-c). Relative to the baseline scenario, a slight increase was observed in average $Q_{Heating}$, from 78.01 to 84.31 kWh/m² in 2020 and from 47.00 to 51.12 kWh/m² in 2080. In contrast, when the baseline and retrofit category B are compared, the average IOD values dropped from 0.33 to 0.21 °C in 2020 and from 1.94 to 1.64 °C in 2080. The reduced solar heat gain through the windows, therefore, had a significant impact on IOD mitigation, but slightly negatively influenced $Q_{heating}$.
- Retrofit category C: The two previous scenarios were combined (Fig. 12-d). Relative to the 2020 baseline scenario, a decrease was observed in $Q_{Heating}$ from 78.01 to 57.33 kWh/m². Relative to the 2080 baseline scenario, $Q_{Heating}$ was observed to decline from 47.00 to 35.36 kWh/m². IOD showed a decline from 0.33 to 0.18 °C in 2020 and from 1.94 to 1.67 °C in 2080. This reveals that when the retrofit scenario categories are applied separately, they demonstrate a favorable impact on either $Q_{Heating}$ or IOD . However, when the scenarios are applied together, it significantly reduces both $Q_{Heating}$ and IOD .

The results show a significant shift in $Q_{Heating}$ due to changes in U-values (category A and C), and another shift in IOD due to changes in SHGC (category B and C). Although category C performed worse than category A in $Q_{Heating}$, the improvement in IOD is significant to combat the negative effects of global warming. The results also point to the need for more drastic strategies for overheating mitigation, such as solar shading, chromogenic glazing, cool envelope materials, green roofs, or thermal mass (Attia et al., 2021). Advanced ventilation techniques can also be explored such as ventilated facades, cross-ventilation, ventilation fans, demand-controlled ventilation or heat recovery systems. The existing UBEMs can be adapted to incorporate these techniques in the future. It should also be noted that the existing ML models cannot predict beyond the input features they already have, which means that new input features should be added to the ML models to reflect these new cooling strategies. Future research is needed for the ML-based evaluation of such strategies.

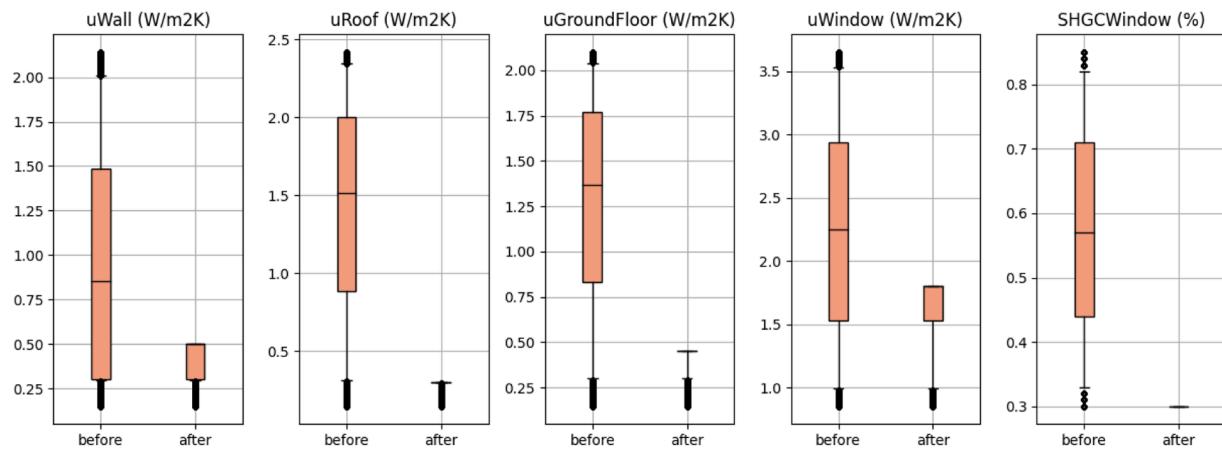


Fig. 11. Retrofit parameters and zone-level distributions before (left) and after (right) retrofit.

Finally, the initial cost of retrofitting each building (replacing all walls, roofs, grounds, and windows) to the thermal insulation levels in the retrofit package A is calculated. It is assumed that the targeted U-values in retrofit package A can be achieved by adding new insulation to the existing construction set. As these existing construction sets are unknown, it is also assumed that all buildings have the same base wall, roof, and ground construction layers, except for an existing insulation layer (if it exists) to be replaced by the new one, as shown in Table 9.

Insulation is only applied if the as-is U-value is higher than the target value. The required insulation thickness is calculated for each insulation layer using the following steps: (i) calculating the target and existing R-values by inserting their corresponding U-values, (ii) determining the additional required R-value (ΔR) by subtracting the existing R-value from the target R-value, and (iii) calculating the required insulation thickness by multiplying the ΔR value with the thermal conductivity (λ) of the insulation material. Insulation materials for wall, roof, and ground constructions are selected based on their thermal performance and suitability for the respective area of application. EPS is selected as the wall insulation material due to its relatively low thermal conductivity, cost-effectiveness, and ease of application. XPS is preferred as roof and ground insulation, due to its high compressive strength and durability. The existing windows are replaced by a glazing system with $U = 1.8 \text{ W/m}^2\text{K}$. After the insulation thickness is calculated, the unit costs of the new insulation layers are acquired from the Unit Price Document 2025 published by the Ministry of Environment, Urbanization, and Climate Change (2025 Unit Price Document, 2025). Finally, the total initial investment cost is calculated by multiplying this unit cost by the total surface area of the corresponding component.

Fig. 13 shows the calculated initial investment costs for retrofitting the 592 buildings. Each bar represents the total retrofit cost per building, broken down by component type. The results show that wall and window retrofits constitute the majority of the total cost, with mean costs reaching €9241 and €5978, respectively. The initial cost of roof and ground retrofit is generally lower, at €2490 and €2521, respectively. This can be attributed to the relatively smaller surface areas of these surfaces compared to walls. While the results only consider the initial cost of retrofit package A, other costs such as installation, transportation, and other indirect costs during cost estimations should also be considered in the future.

4. Discussion and conclusion

Analyzing building performance at the urban scale while considering the impact of climate change is a necessity for effective energy efficiency strategies. ML methods have the capability to predict building performance with lower computational cost, allowing the decision-makers to quickly calculate the effects of different energy-saving scenarios.

However, existing ML methods do not consider the long-term effects, as the models are usually trained without considering future climate scenarios. In addition, recent approaches focused on discrete years in future calculations. In this study, a UBEM-supported, ML-based approach for long-term building performance prediction is presented.

In this study, combined models, $M_{2020+2050+2080}^{QH}$ and $M_{2020+2050+2080}^{IOD}$, have been developed to predict $Q_{heating}$ and IOD by using simulation results from the weather files of the years 2020, 2050, and 2080 as training data. The correct values have been compared to the predictions of the trained models. With $R^2 = \sim 0.98$, the prediction quality for $Q_{heating}$ is very high for all prediction years. Still with $R^2 = \sim 0.96$, the IOD predictions have a lower performance quality. Furthermore, ML models were trained by excluding data from specific years. The evaluations indicate a slight performance decrease in predicting both $Q_{heating}$ and IOD for unseen years. $M_{2020+2050}^{QH}$ performs slightly worse on predicting $Q_{heating}$ in 2080 and $M_{2020+2080}^{QH}$ on predicting $Q_{heating}$ in 2050. In addition, with ~ 0.08 lower R^2 of $M_{2020+2050}^{IOD}$ for 2080 and ~ 0.10 lower R^2 of $M_{2020+2080}^{IOD}$ for 2050, the models have more difficulty in predicting IOD on unseen data. This difference in the performance of prediction $Q_{heating}$ and IOD is attributed to the changing distributions of IOD and $Q_{heating}$ over time. The $Q_{heating}$ distribution shrinks within its own data range, aiding model generalization, while the shifting IOD values affect the model's ability to adapt to unseen samples.

When the time costs of simulation-based and ML-based calculation of building performance are compared and it is observed that the latter is ~ 430 times faster than the former. This advantage of ML-based methods can help facilitate the exploration of a high number of retrofit alternatives in a very quick manner. The reduced time cost is also critical for the development of decision-support tools and their capacity in providing real-time feedback to the users.

After model training, $M_{2020+2050+2080}^{QH}$ and $M_{2020+2050+2080}^{IOD}$ are used to assess the selected neighborhood's $Q_{heating}$ and IOD from 2020 to 2080 in 5-year intervals. These predictions are performed for the baseline scenario representing the as-is condition and three retrofit packages. Compared to the baseline scenario, retrofit package A shows a consistent decrease in average $Q_{heating}$ for each year, from 78.01 to 51.31 kWh/m² in 2020, and from 47 to 31.32 kWh/m² in 2080 but has little effect on average IOD (from 0.33 to 0.32 °C in 2020, and from 1.94 to 2.00 °C in 2080). Retrofit package B slightly increases average $Q_{heating}$ from 78.01 to 84.31 kWh/m² in 2020 and from 47 to 51.12 kWh/m² in 2080 but can reduce average IOD from 0.33 to 0.21 °C in 2020 and from 1.94 to 1.64 °C in 2080. This shows that changing SHGC slightly reduces but does not completely eliminate overheating risks. Therefore, other adaptive measures will be necessary to protect buildings from extreme temperatures. Finally, retrofit package C significantly decreases $Q_{heating}$ from 78.01 to 57.33 kWh/m² in 2020 and from 47 to 35.36 kWh/m² in 2080.

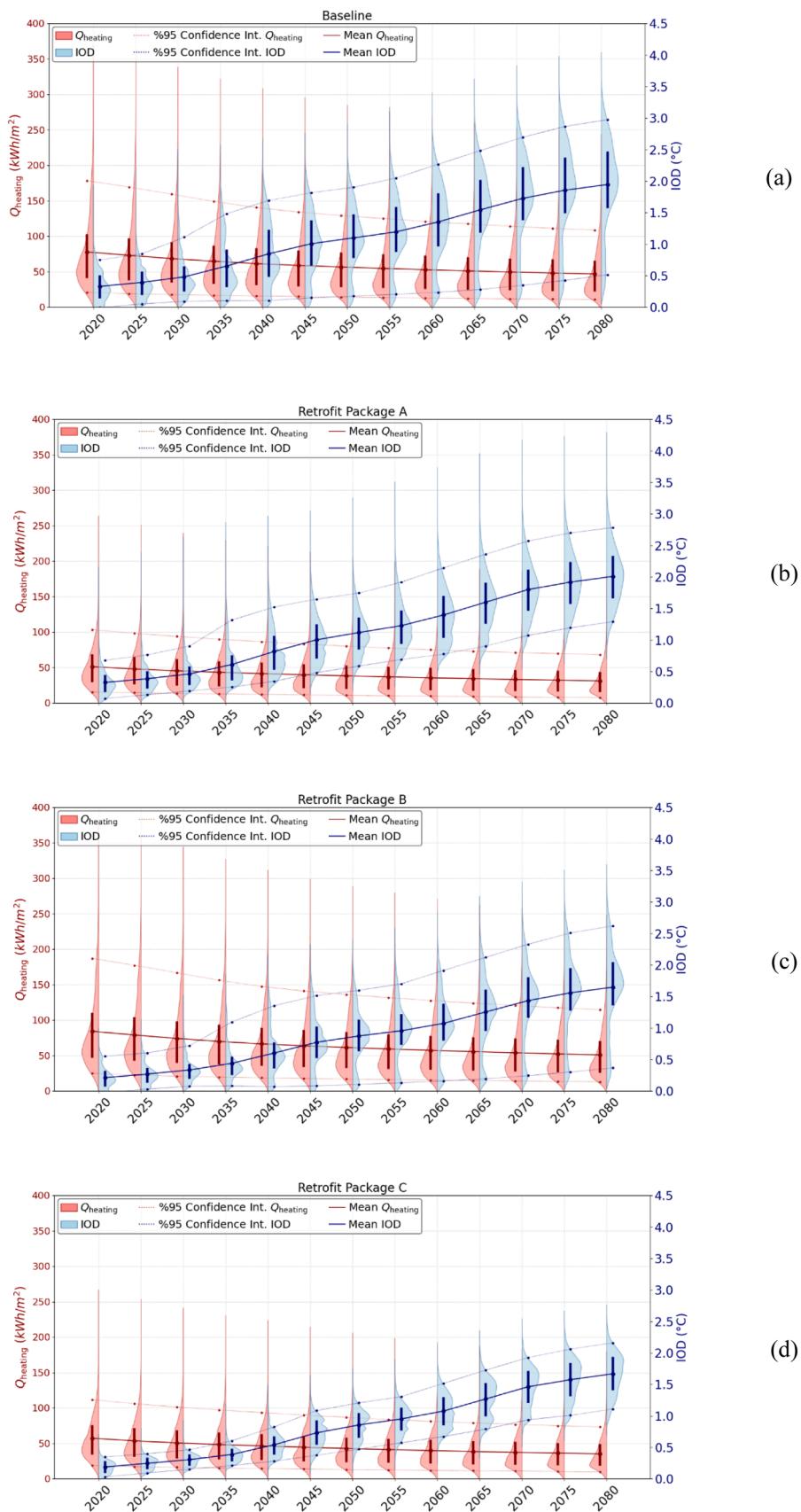


Fig. 12. ML-based Q_{heating} and IOD predictions for (a) the baseline case (b) retrofit package A (c) retrofit package B (d) retrofit package C (the solid rectangles indicate the interquartile ranges).

Table 9

Base construction layers and new insulation properties.

	Base construction		New insulation		
	Layers (from inside out)	Insulation type	Thermal cond. (W/mK)	Thickness (m)	Unit cost (€/m ²)
Wall	0.02 m cement plaster, $R = 0.02^*$ 0.085 m perforated brick, $R = 0.189$ Insulation layer of unknown thickness (if exists) 0.01 m gypsum plaster, $R = 0.029$	EPS	0.035	0.06	18.73
Roof	0.025 m plaster, $R = 0.03$ 0.20 m concrete slab, $R = 0.08$ 0.04 m leveling concrete, $R = 0.13$ 0.003 m waterproof membrane, $R = 0.015$ Insulation layer of unknown thickness (if exists) Separation layer Gravel, $R = 0.36$	XPS	0.035	0.08	7.56
Ground	0.05 m laminate floor, $R = 0.17$ 0.003 m polyester felt, $R = 0.02$ Insulation layer of unknown thickness (if exists) 0.15 m groundwork, $R = 0.06$ 0.1 m protective concrete, $R = 0.06$ 0.003 m waterproof membrane, $R = 0.015$ 0.1 m blinding concrete, $R = 0.06$ Gravel	XPS	0.035	0.08	7.56
Window	Completely discard the	A new window product with $U = 1.8$		31.37	

Table 9 (continued)

	Base construction	New insulation			
	Layers (from inside out)	Insulation type	Thermal cond. (W/mK)	Thickness (m)	Unit cost (€/m ²)
	existing window *Units of R values are in m ² K/W				

IOD decreases from 0.33 to 0.18 °C in 2020 and from 1.94 to 1.66 °C in 2080. The results suggest that further research is needed to explore other strategies for climate change mitigation, especially for overheating.

4.1. Limitations and future work

Predictive models for building energy use estimation have significant potential for fast and responsive decision support tools, especially where real-time feedback is needed—such as in early design stages or retrofit planning processes. Their speed makes them suitable for interactive applications where traditional simulation-based approaches would be too slow. However, this agility often comes with trade-offs. First, prediction accuracy can be limited, especially if the model is trained with narrow datasets, leading to overfitting. We addressed this problem by generating new layers of input data and running simulations. As a result, our model showed good generalization across different scenarios. Secondly, reliability can be an issue when predictive models are used beyond their training dataset, e.g. for different building types with different input features (materials, 3D form, occupancy patterns, urban contexts, climate features etc.). A model trained on a very diverse dataset can help to solve this issue. However, compiling such comprehensive datasets for a wide range of building types can become impractical. Different building types (such as single-family homes, offices, or hospitals) behave very differently in terms of energy consumption, as they differ in terms of size, occupancy patterns, heating/cooling systems, and internal gains. Attempting to incorporate all these features of different building types into a single model can result in unmanageable complexity and lower accuracy, as the model may struggle to learn patterns across the entire dataset. In many cases, a more effective strategy is to focus on a thinner, well-defined slice of the building stock, for example, a specific residential typology for a specific climate zone, and develop a model tailored to that context. This approach can bring a clearer feature relevance and more trustworthy predictions within its own scope. Third, interpretability is also a key challenge, especially for black-box models like MLPs that are used in this study. MLPs can make accurate predictions but offer little insight into the reasons why certain outcomes were achieved. For decision-makers, this lack of transparency can reduce confidence in the calculated results. In the future, interpretable model diagnostics methods can be used (such as Shapley Additive Explanations (Lundberg & Lee, 2017), Local Interpretable Model-Agnostic Explanations (Ribeiro et al., 2016), Integrated Gradients (Sundararajan et al., 2017)).

This study assumes that the building characteristics remain unchanged over 60 years and doesn't take into account the turnover dynamics in buildings, such as buildings being demolished due to old age, changes in occupancy or their behaviors, the adoption of new building technologies, or declining efficiencies of existing building systems. Another future intervention is air conditioning, which is expected to be needed under climate change impact and the resulting indoor overheating. However, in reality, phased interventions are implemented that gradually and incrementally improve a building to minimize high upfront costs or disruption to occupants. This limitation in our study can be overcome by planning various retrofit actions over time by

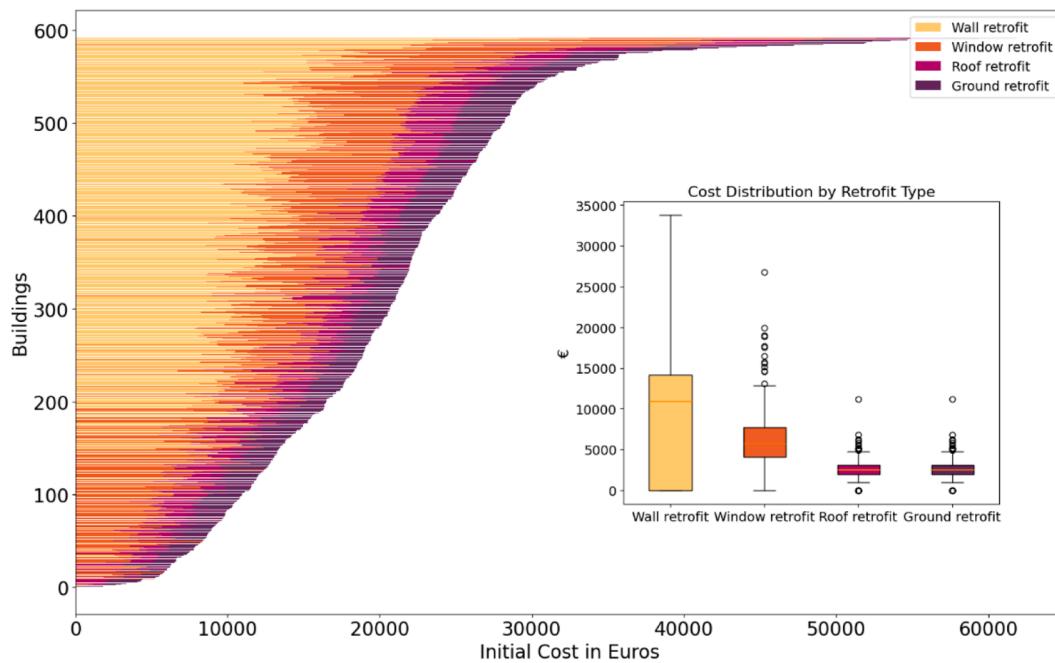


Fig. 13. The initial retrofit cost by building and by component type.

prioritizing them in terms of effectiveness, financial feasibility, or system availability. This can be explored in future work.

In this study, a single climate scenario (SSP5–8.5) was utilized, which represents the worst-case, business-as-usual development. The use of a single climate projection scenario is crucial to ensure that the performance predictions remain consistent throughout the future years during retrofit decision-making processes. However, a variety of climate scenarios can also be used in simulations, and their climate-related input parameters can be integrated into the ML models. As such, a variety of different future scenarios can also be used to evaluate more optimistic futures as a result of stricter standards and the widespread implementation of energy efficiency technologies.

The SSP5–8.5 scenario is selected to represent a worst-case climate situation, making it useful for testing how buildings and energy systems could cope with severe future conditions. This scenario choice plays a critical role in shaping long-term projections of building energy demand. If milder climate scenarios—such as SSP1–2.6 or SSP2–4.5—were used instead, the projected impacts on building performance would likely be different. These scenarios assume a smaller temperature increase and less severe climate extremes, which could lead to a smaller reduction in heating energy consumption and a more modest increase in indoor overheating. This in turn could indicate a reduced need for aggressive retrofit measures compared to what would be required under SSP5–8.5. Although the actual path that global development will take remains uncertain, it is valuable to consider a range of scenarios. From a building retrofit perspective, exploring high and low emissions scenarios can support more resilient and adaptive planning. Accordingly, it would be beneficial for future research to use ensemble simulations that incorporate multiple SSPs to better understand how energy demand might change in different climate future scenarios.

In this study, an existing tool, Future Weather Generator, is used which generates future weather files by modifying a historical baseline weather file using climate change projections from GCMs. These projections are usually expressed as deltas, i.e. changes in climate variables such as temperature and humidity between a historical reference period and a future period. The morphing method applies these deltas linearly to the baseline data, i.e. it adds or multiplies the projected changes evenly over all time steps. However, morphing urban weather files can be problematic because GCMs typically represent large-scale, rural-like

conditions that are not capable of capturing the urban heat island (UHI) effect. UHI is a localized, non-linear phenomenon, and varies by time of day, season and urban form. It also often amplifies night-time temperatures in a way that is not reflected by simple delta values. Consequently, directly applying GCM deltas belonging to rural areas to urban baseline data may misrepresent how UHI will evolve with climate change and lead to inaccurate estimations in urban areas. In particular, night-time temperatures may be underestimated, resulting in lower IOD and higher heating energy demand than in reality. Consequently, these inaccuracies in the weather file can lead to misleading conclusions about the effectiveness of retrofit strategies. A solution to this problem could be to post-process the rural deltas to better reflect urban environments by modifying daytime and nighttime temperature increases separately. This can be addressed as future work.

Moreover, decision-makers can select the climate scenario they find most probable and calculate the heating energy consumption and IOD values that correspond with the selected scenario. Future work can focus on coupled analyses of future building performance linked to different development pathways and energy policies, as linked to the different climatic projections.

Our work demonstrates how the ML model and predictions can be calculated for a simple case by developing retrofit scenarios and projecting the corresponding heating consumption and IOD until 2080. Future work can also consider other retrofit measures and dynamic adaptation strategies. ML models can further be developed to evaluate such different scenarios, which also would require that both the datasets and ML models are extended with new features describing these new systems and their corresponding energy use values. Moreover, the cost-effectiveness of different measures is a critical factor during retrofit decision-making, which is not within the scope of this paper.

While this study quantifies annual thermal discomfort using IOD, it does not specify whether air conditioning (AC) is required to prevent excessive overheating. Future work can focus on the estimation of the exceedance of a comfort threshold, which can determine whether a space needs air conditioning currently or in the future. As a result, the energy models can be modified accordingly with AC installations. The standard CIBSE TM59 (Bateson et al., 2017), which describes a design methodology for the assessment of overheating risk in residential buildings, can help in the determination of a threshold after which a

zone is considered thermally uncomfortable. TM59 proposes the “hours of exceedance” criteria, which state a space has a high risk of overheating when the number of hours that the temperature difference between the upper-temperature limit and the operative temperature is greater than one degree exceeds 3 % of occupied hours. These criteria can be used to determine the exact future year that AC would be needed in a zone.

Additionally, this study quantifies overheating with IOD, which might remain insufficient under different environmental and demographic conditions. Future work can focus on other metrics, such as the Heat Vulnerability Index, which can consider a variety of vulnerability factors such as social/language vulnerability, socioeconomic vulnerability, environmental/urban vulnerability, and elderly/ social isolation (Nayak et al., 2018). HVI can also provide a more nuanced analysis of the sensitivity and adaptive capacity of different occupants based on their age, sex, health condition, social isolation, socioeconomic condition, and level of education (Szagri et al., 2023).

The developed UBEM considers each apartment unit as one thermal zone, without explicitly representing day-occupied and night-occupied spaces. In general, it is accepted that in residential buildings, day-occupied and night-occupied spaces have different comfort models (i.e. bedroom occupants have lower comfort thresholds as compared, and they tend to impose higher penalties for overheating) (Hamdy & Hensen, 2015). The ASHRAE 55–2013 Adaptive Comfort Model used in this study, on the other hand, is developed for all naturally ventilated buildings, regardless of their space type. As a result, it does not capture the differences in comfort requirements between day-occupied and night-occupied zones and may underestimate overheating for the latter. Future work is needed to provide higher-resolution representations of residential building thermal comfort at an urban scale. However, it should also be noted that higher spatial resolution also has disadvantages related to modelling efforts and computational costs of simulations. Another limitation is that the ASHRAE 55–2013 Adaptive Comfort Model, which is not developed particularly for the selected Ankara neighbourhood, might not accurately capture the occupants' individual and cultural differences in temperature preferences and adaptive behaviours. This could result in discrepancies between expected and actual comfort levels for the studied context, as well as an incorrect representation of occupants' thermal comfort needs in this particular context.

While two distinct ML models with one output are developed in this study, it is indeed possible to use a single ML model (e.g., an MLP) to predict two outputs. However, it turns out in the case of $Q_{Heating}$ and IOD that these two output variables are not correlated, and learning them simultaneously (called multi-task learning in the literature) leads to lower performance than using two ML models to learn these two output variables. This suggests that the learning difficulties of the two variables may be different, and the shared information (representation) between the two variables may be low. This calls for more research, which can be the sole focus of another study.

Although accurate predictions (with $R^2 > 0.9$) are obtained with the presented MLP models, they can be improved further to consider inter-building or inter-zone relations. Inter-building relations can be categorized as physics-based (i.e. form shading, microclimate) or occupant-based (Lin et al., 2025) and recent research has proven that the integration of such spatial relationships into data-driven models can improve predictions by providing a more realistic representation of the physical conditions. Alternatively, inter-zonal relationships can also be integrated into ML models, especially when the indoor conditions of two neighboring zones are sufficiently different and the heat losses through internal surfaces are high. A solution, which was previously proposed by the authors of this study, is the use of Graph Neural Networks (GNN), as they can explicitly represent the neighborhood relations (specifically related to the heat transfer) between zones, i.e. by modeling a building as a homogeneous bidirectional graph, where a zone is represented by a node and the relationship between two neighboring zones are represented by a bidirectional edge (Halacli et al., 2023). While such models

with more inductive bias have proven to improve accuracy, the high cost of developing a separate graph representing a building hinders its practical implementation. Additionally, in the future, GNNs can also be used to capture the inter-building relationships to represent context shading.

A significant challenge in the presented approach is the collection of extensive datasets, as well as training a new ML model for every distinct urban setting and climate. Large datasets are needed for high prediction capacity, as they have more diverse samples and reduce overfitting. On the other hand, difficulties in obtaining building data and cumbersome UBEM development processes lead to insufficient dataset sizes and overfitting. A potential direction for future development is transfer learning (TL), which allows the application of existing knowledge from one domain to another, especially in contexts with scarce data (Weiss et al., 2016). Particularly in urban-scale predictions, a well-trained model belonging to a neighborhood can be reused by implementing TL in other neighborhoods/cities /climates. This can result in the effective use of limited data resources and improve the accuracy and applicability of predictive models in real-world scenarios. In another work by the authors, TL was applied to the Ankara MLP model developed in this study to help train two other neighborhoods in cities with different climates in Turkiye (Canli et al., 2024). The results show that TL can successfully utilize what has been learned in a source city for a target city and provide better performance than an ML model trained from scratch. This shows significant potential for future research in establishing a network of cities that support ML-based learning from one another using connected data spaces.

Acknowledgements

This study is funded by Scientific and Technological Research Council of Turkiye (TUBITAK), Grant no: 120M997 and Horizon Europe under grant number 101096405. We also gratefully acknowledge the computational resources kindly provided by METU Center for Robotics and Artificial Intelligence (METU-ROMER) and TRUBA (Turkish National High Performance Computing Center).

CRediT authorship contribution statement

Ilkim Canli Akyol: Writing – original draft, Visualization, Validation, Software, Formal analysis, Data curation, Writing – review & editing, Investigation, Methodology. **Eren Gökberk Halacli:** . **Sevval Ucar:** Writing – original draft, Visualization, Validation, Software, Methodology. **Orcun Koral Iseri:** Writing – original draft, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Feyza Yavuz:** Formal analysis, Data curation. **Dilara Guney:** Writing – original draft, Visualization. **Fatma Ece Gursoy:** Validation. **Ayca Duran:** Data curation. **Cagla Meral Akgul:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Sinan Kalkan:** Writing – review & editing, Writing – original draft, Validation, Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ipek Gursel Dino:** Writing – review & editing, Writing – original draft, Validation, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The dataset, code and trained models presented in this paper are released by the authors in a public GitHub repository to help reproduce

the modeling and validation results (<https://github.com/metu-energ/y/energy-journal>).

Data availability

We share our code in a github repository. Our data can be made available upon request.

References

- Aafaifia, M., Djiar, K. A., Bich-Ngoc, N., & Teller, J. (2021). An energy consumption model for the Algerian residential building's stock, based on a triangular approach: Geographic Information System (GIS), regression analysis and hierarchical cluster analysis. *Sustainable Cities and Society*, 74, Article 10391. <https://doi.org/10.1016/j.scs.2021.103191>. Article.
- Ahmed, W., & Asif, M. (2020). BIM-based techno-economic assessment of energy retrofitting residential buildings in hot humid climate. *Energy and Buildings*, 227, Article 110406. <https://doi.org/10.1016/j.enbuild.2020.110406>. Article.
- Akgüç, A., & Yilmaz, A. Z. (2022). Determining HVAC system retrofit measures to improve cost-optimum energy efficiency level of high-rise residential buildings. *Journal of Building Engineering*, 54, Article 104631. <https://doi.org/10.1016/j.jobe.2022.104631>. Article.
- Ala, G., Orioli, A., & Di Gangi, A. (2019). Energy and economic analysis of air-to-air heat pumps as an alternative to domestic gas boiler heating systems in the South of Italy. *Energy*, 173, 59–74. <https://doi.org/10.1016/j.energy.2019.02.011>
- Ali, U., Bano, S., Shamsi, M. H., Sood, D., Hoare, C., Zuo, W., Hewitt, N., & O'Donnell, J. (2024). Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. *Energy and Buildings*, 303, Article 113768. <https://doi.org/10.1016/j.enbuild.2023.113768>. Article.
- 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, Article 111073. <https://doi.org/10.1016/j.enbuild.2021.111073>. Article.
- Allobaidi, M. H., Chebana, F., & Meguid, M. A. (2018). Robust ensemble learning framework for day-ahead forecasting of household based energy consumption. *Applied Energy*, 212, 997–1012. <https://doi.org/10.1016/j.apenergy.2017.12.054>
- Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192–1205. <https://doi.org/10.1016/j.rser.2017.04.095>
- ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers). (2013). *ANSI/ASHRAE Standard 55-2023 - Thermal Environmental Conditions for Human Occupancy*.
- ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers). (2015). *ANSI/ASHRAE Standard 62.1-2019 - Ventilation for Acceptable Indoor Air Quality*.
- ASHRAE (American Society of Heating, Refrigerating and Air-Conditioning Engineers). (2023). *ANSI/ASHRAE/IES Standard 90.1 - Energy Standard for Sites and Buildings Except Low-Rise Residential Buildings (I-P Edition)*.
- 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, Article 103534. <https://doi.org/10.1016/j.scs.2021.103534>. Article.
- Attia, S., Levinson, R., Ndongo, E., Holzer, P., Berk Kazancı, O., Homaei, S., Zhang, C., Olesen, B. W., Qi, D., Hamdy, M., & Heiselberg, P. (2021). Resilient cooling of buildings to protect against heat waves and power outages: Key concepts and definition. *Energy and Buildings*, 239, Article 110869. <https://doi.org/10.1016/j.enbuild.2021.110869>. Article.
- Baba, F. M., Ge, H., Wang, L.(Leon), & Zmeureanu, R. (2022). Do high energy-efficient buildings increase overheating risk in cold climates? Causes and mitigation measures required under recent and future climates. *Building and Environment*, 219, Article 109230. <https://doi.org/10.1016/j.buildenv.2022.109230>. Article.
- Barbaresi, A., Ceccarelli, M., Menichetti, G., Torreggiani, D., Tassinari, P., & Bovo, M. (2022). Application of Machine Learning Models for Fast and Accurate Predictions of Building Energy Need. *Energies*, 15(4), Article 1266. <https://doi.org/10.3390/en15041266>. Article.
- Bateson, A., Bonfigli, C., Chorafa, M., Diamond, S., Eliades, C., Mylona, A., Taylor, B., & Virk, D. (2017). *TM59 Design methodology for the assessment of overheating risk in homes*. CIBSE.
- Bektaş, Y., & Sakarya, A. (2023). The Relationship between the Built Environment and Climate Change: The Case of Turkish Provinces. *Sustainability*, 15(2), Article 1659. <https://doi.org/10.3390/su15021659>. Article.
- Berardi, U., & Jafarpur, P. (2020). Assessing the impact of climate change on building heating and cooling energy demand in Canada. *Renewable and Sustainable Energy Reviews*, 121, Article 109681. <https://doi.org/10.1016/j.rser.2019.109681>. Article.
- Boait, P. J., Dixon, D., Fan, D., & Stafford, A. (2012). Production efficiency of hot water for domestic use. *Energy and Buildings*, 54, 160–168. <https://doi.org/10.1016/j.enbuild.2012.07.011>
- Brohus, H., Heiselberg, P., Brohus, H., Heiselberg, P., Simonsen, A., & Sørensen, K. C. (2009). Uncertainty of energy consumption assessment of domestic buildings. In *Eleventh International IBPSA Conference*.
- Cáceres, L., Merino, J. I., & Díaz-Díaz, N. (2021). A computational intelligence approach to predict energy demand using random forest in a cloudera cluster. *Applied Sciences (Switzerland)*, 11(18), Article 8635. <https://doi.org/10.3390/app11188635>. Article.
- Canlı, İ., Uçar, Ş., Gürsoy, F. E., Halaçlı, E. G., İşeri, O. K., Yavuz, F., Gürsel Dino, İ., & Kalkan, S. (2024). Transfer Learning-Based Prediction of Heating Energy Consumption Using Urban Building Energy Models (UBEM). In *31st International Workshop on Intelligent Computing in Engineering*.
- Chen, Y., Guo, M., Chen, Z., Chen, Z., & Ji, Y. (2022). Physical energy and data-driven models in building energy prediction: A review. *Energy Reports*, 8, 2656–2671. <https://doi.org/10.1016/j.eegr.2022.01.162>
- Chen, Y. T., Piedad, E., & Kuo, C. C. (2019). Energy consumption load forecasting using a level-based random forest classifier. *Symmetry*, 11(8), Article 956. <https://doi.org/10.3390/sym11080956>. Article.
- D'Agostino, D., Parker, D., Epifani, I., Crawley, D., & Lawrie, L. (2022). How will future climate impact the design and performance of nearly zero energy buildings (NZEBs)? *Energy*, 240, Article 122479. <https://doi.org/10.1016/j.energy.2021.122479>. Article.
- Dino, I. G., & Meral Akgül, C. (2019). Impact of climate change on the existing residential building stock in Turkey: An analysis on energy use, greenhouse gas emissions and occupant comfort. *Renewable Energy*, 141, 828–846. <https://doi.org/10.1016/j.renene.2019.03.150>
- Dogan, T., & Reinhart, C. (2017). Shoeboxer: An algorithm for abstracted rapid multi-zone urban building energy model generation and simulation. *Energy and Buildings*, 140, 140–153. <https://doi.org/10.1016/j.enbuild.2017.01.030>
- European Environment Agency. (2021, June). *EEA Report 16/2020, Urban sustainability in Europe : what is driving cities' environmental change?* <https://www.eea.europa.eu/en-analysis/publications/urban-sustainability-in-europe-what>
- Fayaz, M., & Kim, D. (2018). A prediction methodology of energy consumption based on deep extreme learning machine and comparative analysis in residential buildings. *Electronics*, 7(10), Article 222. <https://doi.org/10.3390/electronics7100222>. Article.
- Ferrando, M., Causone, F., Hong, T., & Chen, Y. (2020). Urban building energy modeling (UBEM) tools: A state-of-the-art review of bottom-up physics-based approaches. *Sustainable Cities and Society*, 62, Article 102408. <https://doi.org/10.1016/j.scs.2020.102408>. Article.
- Flores-Larsen, S., Filippini, C., & Barela, G. (2019). Impact of climate change on energy use and bioclimatic design of residential buildings in the 21st century in Argentina. *Energy and Buildings*, 184, 216–229. <https://doi.org/10.1016/j.enbuild.2018.12.015>
- Gercek, M., & Durmuş Arsan, Z. (2019). Energy and environmental performance based decision support process for early design stages of residential buildings under climate change. *Sustainable Cities and Society*, 48, Article 101580. <https://doi.org/10.1016/j.scs.2019.101580>. Article.
- Guo, Y., Wang, J., Chen, H., Li, G., Liu, J., Xu, C., Huang, R., & Huang, Y. (2018). Machine learning-based thermal response time ahead energy demand prediction for building heating systems. *Applied Energy*, 221, 16–27. <https://doi.org/10.1016/j.apenergy.2018.03.125>
- Halaçlı, E. G., Canlı, İ., İşeri, O. K., Yavuz, F., Akgül, C. M., Kalkan, S., & Dino, I. G. (2023). A Novel Graph Neural Network for Zone-Level Urban-Scale Building Energy Use Estimation. In *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation* (pp. 169–176). <https://doi.org/10.1145/3600100.3623747>
- Hamdy, M., Carlucci, S., Hoes, P. J., & Hensen, J. L. M. (2017). The impact of climate change on the overheating risk in dwellings—A Dutch case study. *Building and Environment*, 122, 307–323. <https://doi.org/10.1016/j.buildenv.2017.06.031>
- Hamdy, M., & Hensen, J. L. (2015). ASSESSMENT OF OVERHEATING RISK IN DWELLINGS. In *Healthy Building Europe 2015 Conference*.
- Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2020). Ten questions on urban building energy modeling. *Building and Environment*, 168, Article 106508. <https://doi.org/10.1016/j.buildenv.2019.106508>. Article.
- Horwitz, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- IEA. (2017). *World Energy Balances 2017*. Paris: IEA, OECD Publishing. https://doi.org/10.1787/world_energy_bal-2017-en
- IEA. (2021). *Empowering Cities for a Net Zero Future: Unlocking resilient, smart, sustainable urban energy systems*. Paris: OECD Publishing. <https://doi.org/10.1787/7a222c8b-en>
- IEA. (2023). *Tracking Clean Energy Progress 2023*. Paris: IEA. <https://www.iea.org/reports/tracking-clean-energy-progress-2023>
- Imam, M., Chen, Y., & Marshall, T. (2023). Integrating Climate Change Projections in Building Performance Simulations: Case Study of 17 North American Cities. *Journal of Architectural Engineering*, 29(3), Article 05023004. <https://doi.org/10.1061/jaeied.aeeng-1593>. Article.
- Jafarpur, P., & Berardi, U. (2021). Effects of climate changes on building energy demand and thermal comfort in Canadian office buildings adopting different temperature setpoints. *Journal of Building Engineering*, 42, Article 102725. <https://doi.org/10.1016/j.jobe.2021.102725>. Article.
- Jentsch, M. F., James, P. A. B., Bourikas, L., & Bahaj, A. B. S. (2013). Transforming existing weather data for worldwide locations to enable energy and building performance simulation under future climates. *Renewable Energy*, 55, 514–524. <https://doi.org/10.1016/j.renene.2012.12.049>
- Jiang, F., Ma, J., Li, Z., & Ding, Y. (2022). Prediction of energy use intensity of urban buildings using the semi-supervised deep learning model. *Energy*, 249, Article 123631. <https://doi.org/10.1016/j.energy.2022.123631>. Article.
- Kamalov, F. (2020). Kernel density estimation based sampling for imbalanced class distribution. *Information Sciences*, 512, 1192–1201. <https://doi.org/10.1016/j.ins.2019.10.017>
- Kamel, E., Sheikh, S., & Huang, X. (2020). Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days. *Energy*, 206, Article 118045. <https://doi.org/10.1016/j.energy.2020.118045>. Article.

- Kingma, D. P., & Ba, J. L. (2014). *Adam: A method for stochastic optimization*. Cornell University. <https://doi.org/10.48550/arxiv.1412.6980>. arXiv.
- Kontokosta, C. E., & Tull, C. (2017). A data-driven predictive model of city-scale energy use in buildings. *Applied Energy*, 197, 303–317. <https://doi.org/10.1016/j.apenergy.2017.04.005>
- Koral Iseri, O., Duran, A., Canlı, I., Meral Akgül, C., Kalkan, S., & Gursel Dino, I. (2025). A method for zone-level urban building energy modeling in data-scarce built environments. *Energy and Buildings*, 337, Article 115620. <https://doi.org/10.1016/j.enbuild.2025.115620>. Article.
- Lei, L., Chen, W., Wu, B., Chen, C., & Liu, W. (2021). A building energy consumption prediction model based on rough set theory and deep learning algorithms. *Energy and Buildings*, 240, Article 110886. <https://doi.org/10.1016/j.enbuild.2021.110886>. Article.
- Li, K., Hu, C., Liu, G., & Xue, W. (2015). Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis. *Energy and Buildings*, 108, 106–113. <https://doi.org/10.1016/j.enbuild.2015.09.002>
- Li, Z., Dai, J., Chen, H., & Lin, B. (2019). An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. *Building Simulation*, 12(4), 665–681. <https://doi.org/10.1007/s12273-019-0538-0>
- Lin, D., Xu, X., Liu, K., Wu, T., Wang, X., & Zhang, R. (2025). Interpretable data-driven urban building energy modeling considering inter-building effect. *Building and Environment*, 274, Article 112688. <https://doi.org/10.1016/j.buildenv.2025.112688>. Article.
- Lin, Q., Liu, K., Hong, B., Xu, X., Chen, J., & Wang, W. (2022). A data-driven framework for abnormally high building energy demand detection with weather and block morphology at community scale. *Journal of Cleaner Production*, 354, Article 131602. <https://doi.org/10.1016/j.jclepro.2022.131602>. Article.
- Liu, Y., Chen, H., Zhang, L., Wu, X., & Wang, X. (2020). Energy consumption prediction and diagnosis of public buildings based on support vector machine learning: A case study in China. *Journal of Cleaner Production*, 272, Article 122542. <https://doi.org/10.1016/j.jclepro.2020.122542>. Article.
- Lundberg, S., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *31st Conference on Neural Information Processing Systems (NIPS 2017)*.
- Maas, A. L., Hannun, A. Y., & Ng, A. Y. (2013). Rectifier Nonlinearities Improve Neural Network Acoustic Models. In *28. 30th International Conference on Machine Learning*.
- McLeod R, Mead K, & Standen M. (2012). *Passivhaus primer: Designer's guide* https://passivhaus-international.org/upload/BRE_Passivhaus_Designers_Guide.pdf.
- Ministry of Environment, Urbanization, and Climate Change (2025). 2025 Unit Price Document.
- Moazami, A., Nik, V. M., Carlucci, S., & Geving, S. (2019). Impacts of future weather data typology on building energy performance – Investigating long-term patterns of climate change and extreme weather conditions. *Applied Energy*, 238, 696–720. <https://doi.org/10.1016/j.apenergy.2019.01.085>
- Na, W., & Wang, M. (2022). A Bayesian approach with urban-scale energy model to calibrate building energy consumption for space heating: A case study of application in Beijing. *Energy*, 247, Article 123341. <https://doi.org/10.1016/j.energy.2022.123341>. Article.
- Nagpal, S., & Reinhart, C. F. (2018). A comparison of two modeling approaches for establishing and implementing energy use reduction targets for a university campus. *Energy and Buildings*, 173, 103–116. <https://doi.org/10.1016/j.enbuild.2018.05.035>
- Nair, V., & Hinton, G. E. (2010). Rectified Linear Units Improve Restricted Boltzmann Machines. In *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*.
- National Glass. (2023). ENERGY PERFORMANCE DATA. Retrieved from <https://www.nationalglass.com.au/assets/main/Energy-Performance-Data-v36-min-2.pdf>. Accessed December 17, 2023.
- Dayak, S. G., Shrestha, S., Kinney, P. L., Ross, Z., Sheridan, S. C., Pantea, C. I., Hsu, W. H., Muscatiello, N., & Hwang, S. A. (2018). Development of a heat vulnerability index for New York State. *Public Health*, 161, 127–137. <https://doi.org/10.1016/j.puhe.2017.09.006>
- Nguyen, A. T., Rockwood, D., Doan, M. K., Le, Dung, & T, K. (2021). Performance assessment of contemporary energy-optimized office buildings under the impact of climate change. *Journal of Building Engineering*, 35, Article 102089. <https://doi.org/10.1016/j.jobe.2020.102089>. Article.
- Noaman, D. S., & El-Ghafour, S. A. (2024). Holistic design of energy-efficient temporary houses: Meeting ventilation, heating, cooling, and lighting demands. *Journal of Building Engineering*, 86, Article 108534. <https://doi.org/10.1016/j.jobe.2024.108534>. Article.
- 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, Article 103406. <https://doi.org/10.1016/j.jobe.2021.103406>. Article.
- Tootkaboni, P., M Ballarini, I., & Corrado, V. (2021). Analysing the future energy performance of residential buildings in the most populated Italian climatic zone: A study of climate change impacts. *Energy Reports*, 7, 8548–8560. <https://doi.org/10.1016/j.egyr.2021.04.012>
- Pan, Y., Zhu, M., Lv, Y., Yang, Y., Liang, Y., Yin, R., Yang, Y., Jia, X., Wang, X., Zeng, F., Huang, S., Hou, D., Xu, L., Yin, R., & Yuan, X. (2023). Building energy simulation and its application for building performance optimization: A review of methods, tools, and case studies. *Advances in Applied Energy*, 10, Article 100135. <https://doi.org/10.1016/j.adapen.2023.100135>. Article.
- Paudel, S., Elmirti, M., Couturier, S., Nguyen, P. H., Kamphuis, R., Lacarrière, B., & Le Corre, O. (2017). A relevant data selection method for energy consumption prediction of low energy building based on support vector machine. *Energy and Buildings*, 138, 240–256. <https://doi.org/10.1016/j.enbuild.2016.11.009>
- Peacock, A. D., Jenkins, D. P., & Kane, D. (2010). Investigating the potential of overheating in UK dwellings as a consequence of extant climate change. *Energy Policy*, 38(7), 3277–3288. <https://doi.org/10.1016/j.enpol.2010.01.021>
- Pham, A. D., Ngo, N. T., Ha Truong, T. T., Huynh, N. T., & Truong, N. S. (2020). Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *Journal of Cleaner Production*, 260, Article 121082. <https://doi.org/10.1016/j.jclepro.2020.121082>. Article.
- Pohoryles, D. A., Maduta, C., Bournas, D. A., & Kouris, I. A. (2020). Energy performance of existing residential buildings in Europe: A novel approach combining energy with seismic retrofitting. *Energy and Buildings*, 223, Article 110024. <https://doi.org/10.1016/j.enbuild.2020.110024>. Article.
- Rahman, A., Srikanth, V., & Smith, A. D. (2018). Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks. *Applied Energy*, 212, 372–385. <https://doi.org/10.1016/j.apenergy.2017.12.051>
- Ramos, D., Faria, P., Morais, A., & Vale, Z. (2022). Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building. *Energy Reports*, 8, 417–422. <https://doi.org/10.1016/j.egyr.2022.01.046>
- Reinhart, C. F., & Davila, C. C. (2015). Urban building energy modeling – A review of a nascent field. *Building and Environment*, 97, 196–202. <https://doi.org/10.1016/j.buldev.2015.12.001>
- Rey-Hernández, J. M., Yousif, C., Gatt, D., Velasco-Gómez, E., San José-Alonso, J., & Rey-Martínez, F. J. (2018). Modelling the long-term effect of climate change on a zero energy and carbon dioxide building through energy efficiency and renewables. *Energy and Buildings*, 174, 85–96. <https://doi.org/10.1016/j.enbuild.2018.06.006>
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why Should I Trust You? In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 1135–1144). <https://doi.org/10.1145/2939672.2939778>
- Rodrigues, E., Fernandes, M. S., & Carvalho, D. (2023). Future weather generator for building performance research: An open-source morphing tool and an application. *Building and Environment*, 233, Article 110104. <https://doi.org/10.1016/j.buldev.2023.110104>. Article.
- Roudsari, M. S., & Pak, M. (2013). LADYBUG: A PARAMETRIC ENVIRONMENTAL PLUG-IN FOR GRASSHOPPER TO HELP DESIGNERS CREATE AN ENVIRONMENTALLY-CONSIOUS DESIGN. In *Proceedings of BS 2013: 13th Conference of the International Building Performance Simulation Association* (pp. 3128–3135).
- Seyedzadeh, S., Pour Rahimian, F., Rastogi, P., & Glesk, I. (2019). Tuning machine learning models for prediction of building energy loads. *Sustainable Cities and Society*, 47, Article 101484. <https://doi.org/10.1016/j.scs.2019.101484>. Article.
- Shcherbakov, M., Kamaev, V., & Shcherbakova, N. (2013). Automated electric energy consumption forecasting system based on decision tree approach. *IFAC Proceedings Volumes*, 46(9), 1027–1032. <https://doi.org/10.3182/20130619-3-RU-3018.00486>
- Shen, P., Braham, W., & Yi, Y. (2019). The feasibility and importance of considering climate change impacts in building retrofit analysis. *Applied Energy*, 233–234, 254–270. <https://doi.org/10.1016/j.apenergy.2018.10.041>
- Sokol, J., Cerezo Davila, C., & Reinhart, C. F. (2017). Validation of a Bayesian-based method for defining residential archetypes in urban building energy models. *Energy and Buildings*, 134, 11–24. <https://doi.org/10.1016/j.enbuild.2016.10.050>
- Somu, N., Raman M R, G., & Ramamirtham, K. (2021). A deep learning framework for building energy consumption forecast. *Renewable and Sustainable Energy Reviews*, 137, Article 110591. <https://doi.org/10.1016/j.rser.2020.110591>. Article.
- Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *Proceedings of the 34th International Conference on Machine Learning*, 70, 3319–3328. <https://proceedings.mlr.press/v70/sundararajan17a.html>
- Swan, L. G., & Ugursal, V. I. (2009). Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. *Renewable and Sustainable Energy Reviews*, 13(8), 1819–1835. <https://doi.org/10.1016/j.rser.2008.09.033>
- Szagri, D., Nagy, B., & Szalay, Z. (2023). How can we predict where heatwaves will have an impact? – A literature review on heat vulnerability indexes. *Urban Climate*, 52, Article 101711. <https://doi.org/10.1016/j.ulclim.2023.101711>. Article.
- Tamer, T., Gürsel Dino, I., & Meral Akgül, C. (2022). Data-driven long-term prediction of building performance under climate change: Building energy demand and BIPV energy generation analysis across Turkey. *Renewable and Sustainable Energy Reviews*, 162, Article 112396. <https://doi.org/10.1016/j.rser.2022.112396>. Article.
- Tian, C., Ye, Y., Lou, Y., Zuo, W., Zhang, G., & Li, C. (2022). Daily power demand prediction for buildings at a large scale using a hybrid of physics-based model and generative adversarial network. *Building Simulation*, 15(9), 1685–1701. <https://doi.org/10.1007/s12273-022-0887-y>
- Todeschi, V., Boghetti, R., Kämpf, J. H., & Mutani, G. (2021). Evaluation of urban-scale building energy-use models and tools—application for the city of Fribourg, Switzerland. *Sustainability (Switzerland)*, 13(4), 1–22. <https://doi.org/10.3390/su13041595>
- Tomrukcu, G., & Ashrafiyan, T. (2024). Climate-resilient building energy efficiency retrofit: Evaluating climate change impacts on residential buildings. *Energy and Buildings*, 316, Article 114315. <https://doi.org/10.1016/j.enbuild.2024.114315>. Article.
- Touzani, S., Granderson, J., & Fernandes, S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. *Energy and Buildings*, 158, 1533–1543. <https://doi.org/10.1016/j.enbuild.2017.11.039>
- Tran, D. H., Luong, D. L., & Chou, J. S. (2020). Nature-inspired metaheuristic ensemble model for forecasting energy consumption in residential buildings. *Energy*, 191, Article 116552. <https://doi.org/10.1016/j.energy.2019.116552>. Article.
- Triana, M. A., Lamberts, R., & Sassi, P. (2018). Should we consider climate change for Brazilian social housing? Assessment of energy efficiency adaptation measures. *Energy and Buildings*, 158, 1379–1392. <https://doi.org/10.1016/j.enbuild.2017.11.003>

- Troup, L., Eckelman, M. J., & Fannon, D. (2019). Simulating future energy consumption in office buildings using an ensemble of morphed climate data. *Applied Energy*, 255, Article 113821. <https://doi.org/10.1016/j.apenergy.2019.113821>. Article.
- TSE. (2013). *TS825 Binalarda isti yahutum kuralları*.
- TURKSTAT. (2021). *Türkiye Family Structure Survey*. https://data.tuik.gov.tr/Buletin/Inde_x?p=Turkiye-Family-Structure-Survey-2021-45813&dil=2.
- Wang, Z., Hong, T., Li, H., & Ann Piette, M. (2021). Predicting city-scale daily electricity consumption using data-driven models. *Advances in Applied Energy*, 2, Article 100025. <https://doi.org/10.1016/j.adapen.2021.100025>. Article.
- Wang, Z., Wang, Y., Zeng, R., Srinivasan, R. S., & Ahrentzen, S. (2018). Random Forest based hourly building energy prediction. *Energy and Buildings*, 171, 11–25. <https://doi.org/10.1016/j.enbuild.2018.04.008>
- Waskom, M. (2024). seaborn: statistical data visualization. *The Journal of Open Source Software*, 6(60), Article 3021. <https://doi.org/10.21105/joss.03021>. Article.
- WeatherShift. (n.d.). Retrieved from <https://www.weathershift.com/>. Accessed December 16, 2023.
- Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, 3(1), Article 9. <https://doi.org/10.1186/s40537-016-0043-6>. Article.
- Ye, Y., Strong, M., Lou, Y., Faulkner, C. A., Zuo, W., & Upadhyaya, S. (2022). Evaluating performance of different generative adversarial networks for large-scale building power demand prediction. *Energy and Buildings*, 269, Article 112247. <https://doi.org/10.1016/j.enbuild.2022.112247>. Article.
- Yu, S., Wang, K., & Wei, Y. M. (2015). A hybrid self-adaptive Particle Swarm Optimization-Genetic Algorithm-Radial Basis Function model for annual electricity demand prediction. *Energy Conversion and Management*, 91, 176–185. <https://doi.org/10.1016/j.enconman.2014.11.059>
- Zhai, Z. J., & Helman, J. M. (2019). Implications of climate changes to building energy and design. *Sustainable Cities and Society*, 44, 511–519. <https://doi.org/10.1016/j.scs.2018.10.043>
- Zou, Y., Xiang, K., Zhan, Q., & Li, Z. (2021). A simulation-based method to predict the life cycle energy performance of residential buildings in different climate zones of China. *Building and Environment*, 193, Article 107663. <https://doi.org/10.1016/j.buildenv.2021.107663>. Article.