Sustainable Cities and Society 114 (2024) 105723

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

Passive over active: How low-cost strategies influence urban energy equity

Siavash Ghorbany [a,*], Ming Hu [b,c], Matthew Sisk [d], Siyuan Yao [e], Chaoli Wang [f]

[a] Department of Civil and Environmental Engineering and Earth Sciences, College of Engineering, University of Notre Dame, Notre Dame, IN 46556, United States
[b] Associate Dean for Research, Scholarship, and Creative Work, School of Architecture, Walsh Family Hall of Architecture, University of Notre Dame, Notre Dame, IN 46556, United States
[c] Faculty of Architecture, Civil Engineering and Applied Arts, Academy of Silesia, Rolan 43, Katowice 40-555, Poland
[d] Associate Professor of the Practice, Lucy Family Institute for Data & Society, University of Notre Dame, Notre Dame, IN 46556, United States
[e] Department of Computer Science and Engineering, College of Engineering, University of Notre Dame, Notre Dame, IN 46556, United States
[f] Department of Computer Science and Engineering, University of Notre Dame, Notre Dame, IN 46556, United States

ARTICLE INFO

Keywords:
Energy burden
Sustainable cities
Machine learning
Neural networks
Energy justice

ABSTRACT

This study delves into the energy burden on households, a crucial aspect of energy justice, influenced by urban environment factors and buildings' passive and active designs. It evaluates the effects of passive and active design features on household energy expenditures at the census tract scale. Applying advanced Machine Learning techniques, including multiple and decision tree regressions, random forests, support vector machines, XGBoost, and Neural Networks, the research assesses the impact of various factors on the energy burden. Findings reveal that passive design elements significantly outweigh active ones in reducing energy costs at the urban scale, as confirmed by a model with a 94.8 % R2 accuracy. The insights provided are vital for policymakers, urban planners, architects, and researchers, pushing for sustainable urban planning and energy justice by prioritizing effective design strategies. This contributes to a broader understanding and implementation of energy-efficient measures in urban development.

1. Introduction and background

1.1. *Energy burden*

Energy burden is defined as the value of energy-related expenditures that each household pays compared to their gross income (Department of Energy, 2023). The energy-related expenditure primarily refers to the utility bill (e.g., heating and cooling). The Average Energy Burden (AEB) is calculated from the average value of these expenditures in a specific region, such as census tracts. A census tract is a relatively permanent subdivision of a county that the U.S. Census Bureau uses to help organize population data for statistical purposes, typically containing between 1200 and 8000 residents (United States Census Bureau, 2024). Over 46 million people in the United States, about 40 % of the population, pay a significant proportion of their income on the energy burden, sometimes exceeding 10 % of their gross income (Drehobl et al., 2020). This difference not only impacts energy justice in the United States such as affording the necessary heating and cooling appliances, but also leads to health issues which are constantly arising due to the current climate change patterns and extreme temperature events (Drehobl et al., 2020; Ghorbany et al., 2024b; Guo et al., 2018). Moreover, studies show that some specific demographic groups are more prone to energy justice issues, which leads to equity issues in society (Drehobl et al., 2020). Directly addressing poverty requires high-level, multidimensional decisions, making solutions difficult and expensive (Ajebe, 2024; Y. Su et al., 2023).

1.2. *Passive and active strategies*

Passive and active strategies are two distinct principles in building science and design. Passive strategy employs solutions that maximize the efficiency and comfort of the built environment by utilizing natural resources such as sunlight and ventilation (Ochoa and Capeluto, 2008). Active strategies, conversely, involve the use of mechanical and electrical devices, including heating systems, to provide comfort for building occupants (Ochoa & Capeluto, 2008; Wigginton & Harris, 2013).

Historically, passive approaches have significantly lessened energy consumption before the advent of modern mechanical systems. Contemporary studies affirm passive strategies' efficacy, particularly in hot climates, with findings suggesting they can decrease cooling loads

* Corresponding author.
E-mail address: sghorban@nd.edu (S. Ghorbany).

https://doi.org/10.1016/j.scs.2024.105723
Received 16 May 2024; Received in revised form 17 July 2024; Accepted 30 July 2024
Available online 4 August 2024

and energy usage by an average of 31 % and 29 %, respectively, (Hu et al., 2023). Other studies in the United States showed that insulated building materials in roofs and ceilings potentially reduce energy consumption by 35 to 45 % while wall insulation can add another 15 % (Shahee et al., 2024). The features of passive strategies are renewable and require low technology, thus reducing energy consumption without imposing additional financial burdens on households (Bhamare et al., 2019; Campaniço et al., 2014; Chen & Yang, 2017). In the contemporary context, the importance of these two principles on energy consumption in buildings has always been an area of debate (Ochoa & Capeluto, 2008; Salameh & Touqan, 2022; Zune et al., 2020). In addition, the lack of data on some of the passive design features, such as the window-to-wall ratio (WWR) and external shading up to recently, has made performing a comprehensive comparison a challenge (Ghorbany et al., 2024b). Other studies related to WWR and external shading were either concentrated on an individual building and not the urban scale (Ashrafian & Moazzen, 2019; Harmati & Magyar, 2015; Javad & Navid, 2019), or used simplified methods to assess the outdoor shading at the design stage and not the actual existing buildings in the cities (Elrefai & Nikolopoulou, 2023).

Prior research on other passive strategies have also focused mainly on isolated buildings, often overlooking its impact on larger urban areas and demographic variations at the census tract level. Some of these studies also considered the passive and active design features together. Sun et al. (2018) worked on a case study in Southeast Asia as the first zero-energy building and investigated the cost efficiency of passive and active design strategies. Other researchers in China worked on the collaborative optimization for the heating systems and passive design to reduce the cost and energy consumption (Wang et al., 2020). Kang et al. (2015) worked on comparing the impact of passive design against the active strategies and showed that passive strategies are more beneficial in terms of energy saving. However, this research was focused on a school building (Kang et al., 2015). What is worth noting is that these studies predominantly originate from developing nations, such as Mexico (Vargas & Hamui, 2021), Nigeria (Onyenokporo & Ochedi, 2019), Egypt (Abdallah, 2022; B. Su, 2011), Spain (Diz-Mellado et al., 2023), and China (Cheung et al., 2005; Han et al., 2017; Ling & Jin, 2018; Yao et al., 2018) while only a few studies concentrated on developed countries such as the United States (Ghorbany et al., 2024b; Hu et al., 2023).

Furthermore, a few studies worked on passive and active design strategies on an urban scale. Bouketta, (2023) worked on investigating the impact of Urban Cool Island as a passive strategy for improving the summer conditions (Bouketta, 2023). However, this research neither directly worked on energy nor concentrated on buildings' impact in urban cool islands. Another study also worked on the impact of passive design in thermal conditions but this study was not focused on active and passive design comparison and its impact on energy (Shaeri et al., 2018). The current literature shows that current studies were not directly related to energy burden behavior and prediction and were mostly considered only one of the passive or active design strategies (Azimi Fereidani et al., 2021; Cheshmehzangi & Dawodu, 2020; Salameh & Touqan, 2022). Therefore, a research to cover this gap and propose a framework to compare these two strategies on an urban scale is necessary.

In addressing urban energy burdens and the effectiveness of different strategies, Machine Learning (ML) has emerged as a powerful tool for predictive analysis, which is due to its exceptional effectiveness in deciphering the intricate relationship between variables and producing predictions (Bukkapatnam et al., 2019; Ghorbany et al., 2024a; Rebala et al., 2019). In the last decades, Fan et al., (2017) worked on predicting cooling factors using deep learning algorithms (Fan et al., 2017) while Bektas Ekici and Aksoy (2011) utilized an Adaptive Network Based Inference System (ANFIS) for predicting energy consumption in both cooling and heating systems and could find a good combination between these two energies (Bektas Ekici & Aksoy, 2011). Another study in this field used artificial neural networks for cooling load prediction (Kwok & Lee, 2011). Sholahudin and Han (2016) used neural networks in heating energy consumption prediction in individual buildings (Sholahudin & Han, 2016). Another study used LSTM for predicting building energy consumption (Li et al., 2022). Although these studies found that machine learning models can be efficient models in energy use cases, their application has been primarily confined to individual energy types and early design stages rather than holistic urban energy consumption predictions incorporating building characteristics.

This research seeks to bridge the gap by comparing passive and active strategies at an urban scale and employing ML to identify the most influential factors for predicting AEBs. The passive design factors investigated in this research include the wall material, roof material, window-to-wall ratio, external shade, and land coverage while the active design factors included the proportion of heating and cooling systems such as central air, warm air heating, hot water steam heating, unit heater heating systems, floor furnaces, and solar.

To address this issue, this research aims to (1) conduct a comprehensive comparison between the passive design factors and active design main components by combining the different data sources and creating a first-of-it-kind dataset for statistical analysis, (2) identify the most significant variables that are impacting the AEB in the census tracts using the interpretable regression methods, and (3) develop an ML model to accurately predict AEB in Chicago Cook County census tracts. The findings of this research indicate that passive, rather than active design factors play a more crucial role in changing energy expenditures on the urban scale. The findings of this research provide insights for policymakers, urban planners, architects, and future researchers to move toward a sustainable and equitable urban environment and the act of energy justice.

2. Methodology

Fig. 1 demonstrates the general flowchart of this study. This research methodology includes three major steps, namely data collection, data processing, and data analysis. The proposed methodology was conducted on the data from the metropolitan area of Chicago, located in a hot-summer humid continental climate zone in the Midwest region of the United States as a representative of a developed country major city. In the data collection section different data sources including the GSV data (Ghorbany et al., 2024b), National Structure Inventory (NSI) data (NSI, 2024), Cook County Open Data (Cook County Assessor's Office, 2022), and LEAD and ACS data (Ma et al., 2019) were examined to extract the required variables for assessing the passive design and building characteristics' influence on AEB. To merge these different data sources without data loss, a pre-processing stage was conducted. Finally, in the data analysis stage, the prepared dataset was investigated by statistical and ML methods to assess the impact of each variable on the AEB on the households and develop the best model for predicting it. The details of each step have been provided in the following sections.

2.1. Data sources

The dispersion of the passive design and building characteristics data is one of the bottlenecks in the AEB analysis that needs to be addressed before any further analysis. Different data resources are used in this research to overcome this challenge. Table 1 demonstrates the variables extracted from each of the data sources and the description of these datasets can be found in APPENDIX 2.

2.2. Data processing

After collecting the required datasets, the data needed to be changed from the individual building scale to the census tract scale to match the energy burden. To accomplish this, each dataset was first treated separately. In the beginning, the missing data were cleaned from the

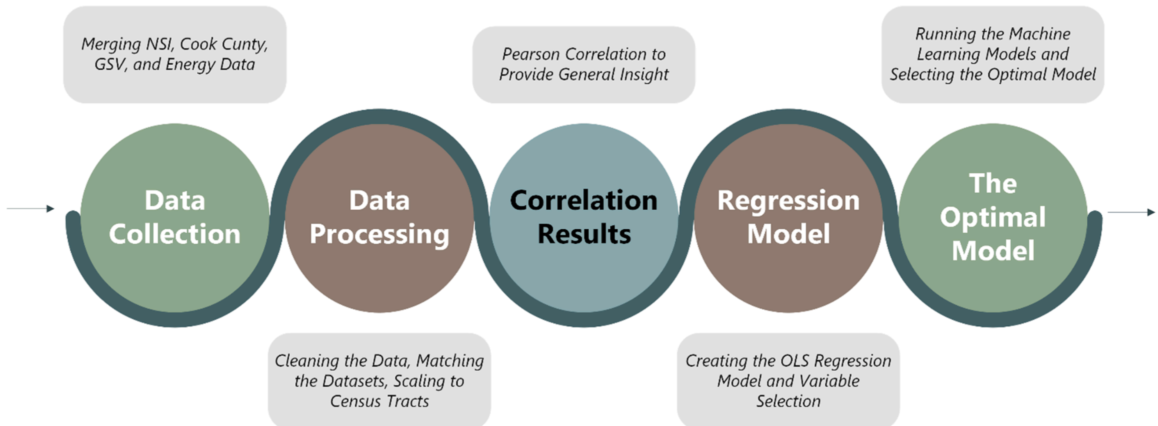

Fig. 1. The research methodology flowchart.

Table 1
The variables distribution among the data sources.

Dataset	Category	#Variables	Variables used
(Ghorbany et al., 2024b)	GSV Data	2	Window to Wall Ratio (WWR), External Shading
(NSI, 2024)	NSI Data	6	Structure Type, Foundation Type, Foundation Height, Square Footage, Number of Stories, Year Built
(Cook County Assessor's Office, 2022)	Cook County	9	Longitude, Latitude, Wall Material, Roof Material, Building Square Feet, Land Square Feet, Central Heating, Other Heating, Central Air
(Ma et al., 2019)	LEAD Data	1	AEB
(U.S. Census Bureau, 2022)	ACS Data	2	Population with Poverty Status Percentage, The Over 65 Years Population Percentage
Total		20	

datasets. Afterward, the land cover was calculated from the building and land square feet for the Cook County Data. Then, the average of the central air variable was calculated for each census tract. For the categorical variables, which included wall material, roof material, central heating, and other heating, the proportion of each category for the census tract was calculated. Then, all the datasets were grouped by the census tract FIPS code. FIPS is a unique code assigned to each census tract in the United States.

The same steps were done on the NSI dataset. After cleaning the missing data, the categorical variables, including the building structure type and building foundation type, were calculated based on the proportion of each category in these variables in each Census Tract. The other variables were averaged and grouped by the census tract FIPS code. Moreover, the extracted data from CNN GSV model were averaged based on the census tracts in Illinois. FIPS code was also assigned to each extracted data from images to make the data merging feasible in future steps. After calculating the required parameters and transforming each dataset to the census tract level, the data was merged based on the FIPS code. In total, 1135 rows of data remained, meaning that there were 1135 out of 1319 census tract in Chicago that match between four datasets for Chicago city. The statistical description of the merged dataset is demonstrated in APPENDIX 1. In this table, the mean value, standard deviation of each variable, minimum, maximum, and quantiles of the data are reported.

2.3. Data analysis

2.3.1. Correlation and multicollinearity

After creating the research data frame, Python was used to implement the statistical analysis. To generate a general insight into the variable's relationship and develop the initial hypothesis, the Pearson Correlation was used (Saidi et al., 2019).

Afterward, the Variance Inflation Factor (VIF) was checked to assess the multicollinearity among the variables. Since the linear regression model coefficients are interpreted with the assumption of keeping other variables constant while changing one independent variable, the multicollinearity is one of the factors that violate the Ordinary Least Squares (OLS) assumptions and impacts its interpretability (Ghorbany et al., 2024b; Lee et al., 2023). The VIF is calculated through Eq. (1) where the coefficient of determination, or R2, is found by regressing the main independent variable in the model against each of the other independent variables.

$$VIF = \frac{1}{1 - R^2} \tag{1}$$

The estimation of VIF gives an idea about the variables that need to be eliminated to keep the model's interpretability.

2.3.2. OLS regression and machine learning

OLS regression is the most interpretable ML model that provides an idea about the interaction of the variables and how they are impacting the dependent variable, in here AEB (Burton, 2021). Therefore, the OLS was used in this study to extract the influential passive design factors on AEB among all the defined variables in Section 2.1. The coefficients of the OLS method are calculated based on Eq. (2) where X is the matrix of the independent variables and Y is the vector of dependent variables.

$$\widehat{\beta} = \left(X^T X\right)^{-1} X^T Y \tag{2}$$

After extracting the $\widehat{\beta}$ of the OLS regression model, the P-Value of the coefficients is calculated and the values below 0.05 are considered significant. Otherwise, the variables can be deleted upon consideration of R^2 and R^2 adjusted difference and the importance of model's interpretability.

After selecting the appropriate variables and interpreting the results, more advanced ML regression models were run to check whether there is a higher prediction power can be achieved through these models. The Support Vector Regression (SVR), decision tree regression, random forest regression, and XGBoost regression models were used to find the most optimal model for average energy prediction based on building characteristics. All the models went into a fine-tuning grid search to find the best parameters and were tested through a 5-fold cross-validation to validate the accuracy and error of the models (Ghorbany et al., 2024b;

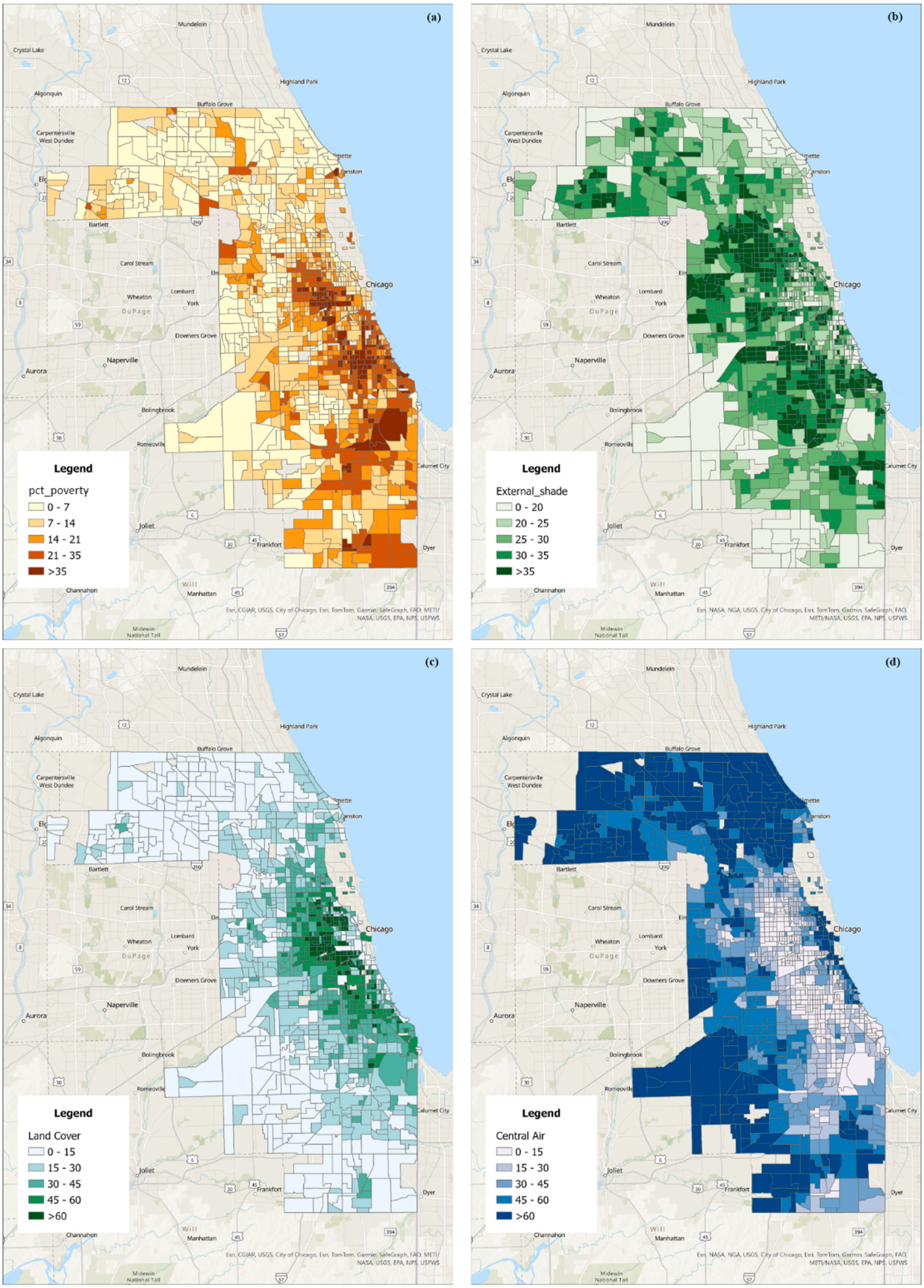

Fig. 2. Population Below Poverty Level (pct_poverty) (a), External Shading (b), Land Coverage (c), and Central Air Condition (d) Data Distribution in Chicago.

He et al., 2023; Nti et al., 2021). The R^2, Root Mean Square Error (RMSE), Mean Square Error (MSE), and Mean Absolute Error (MAE) were used as the performance measures of these models in this study (see Eqs. (3)–(6)) (Ghorbany et al., 2023; Pham, 2019).

$$R^2 = 1 - \frac{RSS}{TSS} \tag{3}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i^p\right)^2}{n}} \tag{4}$$

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i^p - y_i\right)^2 \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i^p - y_i\right| \tag{6}$$

In these equations, y_i indicates the predicted value, y_i^p is the estimated data, $\overline{y}^p$ demonstrate the mean value, and n indicates the sample size. Moreover, RSS is the sum of squares of residuals and TSS is the total sum of squares.

3. Housing physical condition data

3.1. Descriptive data analysis results

The extraction of general statistical characteristics of the data was the first step to provide a general insight into the dataset. APPENDIX 1 demonstrates the results of the statistical description of the merged dataset. Only 1133 census tracts were used for data analysis as building square footage and land cover data were only available for those tracts out of the 1135 tracts identified in Cook County.

In Cook County, wall material proportions were categorized into wood, masonry, combined wood and masonry, and stucco, with masonry being the most common (average 37.67 %) and stucco least common (average 1.14 %). Roofing materials were grouped into six types, including shingle/asphalt, tar and gravel, shake roof, tile roof, slate roof, and other roofs. Shingle/asphalt was the most prevalent,

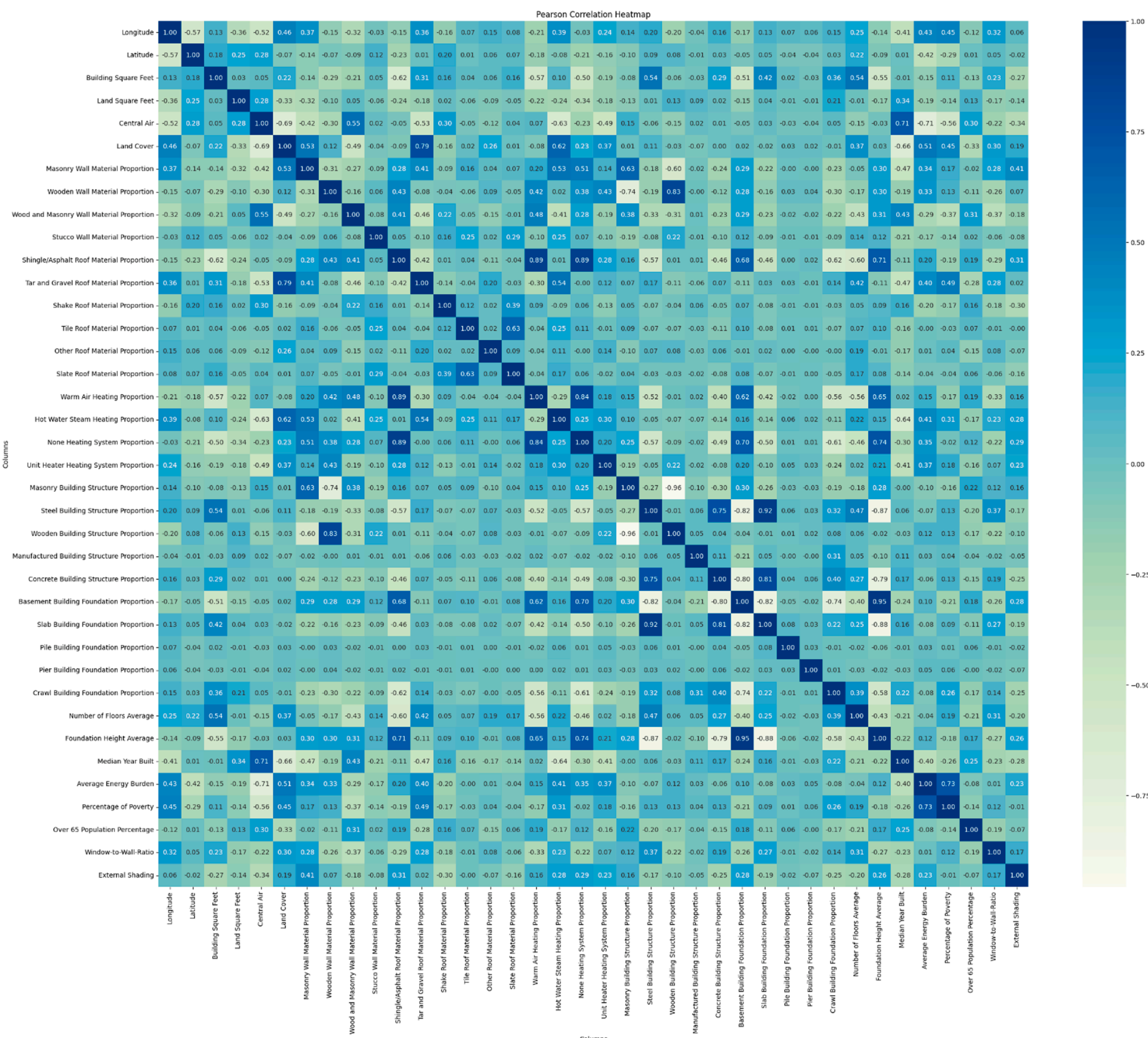

Fig. 3. Pearson correlation among the study variables.

averaging 67.87 %, whereas slate was the least, at only 0.32 % average but reaching 14.69 % in some areas. This indicates a preference for asphalt and shingle roofs. Central heating types included warm air, hot water steam, and electric; notable was the absence of specific heating categories in local use. In terms of ventilation, about 39.19 % of buildings had central air conditioning that also controls the mechanical ventilation, window-to-wall ratios ranged from 15.15 % to 25.68 %. As for external shading, it was present in 30.27 % on average, up to 62.24 % in certain areas, reflecting diverse natural and mechanical ventilation practices in Cook County, Chicago. Fig. 2 demonstrates the distribution of some of these variables in Chicago.

3.2. Statistical data analysis results

3.2.1. Correlation insights: socioeconomic and geographical determinants of energy burden

Upon examining the data distribution, Pearson Correlation was deployed to provide a perspective on the relationship between the covariates and AEB as well as the collinearity between the variables themselves. Fig. 3 shows poverty (pct_poverty) and central air conditioning systems have the most substantial correlation to energy burden, at 0.73 and -0.71, respectively. Poverty's positive correlation with energy burden suggests that it concurrently rises. Conversely, central air conditioning's negative correlation implies more such systems correspond with lower energy burden, potentially reflecting the newer buildings with enhanced passive strategies (appropriate window-to-wall ratio), as indicated by the 0.71 correlation between building year and air conditioning. This may be due to the income and socioeconomic parameters since there is an interesting pattern between the central air systems usage and regions with lower poverty as demonstrated in Fig. 2. It is noteworthy that the correlation does not necessarily mean a cause-and-result relationship between the variables.

Additional variables like median year built, wood wall materials, masonry wall materials, wood and masonry combination wall material, latitude, longitude, heating systems except warm air, tar and gravel roof material, and land cover also show notable correlation with energy burden in Table 2. These relationships provide a framework for understanding factors influencing energy expenses and burdens.

As shown in Table 2, latitude and longitude indicate the geographical location of the census tracts, which is another factor that has a correlation larger than 0.4 with the AEB. Longitude correlates positively (0.43), indicating that as census tracts shift eastward, the energy burden tends to increase. This is while as demonstrated in Fig. 2, the eastern areas in Chicago have denser land covers, suggesting the increase in building density might be a potential reason for more energy demand in these areas. The latitude, conversely, has a negative correlation (-0.42), suggesting a decreased energy burden when moving northward, as shown in Fig. 3. These trends may reflect climatic variations like wind and humidity, which might be influenced due to the alignment of this city to Lake Michigan and can create microclimates inside the city (Alfraihat et al., 2016). This also might be due to socioeconomic patterns such as wealth distribution near or further to water bodies. To elaborate on this, as demonstrated in Fig. 2, the northern and western areas of Chicago have generally lower poverty percentages which suggests that this trend can be accelerated by the financial issues in these areas and also it might be impacting the marginalized areas more. Moreover, Fig. 2 shows that these areas with more poverty have lower central air conditioning systems suggesting that other types of cooling and heating is responsible for the energy burden in these areas. Additionally, the "combined wood and masonry" wall category in Fig. 3 highlights that the combination of wood and masonry as wall materials (e.g., bricks being used to a certain height of the building and wood for the rest) correlates differently with AEB than when these materials are used separately. Therefore, the significance and influence of these variables are further examined using the Ordinary Least Squares (OLS) model to understand the impact of those variables.

Table 2
Variables with relatively high (higher than 0.3) negative and positive correlations with AEB.

Mode	Variables
Positive Correlation	heating systems except for warm air, tar and gravel roof material, wood wall materials, masonry wall materials, land cover, longitude, poverty percentage
Negative Correlation	Median year built, wood and masonry combination wall material, latitude, central air conditioning

3.2.2. OLS regression results

To initial the analysis, a regression model with AEB as the outcome and 36 predictors was conducted (see Table 3). The model is significant as indicated by a Prob F-statistic below 0.005 and an F-statistic value of 128, well above the usual significance threshold of 10-20 (Sureiman & Mangera, 2020). Consequently, the null hypothesis, stating that the variables are not associated with AEB, is rejected indicating the variables collectively are associated with the AEB. However, many P-values above 0.05 suggest some variables may not significantly differ from zero, implying a lack of individual impact. The closeness of R-squared to adjusted R-squared suggests these discrepancies are likely due to multicollinearity, advising caution before removing any variables.

Considering the multicollinearity presence, variables such as latitude, median year built, the percentage of poverty, population over 65 years old, window-to-wall ratio, external shading, central air conditioning, and land cover ratio remained significant, indicating certain association with the AEB. While many were anticipated by the Pearson correlation analysis (refer to Fig. 3), factors like age influence, window-to-wall ratio, and external shading introduced new insights. That means even though these variables were not among the highly correlated values in the correlation results (Table 2), OLS as a more robust model proved that these variables are significant to the energy burden.

To refine the model, variables with a P-value below 0.05 were retained, and then their variances inflation factor (VIF) was monitored. The adjustments in R-squared and Bayesian information criterion (BIC) were the criteria for the inclusion or removal of variables. Lower BIC values indicate an improved model over the previous iterations (Chakrabarti & Ghosh, 2011; Ghorbany et al., 2022). Table 4 shows the model after excluding variables that caused multicollinearity. Excluding Slate Roof Material Proportion, all retained variables have P-values below 0.05. Despite the marginal P-value of Slate Roof Material Proportion, the close R-squared and adjusted R-squared values justify keeping the current variables. This revised model enhances interpretability, and a comparison of the R2 values in Tables 3 and 4 shows it maintains predictive efficacy. Moreover, the Akaike information criterion (AIC) and BIC values are lower than the previous model, suggesting that this is a better fit, and the F-statistic increase from 128 to 291 supports this.

The regression model fits the data well, surpassing prior results by Ghorbany et al. (2024b), but the intercept's physical plausibility is questionable – suggesting a 34 % energy burden in the absence of urban development. The high P-value for the intercept in Table 3, however, hints at a plausible zero value in undeveloped areas. Removing the constant from the model, Table 5 shows the final model. The F-statistic rose to 1355 and both R-squared and adjusted R-squared values improved to 0.948 and 0.947, respectively. The AIC and BIC values are nearly unchanged. These measures, alongside logical model interpretation, suggest this is the superior model. It's R-squared raised significantly from 0.785 to 0.948, outperforming prior models by over 20 % (Ghorbany et al., 2024b).

In the developed model, the P-value of all the variables is below 0.05 suggesting that there is enough evidence to reject the null hypothesis and accept all variables associated with AEB. According to the coefficients presented in Table 5, the geographical location of the census tracts has the most influence on the AEB, suggesting that potentially

Table 3
General model with all variables and AEB as dependent variable.

Prob > F = 0			Dep. Variable: AEB			
F-statistic: 128.8			R-squared: 0.789			
No. Observations: 1133			Adj. R-squared: 0.783			
			AIC: 2693			
			BIC: 2859			
Variables	Coefficient	Std. Error	t-value	P-value	CI Lower	CI Upper
const	7.533	36.265	0.208	0.835	-63.622	78.689
Longitude	-0.072	0.349	-0.207	0.836	-0.756	0.612
Latitude	-1.223	0.263	-4.658	0.000	-1.739	-0.708
Building Square Feet	0.000	0.000	-1.717	0.086	0.000	0.000
Masonry Wall Material Proportion	0.243	0.579	0.420	0.674	-0.893	1.380
Wood Wall Material Proportion	0.249	0.579	0.430	0.668	-0.887	1.384
Wood and Masonry Wall Material Proportion	0.242	0.579	0.417	0.676	-0.894	1.378
Stucco Wall Material Proportion	0.195	0.579	0.338	0.736	-0.940	1.331
Shingle and Asphalt Roof Material Proportion	0.159	0.386	0.413	0.680	-0.598	0.917
Tar and Gravel Roof Material Proportion	0.155	0.386	0.400	0.689	-0.603	0.912
Shake Roof Material Proportion	0.166	0.386	0.430	0.667	-0.592	0.924
Tile Roof Material Proportion	0.120	0.387	0.311	0.756	-0.638	0.879
Other Roof Material Proportion	0.089	0.387	0.231	0.818	-0.670	0.849
Slate Roof Material Proportion	0.240	0.387	0.619	0.536	-0.520	0.999
Wooden Building Structure Proportion	0.082	7.354	0.011	0.991	-14.348	14.513
Manufactured Building Structure Proportion	9.735	8.586	1.134	0.257	-7.112	26.582
Masonry Building Structure Proportion	0.031	7.288	0.004	0.997	-14.269	14.332
Steel Building Structure Proportion	-1.361	7.408	-0.184	0.854	-15.897	13.176
Concrete Building Structure Proportion	-0.954	7.431	-0.128	0.898	-15.534	13.626
Basement Building Foundation Proportion	22.426	40.54	0.553	0.580	-57.118	101.971
Pier Building Foundation Proportion	-22.767	104.492	-0.218	0.828	-227.793	182.258
Pile Building Foundation Proportion	-37.058	169.691	-0.218	0.827	-370.013	295.897
Slab Building Foundation Proportion	23.582	40.482	0.583	0.560	-55.848	103.012
Crawl Building Foundation Proportion	21.350	40.526	0.527	0.598	-58.168	100.868
Number of Stories Mean	-0.022	0.135	-0.166	0.868	-0.288	0.243
Foundation Height Mean	0.002	0.556	0.004	0.997	-1.089	1.094
Median Year Built	0.009	0.003	2.937	0.003	0.003	0.015
Percentage of Poverty	0.069	0.003	24.515	0.000	0.063	0.074
Percentage of Over 65 Years	0.017	0.004	4.451	0.000	0.01	0.025
WWR	-0.069	0.019	-3.631	0.000	-0.106	-0.032
Land Cover	0.008	0.003	2.941	0.003	0.003	0.013
Central Air	-0.023	0.002	-10.416	0.000	-0.027	-0.018
Warm Air Heating Proportion	0.462	1.158	0.399	0.69	-1.809	2.734
Hot Water Steam Heating Proportion	0.467	1.158	0.403	0.687	-1.805	2.739
None Heating System Proportion	-0.854	2.123	-0.402	0.688	-5.019	3.311
Unit Heater Heating System Proportion	-0.862	2.123	-0.406	0.685	-5.028	3.303
External Shading	0.010	0.004	2.618	0.009	0.002	0.017

climate (e.g., microclimates caused by Lake Michigan or wind factors), and socioeconomic conditions are highly associated with energy use. Moreover, the latitude of the census tracts has the highest coefficient among all variables. That means a 1 1-degree increase in the location of the census tracts on average, decreases the AEB by 1.328 % in the census tract while the amount of this change is, on average, 0.455 % for the longitude.

As expected, poverty is the next most influential factor on the AEB, meaning a 1 % increase in the population below poverty, on average, increases the energy burden by 0.067 %. Moreover, with 95 % confidence, this change is between 0.062 to 0.072 % on the AEB. However, changing the poverty level requires a difficult multifactor economic decision which might have an adverse effect on many factors in the long time (Cheng & Ngok, 2020; Sachs, 2015; Šileika & Bekerytė, 2013). The findings from Table 5 suggest that passive design elements, specifically the window-to-wall ratio (WWR) and roofing material, exert a comparable influence on poverty-related changes. Notably, these passive design parameters can be adjusted relatively easily during the design and renovation processes, potentially mitigating additional financial strain on households (Ghorbany et al., 2024b). Moreover, the passive design factors investigated in this research (e.g., wall material, window-to-wall ratio) collectively as a group, impact the AEB more than the poverty ratio. This finding provides a substantial perspective for the policymakers and energy department for developing sustainable cities and expanding energy justice through passive design policymaking. According to the multiple regression results, the AEB can be calculated through Eq. (7).

$$
\begin{aligned}
\textit{Average Energy Burden} = {} & (-0.455)\times \text{Longitude} + (-1.328)\times \text{Latitude} \\
& + (0.067)\times \text{pct\_poverty} + (0.019) \\
& \times \text{Over 65 Years Population Percentage} \\
& + (-0.060)\times \text{WWR} + (-0.024)\times \text{Central Air} \\
& + (0.011)\times \text{External Shading} + (0.015) \\
& \times \text{Masonry Wall Material Proportion} + (0.020) \\
& \times \text{Wooden Wall Material Proportion} + (0.012) \\
& \times \text{Wood and Masonry Wall Material Proportion} \\
& + (-0.028)\times \text{Stucco Wall Material Proportion} \\
& + (0.009)\times \text{Median Year Built} + (0.004) \\
& \times \text{Land Cover} + (-0.062) \\
& \times \text{Other Roof Material Proportion} + (0.053) \\
& \times \text{Slate Roof Material Proportion}
\end{aligned} \tag{7}
$$

As evident in the P-Values from Table 5, among active design factors, only air conditioning association with energy burden is statistically significant and therefore linked to energy burden on a census tract scale, indicating that passive design elements are more significantly impactful for urban policy. Policymakers, architects, and urban planners should, therefore, focus more on passive design factors. The comparison in Fig. 4 shows that passive design elements have the greatest effect on

Table 4
Optimized model after removing the redundant variables.

Prob > F = 0 F-statistic: 291.3 No. Observations: 1133			Dep. Variable: AEB R-squared: 0.785 Adj. R-squared: 0.782 AIC: 2680 BIC: 2756			
Variables	Coefficient	Std. Error	t-value	P-value	CI Lower	CI Upper
const	34.094	10.347	3.295	0.001	13.793	54.396
Latitude	-1.186	0.187	-6.329	0.000	-1.553	-0.818
pct_poverty	0.067	0.003	26.190	0.000	0.062	0.072
Over 65 Years Population Percentage	0.018	0.004	4.829	0.000	0.011	0.026
WWR	-0.066	0.018	-3.701	0.000	-0.101	-0.031
Central Air	-0.023	0.002	-11.957	0.000	-0.027	-0.019
External Shading	0.011	0.004	3.153	0.002	0.004	0.018
Masonry Wall Material Proportion	0.015	0.002	9.332	0.000	0.012	0.019
Wooden Wall Material Proportion	0.021	0.002	13.415	0.000	0.018	0.024
Wood and Masonry Wall Material Proportion	0.013	0.002	6.014	0.000	0.009	0.017
Stucco Wall Material Proportion	-0.029	0.008	-3.493	0.000	-0.045	-0.013
Median Year Built	0.009	0.003	3.265	0.001	0.004	0.014
Land Cover	0.004	0.002	2.027	0.043	0.000	0.007
Other Roof Material Proportion	-0.068	0.025	-2.719	0.007	-0.117	-0.019
Slate Roof Material Proportion	0.048	0.025	1.92	0.055	-0.001	0.097

Table 5
The final model with the highest accuracy.

Prob > F = 0 F-statistic: 1355 No. Observations: 1133			Dep. Variable: AEB R-squared: 0.948 Adj. R-squared: 0.947 AIC: 2682 BIC: 2757			
Variables	Coefficient	Std. Error	t-value	P-value	CI Lower	CI Upper
Longitude	-0.455	0.149	-3.053	0.002	-0.748	-0.163
Latitude	-1.328	0.240	-5.521	0.000	-1.800	-0.856
Percentage of Poverty	0.067	0.003	26.422	0.000	0.062	0.072
Percentage of Over 65 Years	0.019	0.004	4.957	0.000	0.011	0.026
WWR	-0.060	0.018	-3.402	0.001	-0.095	-0.026
Central Air	-0.024	0.002	-12.716	0.000	-0.027	-0.020
External Shading	0.011	0.004	3.067	0.002	0.004	0.018
Masonry Wall Material Proportion	0.015	0.002	9.123	0.000	0.012	0.019
Wood Wall Material Proportion	0.020	0.002	12.379	0.000	0.017	0.024
Wood and Masonry Wall Material Proportion	0.012	0.002	5.707	0.000	0.008	0.017
Stucco Wall Material Proportion	-0.028	0.008	-3.427	0.001	-0.044	-0.012
Median Year Built	0.009	0.003	3.158	0.002	0.003	0.014
Land Cover	0.004	0.002	2.262	0.024	0.001	0.008
Other Roof Material Proportion	-0.062	0.025	-2.488	0.013	-0.111	-0.013
Slate Roof Material Proportion	0.053	0.025	2.131	0.033	0.004	0.102

household energy burdens, although they also exhibit higher variability.

Furthermore, among the wall materials, the stucco, a relatively rare wall type at the census tract scale, had a negative coefficient. The use of slate materials in the roofs seems to have an adverse effect on AEB. Among the passive design factors, WWR's impact on the AEB is another interesting finding of this study. The analysis showed that the impact of changing in WWR in the census tracts is almost equal to changing the poverty status in the region. That means increasing the average WWR in the census tracts scale by 1 %, on average, decreases the energy burden in the census tract by 0.06 %, and with 95 % confidence, its impact is between 0.026 to 0.095 in reducing the energy burden. Another noteworthy finding is that the age groups were another demographic variable that indicated an increase in the over 65-year-old population ratio is correlated with energy expenditure growth. This can potentially be explained by the older population's spending more time in the home and requiring a more controlled indoor thermal environment.

3.2.3. ML models comparison

After interpreting the results of the OLS multiple regression model as the most interpretable ML model, different ML models were investigated to find the best possible model for the AEB prediction based on the defined variables. Therefore, the SVR, XGBoost, Random Forest regression, and Decision Tree regression models were investigated. Moreover, a Neural Network model with three hidden layers each layer including 8 neurons and sigmoid activation function in 500 epochs was tested on the model. Table 6 demonstrates the results of these models.

As demonstrated in the Table 6, the developed multiple regression model has the highest R-squared value among all represented models. However, the MSE in some of the models, including the Neural Network model, is lower. MSE is defined as the summation of bias^2 and variance (Lin & Dobriban, 2020). The lower MSE value with lower R^2 in the neural network indicates that this model probably has a lower bias. Still, its difference from the OLS model is not significant, and it covers less variance as well. In other words, this model is a good fit for the ordinary predictions of the AEB; however, it might not perform as well as OLS in dealing with the outliers. Considering this fact and the power of interpretability in the OLS models, this model can still be announced as the most suitable model for AEB prediction. Moreover, all of the presented models, except the decision tree, generated lower errors compared to the best energy burden prediction models presented in the previous studies (Ghorbany et al., 2024b).

4. Contributions

These findings contribute to urban planning and energy policy-making, highlighting the superior role of passive design strategies over active design features in reducing energy burdens at the urban scale. This research provides valuable insights into energy justice, stressing the importance of passive design in addressing energy costs and sustainability in urban areas. It also emphasizes the need to consider geographical, socioeconomic, and demographic factors in urban planning.

By demonstrating that passive design features like window-to-wall ratio, external shade, and building materials have a significant effect on energy expenditures, the study advocates for a shift towards more low-cost, passive, sustainable, energy-efficient, and equitable urban design practices. The high R^2 value (94.8 %) indicates a strong model for

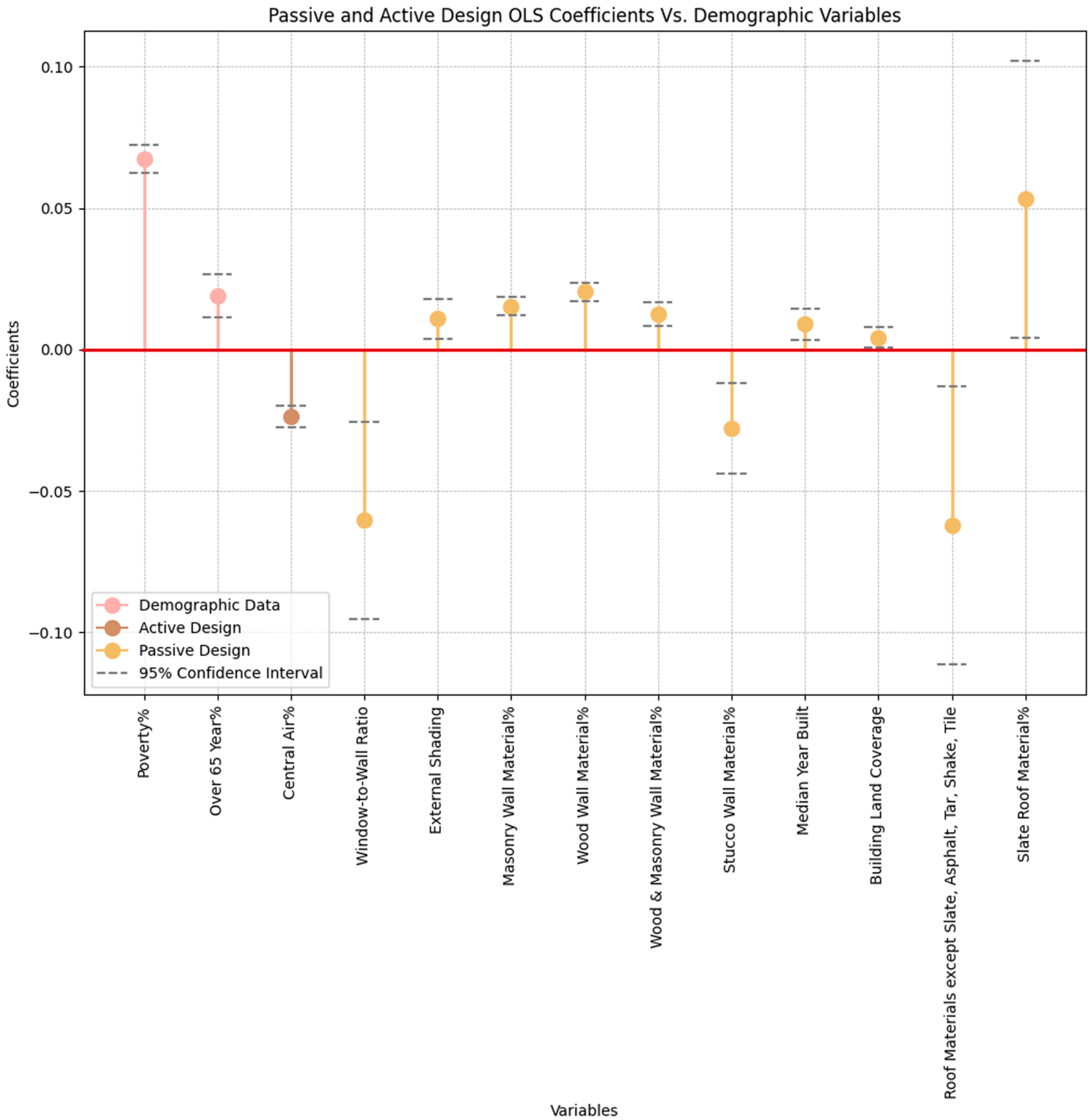

Fig. 4. Passive and active design comparison with demographic variables. Coefficients and confidence intervals.

Table 6
ML models performance measurement indicators.

Model	R-Squared	RMSE	MSE	MAE
Multiple Regression Model (OLS)	94.8 %	0.78	0.61	0.60
Neural Network Model	83.3 %	0.70	0.49	0.52
SVR Model	69.9 %	0.78	0.62	0.60
XGBoost Model	73.4 %	0.74	0.56	0.57
Random Forest Model	69.9 %	0.78	0.65	0.60
Decision Tree Model	53.2 %	0.97	0.99	0.74

predicting AEB, offering a clear, actionable framework for improving urban design and policy to reduce energy costs and enhance energy justice. This research not only deepens our understanding of the impact of design features on urban household energy burden but also provides concrete evidence for enhancing energy efficiency through informed passive design.

5. Conclusion

This research aimed to investigate the passive design and active design variables to (1) identify which variables from these categories are associated with AEB, (2) which of the categories has greater influence on the AEB in census tracts scale, and (3) develop a prediction model for AEB. To achieve these goals, this study started by merging four different datasets, including the Google Street View extracted information using Convolutional Neural Network, and preprocessing this dataset. Afterward, the statistical analysis was performed using a combination of ML methods, including multiple regression, decision tree regression, random forest regression, support vector regression (SVR), XGBoost regression, and neural networks to assess the variables' impact on AEB and provide a robust prediction model for this issue.

The results indicated that the passive design factors play a more crucial role in the energy burden of urban-scale households. From the mechanical devices in the building, only the central air conditioner was associated with the energy burden. Meanwhile, most of the passive design factors, including wall materials, roof materials, WWR, external shad, and land cover, were influencing the energy expenditures of the households. Among these variables, the WWR had one of the highest influences on energy expenditures, which was comparable to the population poverty ratio. Moreover, the geographical location of the census tracts and the elderly population of these regions were other factors that proved to be associated with the energy burden.

The developed model was able to predict the energy burden with 94.8 % accuracy, which is the best-developed model for energy burden prediction up to this day and to the authors' knowledge. Overall, the findings of this research emphasize the importance of passive design strategies on the urban scale and prove these variables to be more significant than the mechanical systems in the buildings. Moreover, it provided an easy-to-use and highly interpretable model to predict the energy burden with high accuracy. The findings of this research can be widely used by future researchers, urban policymakers, urban planners, and architects. Future studies are suggested to include the insulation data and investigate the results on a national scale. Moreover, future studies can consider adding the buildings' direction, especially in coastal cities, and examine the accuracy of the provided model for such conditions. Future studies are also encouraged to investigate and quantify the impact of passive design in census tracts with active solar energy techniques.

Funding

This research project was funded by the National Science Foundation (# 2317971) and the University of Notre Dame, Lucy Institute of Data and Society.

CRediT authorship contribution statement

Siavash Ghorbany: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ming Hu:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Matthew Sisk:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Conceptualization. **Siyuan Yao:** Writing – review & editing, Validation, Conceptualization. **Chaoli Wang:** Writing – review & editing, Validation, Supervision, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Data availability

Data will be made available on request.

Acknowledgment

We thank all faculty and staff, and students in the School of Architecture, College of Engineering, Department of Civil and Environmental Engineering and Earth Science, Lucy Institute of Data and Society at the University of Notre Dame for their support of this project.

The authors thank colleagues in our university research group, paper reviewers, and the journal editor for constructive feedback that has helped improve the article.

Appendixes

Appendix 1
The merged dataset statistical description.

	count	mean	std	min	25 %	50 %	75 %	max
FIPS	1135							
Building Square Feet	1133	2007.87	913.80	821.42	1482.54	1832.42	2313.40	19992.00
Land Square Feet	1135	35062.02	121213.62	2435.91	4046.53	5999.52	16409.78	2505312.31
Building Square Meters	1133	186.57	84.89	76.30	137.71	170.26	214.94	1857.21
Land Square Meters	1135	3256.91	11258.91	226.24	375.99	557.60	1524.44	232718.13
Central Air	1135	39.19	25.90	0.00	15.46	36.27	59.27	100.00
Land Cover	1133	30.48	21.71	0.07	12.68	27.99	45.67	100.00
Masonry Wall Material Proportion	1135	37.67	22.57	0.00	19.74	35.50	54.14	94.51
Wooden Wall Material Proportion	1135	25.76	19.62	0.00	9.85	22.18	38.46	96.82
Wood and Masonry Wall Material Proportion	1135	14.26	15.16	0.00	4.20	8.14	19.84	84.24
Stucco Wall Material Proportion	1135	1.14	3.23	0.00	0.00	0.23	0.71	36.34
Shingle/Asphalt Roof Material Proportion	1135	67.87	26.06	0.00	52.67	76.10	88.30	99.67
Tar and Gravel Roof Material Proportion	1135	9.34	11.31	0.00	1.11	4.05	14.57	59.48
Shake Roof Material Proportion	1135	0.40	1.76	0.00	0.00	0.00	0.13	19.37
Tile Roof Material Proportion	1135	0.45	1.23	0.00	0.00	0.17	0.47	27.28
Other Roof Material Proportion	1135	0.44	0.99	0.00	0.00	0.14	0.48	17.33
Slate Roof Material Proportion	1135	0.32	1.02	0.00	0.00	0.10	0.29	14.69
Warm Air Heating Proportion	1135	63.19	23.73	0.00	49.28	67.95	81.23	99.76
Hot Water Steam Heating Proportion	1135	15.64	13.80	0.00	4.61	12.14	23.37	70.81
None Heating System Proportion	1135	75.31	22.67	0.00	66.04	84.55	91.60	99.31
Unit Heater Heating System Proportion	1135	3.52	3.91	0.00	0.85	2.57	4.82	35.02
Masonry Building Structure Proportion	1135	59.18	19.70	2.97	44.41	60.28	75.01	96.94
Steel Building Structure Proportion	1135	4.06	3.77	0.00	1.84	3.13	5.16	36.10
Wooden Building Structure Proportion	1135	34.54	18.75	2.12	20.01	32.26	46.33	96.52
Manufactured Building Structure Proportion	1135	0.05	0.51	0.00	0.00	0.00	0.00	12.44
Concrete Building Structure Proportion	1135	2.17	2.32	0.00	0.79	1.38	2.74	23.97
Basement Building Foundation Proportion	1135	85.10	14.89	1.14	82.00	90.28	94.30	100.00
Slab Building Foundation Proportion	1135	9.56	10.34	0.00	4.04	6.48	11.21	97.31
Pile Building Foundation Proportion	1135	0.00	0.01	0.00	0.00	0.00	0.00	0.34
Pier Building Foundation Proportion	1135	0.00	0.02	0.00	0.00	0.00	0.00	0.66

(*continued on next page*)

Appendix 1 (*continued*)

	count	mean	std	min	25 %	50 %	75 %	max
Crawl Building Foundation Proportion	1135	5.34	8.65	0.00	0.68	1.95	5.80	75.89
Median Year Built	1135	1954.85	15.26	1939.00	1939.00	1953.00	1964.21	2005.97
avg_energyburden	1135	2.97	1.68	1.00	2.00	2.00	4.00	10.00
Pct_poverty	1135	15.98	12.41	0.39	6.45	12.05	22.71	72.55
Over 65 Years Population Percentage	1135	15.00	6.82	0.67	10.30	14.47	18.67	54.98
WWR Ratio	1135	15.15	1.57	10.88	14.06	14.98	16.01	25.68
External Shading	1135	30.27	8.18	6.79	24.82	30.31	35.89	62.24

Appendix 2

Data sources

The google street view data

The external shading and window-to-wall ratio (WWR) has been proven to be two crucial variables in passive design (Ghorbany et al., 2024b; Hu et al., 2023). That said, this data is not available through conventional datasets. Therefore, the first dataset used in this study is based on the authors' previous research, a Convolutional Neural Network (CNN) model that is able to extract the external shading and WWR data from the Google Street View images (Ghorbany et al., 2024b). The extracted data from this method is linked to the building's coordination, allowing to define the corresponding census tract as the mutual identification value to merge the datasets. The WWR determines the ratio of the windows area to the wall area and the external shading is a binary value determining what proportion of windows have external shading in the building.

National Structure Inventory (NSI) data

In addition to the ventilation characteristics of the buildings, other factors impact the passive design features of the built environment. Some of these variables include the buildings' structure type, building foundation type and its depth, the square footage of the building, the number of stories, and the building year built (NSI, 2024). The NSI is a dataset created and maintained by U. S. Army Corps of Engineers (USACE) to assist in the assessment of the effects of both natural and man-made disasters, make point-based building inventories with attribution consistently accessible across the country, and provide access all federal departments eager to work together on structural inventory data. The building structure type in this dataset includes masonry, wood, manufactured, and steel. The foundation type, on the other hand, is categorized in 7 groups, namely crawl, full basement, slab, pier, pile, fill, and solid wall.

Cook county open data

The envelope of the buildings including walls and roofs is another factor that influences the energy usage of the building for heating and cooling purposes (Manso et al., 2021; Staszczuk & Kuczyński, 2021). Moreover, the location of the buildings (longitude and latitude), the isolation of the building (land coverage), and the heating systems used in the building are claimed as the design factors that impact the buildings' energy usage (Bokaie et al., 2016; Mahmoud et al., 2021; Pineau et al., 2013; Pulselli et al., 2009). To access these data and assess their influence on the AEB, the open data from Cook County Government provided by the Cook County Assessor's Office were used (Cook County Assessor's Office, 2022). The variables "Longitude", "Latitude", "Wall Material", "Roof Material", "Building Square Feet", "Land Square Feet", "Central Heating", "Other Heating", and "Central Air" from this dataset were used. To add the land cover, the proportion of building square footage to the land square footage was calculated for the buildings. The wall material data included four types of external wall materials and roof types included 6 different types. Central heating includes four different types of heating including warm air, hot water steam, electric, and other and other heating systems including floor furnaces, unit heaters, stoves, solar, and none categories. The central air is a binary variable determining whether the building is equipped with a central air conditioning system or not.

LEAD and ACS data

As the target variable of this research, the AEB for each census tract was extracted and calculated from the LEAD data (Ma et al., 2019). This dataset includes the different energy costs for each census tract and the gross income information. Therefore, the AEB can be calculated from these data. Moreover, the percentage of people in poverty status and the elderly population ratio were used as the demographic and economic benchmarks extracted from ACS 5 year aggregated data, to assess the impact of building properties compared to these factors (U.S. Census Bureau, 2022). This dataset includes demographic, socioeconomic, and housing characteristics of the United States population (U.S. Census Bureau, 2022).

References

Abdallah, A. S. H. (2022). Passive design strategies to improve student thermal comfort in Assiut University: A field study in the faculty of physical education in hot season. In *Sustainable cities and society, 86*, Article 104110. https://doi.org/10.1016/j.scs.2022.104110. August.

Ajebe, M. (2024). Elusive trade-off: The solution to energy poverty and GHG emissions in Africa. *Environmental and Sustainability Indicators, 21*, Article 100320. https://doi.org/10.1016/j.indic.2023.100320. November 2023.

Alfraihat, R., Mulugeta, G., & Gala, T. S. (2016). Ecological evaluation of urban heat island in Chicago City, USA. *Journal of Atmospheric Pollution, 4*(1), 23–29. https://doi.org/10.12691/jap-4-1-3

Ashrafian, T., & Moazzen, N. (2019). The impact of glazing ratio and window configuration on occupants' comfort and energy demand: The case study of a school building in Eskisehir, Turkey. *Sustainable Cities and Society, 47*, Article 101483. https://doi.org/10.1016/j.scs.2019.101483. March.

Azimi Fereidani, N., Rodrigues, E., & Gaspar, A. R. (2021). A review of the energy implications of passive building design and active measures under climate change in the Middle East. *Journal of Cleaner Production, 305*. https://doi.org/10.1016/j.jclepro.2021.127152

Bektas Ekici, B., & Aksoy, U. T. (2011). Prediction of building energy needs in early stage of design by using ANFIS. *Expert Systems with Applications, 38*(5), 5352–5358. https://doi.org/10.1016/j.eswa.2010.10.021

Bhamare, D. K., Rathod, M. K., & Banerjee, J. (2019). Passive cooling techniques for building and their applicability in different climatic zones—The state of art. *Energy and Buildings, 198*, 467–490. https://doi.org/10.1016/j.enbuild.2019.06.023

Bokaie, M., Zarkesh, M. K., Arasteh, P. D., & Hosseini, A. (2016). Assessment of urban heat island based on the relationship between land surface temperature and Land

Use/Land Cover in Tehran. *Sustainable Cities and Society, 23*, 94–104. https://doi.org/10.1016/j.scs.2016.03.009

Bouketta, S. (2023). Urban Cool Island as a sustainable passive cooling strategy of urban spaces under summer conditions in Mediterranean climate. In *Sustainable cities and society, 99*, Article 104956. https://doi.org/10.1016/j.scs.2023.104956. September.

Bukkapatnam, S. T. S., Afrin, K., Dave, D., & Kumara, S. R. T. (2019). Machine learning and AI for long-term fault prognosis in complex manufacturing systems. *CIRP Annals, 68*(1), 459–462. https://doi.org/10.1016/j.cirp.2019.04.104

Burton, A. L. (2021). OLS (Linear) regression. *The encyclopedia of research methods in criminology and criminal justice* (pp. 509–514). Wiley. https://doi.org/10.1002/9781119111931.ch104

Campaniço, H., Hollmuller, P., & Soares, P. M. M. (2014). Assessing energy savings in cooling demand of buildings using passive cooling systems based on ventilation. *Applied Energy, 134*, 426–438. https://doi.org/10.1016/j.apenergy.2014.08.053

M. R. B. T.-P. of Chakrabarti, A., Ghosh, J. K., Bandyopadhyay, P. S., & Forster, S. (2011). AIC, BIC and recent advances in model selection. In *Philosophy of statistics, 7* pp. 583–605). Elsevier. https://doi.org/10.1016/B978-0-444-51862-0.50018-6.

Chen, X., & Yang, H. (2017). Sensitivity analysis and optimization of a typical passively designed residential building with hybrid ventilation in hot and humid climates. *Energy Procedia, 142*, 1781–1786. https://doi.org/10.1016/j.egypro.2017.12.563

Cheng, Q., & Ngok, K. (2020). Welfare attitudes towards anti-poverty policies in China: Economical individualism, social collectivism and institutional differences. *Social Indicators Research, 150*(2), 679–694. https://doi.org/10.1007/s11205-020-02313-y

Cheshmehzangi, A., & Dawodu, A. (2020). Passive cooling energy systems: Holistic SWOT analyses for achieving urban sustainability. *International Journal of Sustainable Energy, 39*(9), 822–842. https://doi.org/10.1080/14786451.2020.1763348

Cheung, C. K., Fuller, R. J., & Luther, M. B. (2005). Energy-efficient envelope design for high-rise apartments. *Energy and Buildings, 37*(1), 37–48. https://doi.org/10.1016/j.enbuild.2004.05.002

Cook County Assessor's Office. (2022). Assessor residential property characteristics | Cook county open data. *Cook County Assessor's Office Data Department*. https://datacatalog.cookcountyil.gov/Property-Taxation/Assessor-Archived-05-11-2022-Residential-Property-/bcnq-qi2z/about_data.

Department of Energy. (2023). Low-income community energy solutions. *Energy.Gov*. https://www.energy.gov/scep/slsc/low-income-community-energy-solutions.

Diz-Mellado, E., López-Cabeza, V. P., Roa-Fernández, J., Rivera-Gómez, C., & Galán-Marín, C. (2023). Energy-saving and thermal comfort potential of vernacular urban block porosity shading. *Sustainable Cities and Society, 89*, Article 104325. https://doi.org/10.1016/j.scs.2022.104325. July 2022.

Drehobl, A., Ross, L., & Ayala, R. (2020). How high are household energy burdens. *An Assessment of National and Metropolitan Energy Burdens across the US*.

Elrefai, R., & Nikolopoulou, M. (2023). A simplified outdoor shading assessment method (OSAM) to identify outdoor shading requirements over the year within an urban context. *Sustainable Cities and Society, 97*, Article 104773. https://doi.org/10.1016/j.scs.2023.104773. March.

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

Ghorbany, S., Hu, M., Yao, S., & Wang, C. (2024a). Towards a sustainable urban future: A comprehensive review of urban heat island research technologies and machine learning approaches. *Sustainability, 16*(11), 4609. https://doi.org/10.3390/su16114609

Ghorbany, S., Hu, M., Yao, S., Wang, C., Nguyen, Q. C., Yue, X., Alirezaei, M., Tasdizen, T., & Sisk, M. (2024b). Examining the role of passive design indicators in energy burden reduction: Insights from a machine learning and deep learning approach. *Building and Environment, 250*, Article 111126. https://doi.org/10.1016/j.buildenv.2023.111126 (December 2023).

Ghorbany, S., Noorzai, E., & Yousefi, S. (2023). BIM-based solution to enhance the performance of public-private partnership construction projects using copula bayesian network. *Expert Systems with Applications, 216*, Article 119501. https://doi.org/10.1016/j.eswa.2023.119501. May 2022.

Ghorbany, S., Yousefi, S., & Noorzai, E. (2022). Evaluating and optimizing performance of public–private partnership projects using copula Bayesian network. *Engineering, Construction and Architectural Management*. https://doi.org/10.1108/ECAM-05-2022-0492

Guo, Y., Gasparrini, A., Li, S., Sera, F., Vicedo-Cabrera, A. M., de Sousa Zanotti Stagliorio Coelho, M., Saldiva, P. H. N., Lavigne, E., Tawatsupa, B., & Punnasiri, K. (2018). Quantifying excess deaths related to heatwaves under climate change scenarios: A multicountry time series modelling study. *PLoS Medicine, 15*(7), Article e1002629.

Han, R., Xu, Z., & Qing, Y. (2017). Study of passive evaporative cooling technique on water-retaining roof brick. *Procedia Engineering, 180*, 986–992. https://doi.org/10.1016/j.proeng.2017.04.258

Harmati, N., & Magyar, Z. (2015). Influence of WWR, WG and glazing properties on the annual heating and cooling energy demand in buildings. *Energy Procedia, 78*, 2458–2463. https://doi.org/10.1016/j.egypro.2015.11.229

He, X., Li, J., Zhou, Q., Lu, G., & Xu, H. (2023). Robust key parameter identification of dedicated hybrid engine performance indicators via K-fold filter collaborated feature selection. *Engineering Applications of Artificial Intelligence, 126*, Article 107114. https://doi.org/10.1016/j.engappai.2023.107114. November 2022.

Hu, M., Zhang, K., Nguyen, Q., & Tasdizen, T. (2023). The effects of passive design on indoor thermal comfort and energy savings for residential buildings in hot climates: A systematic review. *Urban Climate, 49*, Article 101466. https://doi.org/10.1016/j.uclim.2023.101466

Javad, K., & Navid, G. (2019). Thermal comfort investigation of stratified indoor environment in displacement ventilation: Climate-adaptive building with smart windows. *Sustainable Cities and Society, 46*, Article 101354. https://doi.org/10.1016/j.scs.2018.11.029. November 2018.

Kang, J., Ahn, K., Park, C., & Schuetze, T. (2015). Assessment of passive vs. active strategies for a school building design. *Sustainability, 7*(11), 15136–15151. https://doi.org/10.3390/su71115136

Kwok, S. S. K., & Lee, E. W. M. (2011). A study of the importance of occupancy to building cooling load in prediction by intelligent approach. *Energy Conversion and Management, 52*(7), 2555–2564. https://doi.org/10.1016/j.enconman.2011.02.002

Lee, K. K. G., Kasim, H., Zhou, W. J., Sirigina, R. P., & Hung, G. G. T. (2023). Feature redundancy assessment framework for subject matter experts. *Engineering Applications of Artificial Intelligence, 117*, Article 105456. https://doi.org/10.1016/j.engappai.2022.105456. February 2022.

Li, Y., Tong, Z., Tong, S., & Westerdahl, D. (2022). A data-driven interval forecasting model for building energy prediction using attention-based LSTM and fuzzy information granulation. *Sustainable Cities and Society, 76*, Article 103481. https://doi.org/10.1016/j.scs.2021.103481. August 2021.

Lin, L., & Dobriban, E. (2020). What causes the test error? Going beyond bias-variance via ANOVA. *Journal of Machine Learning Research, 22*, 1–82. http://arxiv.org/abs/2010.05170.

Ling, W., & Jin, H. (2018). Coupling of active technologies and passive technologies: Low energy design strategies for northern public buildings with different space characteristics in China. *IOP Conference Series: Materials Science and Engineering, 423* (1), Article 012186. https://doi.org/10.1088/1757-899X/423/1/012186

Ma, O., Laymon, K., Day, M. H., Bracho, R., Weers, J. D., & Vimont, A. J. (2019). *Low-Income Energy Affordability Data (LEAD) tool methodology* (NREL/TP-6A20-74249, 1545589).

Mahmoud, M., Ramadan, M., Naher, S., Pullen, K., & Olabi, A.-G. (2021). The impacts of different heating systems on the environment: A review. *Science of The Total Environment, 766*, Article 142625. https://doi.org/10.1016/j.scitotenv.2020.142625

Manso, M., Teotónio, I., Silva, C. M., & Cruz, C. O. (2021). Green roof and green wall benefits and costs: A review of the quantitative evidence. *Renewable and Sustainable Energy Reviews, 135*, Article 110111. https://doi.org/10.1016/j.rser.2020.110111. August 2020.

NSI. (2024, February 6). *NSI technical references*. https://www.hec.usace.army.mil/confluence/nsi/technicalreferences/latest.

Nti, I. K., Nyarko-Boateng, O., & Aning, J. (2021). Performance of machine learning algorithms with different K values in K-fold cross validation. *International Journal of Information Technology and Computer Science, 13*(6), 61–71. https://doi.org/10.5815/ijitcs.2021.06.05

Ochoa, C. E., & Capeluto, I. G. (2008). Strategic decision-making for intelligent buildings: Comparative impact of passive design strategies and active features in a hot climate. *Building and Environment, 43*(11), 1829–1839. https://doi.org/10.1016/j.buildenv.2007.10.018

Onyenokporo, N. C., & Ochedi, E. T. (2019). Low-cost retrofit packages for residential buildings in hot-humid Lagos, Nigeria. *International Journal of Building Pathology and Adaptation, 37*(3), 250–272. https://doi.org/10.1108/IJBPA-01-2018-0010

Pham, H. (2019). A new criterion for model selection. *Mathematics, 7*(12), 1215. https://doi.org/10.3390/math7121215

Pineau, D., Rivière, P., Stabat, P., Hoang, P., & Archambault, V. (2013). Performance analysis of heating systems for low energy houses. *Energy and Buildings, 65*, 45–54. https://doi.org/10.1016/j.enbuild.2013.05.036

Pulselli, R. M., Simoncini, E., & Marchettini, N. (2009). Energy and emergy based cost–benefit evaluation of building envelopes relative to geographical location and climate. *Building and Environment, 44*(5), 920–928. https://doi.org/10.1016/j.buildenv.2008.06.009

Rebala, G., Ravi, A., & Churiwala, S. (2019). Machine learning definition and basics BT - an introduction to machine learning. *An introduction to machine learning*. https://doi.org/10.1007/978-3-030-15729-6_1

Sachs, I. (2015). Entering the anthropocene: The twofold challenge of climate change and poverty eradication. In F. Mancebo, & I. Sachs (Eds.), *Transitions to sustainability* (pp. 7–18). Netherlands: Springer. https://doi.org/10.1007/978-94-017-9532-6_2.

Saidi, R., Bouaguel, W., Essoussi, N., & Hassanien, A. E. (2019). *Hybrid feature selection method based on the genetic algorithm and pearson correlation coefficient BT - machine learning paradigms: theory and application* (pp. 3–24). Springer International Publishing. https://doi.org/10.1007/978-3-030-02357-7_1

Salameh, M., & Touqan, B. (2022). Traditional passive design solutions as a key factor for sustainable modern urban designs in the hot, arid climate of the United Arab Emirates. *Buildings, 12*(11), 1811. https://doi.org/10.3390/buildings12111811

Shaeri, J., Aflaki, A., Yaghoubi, M., & Janalizadeh, H. (2018). Investigation of passive design strategies in a traditional urban neighborhood: A case study. *Urban Climate, 26*, 31–50. https://doi.org/10.1016/j.uclim.2018.08.003. January.

Shahee, A., Abdoos, M., Aslani, A., & Zahedi, R. (2024). Reducing the energy consumption of buildings by implementing insulation scenarios and using renewable energies. *Energy Informatics, 7*(1), 18. https://doi.org/10.1186/s42162-024-00311-9

Sholahudin, S., & Han, H. (2016). Simplified dynamic neural network model to predict heating load of a building using Taguchi method. *Energy, 115*, 1672–1678. https://doi.org/10.1016/j.energy.2016.03.057

Šileika, A., & Bekerytė, J. (2013). The theoretical issues of unemployment, poverty and crime coherence in the terms of sustainable development. *Journal of Security and Sustainability Issues, 2*(3), 59–70. https://doi.org/10.9770/jssi.2013.2.3(5

Staszczuk, A., & Kuczyński, T. (2021). The impact of wall and roof material on the summer thermal performance of building in a temperate climate. *Energy, 228*, Article 120482. https://doi.org/10.1016/j.energy.2021.120482

Su, B. (2011). The impact of passive design factors on house energy efficiency. *Architectural Science Review, 54*(4), 270–276. https://doi.org/10.1080/00038628.2011.613638

Su, Y., Song, J., Lu, Y., Fan, D., & Yang, M. (2023). Economic poverty, common prosperity, and underdog entrepreneurship. *Journal of Business Research, 165*, Article 114061. https://doi.org/10.1016/j.jbusres.2023.114061

Sun, X., Gou, Z., & Lau, S. S.-Y. (2018). Cost-effectiveness of active and passive design strategies for existing building retrofits in tropical climate: Case study of a zero energy building. *Journal of Cleaner Production, 183*, 35–45. https://doi.org/10.1016/j.jclepro.2018.02.137

Sureiman, O., & Mangera, C. M. (2020). F-test of overall significance in regression analysis simplified. *Journal of the Practice of Cardiovascular Sciences, 6*(2). https://journals.lww.com/jpcs/fulltext/2020/06020/f_test_of_overall_significance_in_regression.6.aspx.

U.S. Census Bureau. (2022, February 8). *American community survey 5-year data (2009-2022). Census.Gov*. https://www.census.gov/data/developers/data-sets/acs-5year.html.

United States Census Bureau. (2024). Census.gov. *In Census.gov*. https://www.census.gov/en.html.

Vargas, A. P., & Hamui, L. (2021). Thermal energy performance simulation of a residential building retrofitted with passive design strategies: A case study in Mexico. *Sustainability, 13*(14), 8064. https://doi.org/10.3390/su13148064

Wang, X., Mai, X., Lei, B., Bi, H., Zhao, B., & Mao, G. (2020). Collaborative optimization between passive design measures and active heating systems for building heating in Qinghai-Tibet plateau of China. *Renewable Energy, 147*, 683–694. https://doi.org/10.1016/j.renene.2019.09.031

Wigginton, M., & Harris, J. (2013). *Intelligent skins*. Routledge. https://doi.org/10.4324/9780080495446

Yao, R., Costanzo, V., Li, X., Zhang, Q., & Li, B. (2018). The effect of passive measures on thermal comfort and energy conservation. A case study of the hot summer and cold winter climate in the Yangtze River region. *Journal of Building Engineering, 15*, 298–310. https://doi.org/10.1016/j.jobe.2017.11.012

Zune, M., Rodrigues, L., & Gillott, M. (2020). Vernacular passive design in Myanmar housing for thermal comfort. *Sustainable Cities and Society, 54*, Article 101992. https://doi.org/10.1016/j.scs.2019.101992