# Creating and Analyzing a Multimedia Dataset for Building Energy Efficiency Estimation *

Gong Qi Chen[1] gchen1@gradcenter.cuny.edu, Ninad Kashyap[2] nkashya001@citymail.cuny.edu, Zihao Zhang[3] zzhang@ccny.cuny.edu, Yana Kucheva[4] ykucheva@ccny.cuny.edu, Michael Bobker[5] mbobker@ccny.cuny.edu, and Zhigang Zhu[1,6] zzhu@ccny.cuny.edu

[1] Department of Computer Science, The CUNY Graduate Center, 365 5th Ave, New York, NY 10016
[2] Sustainability in the Urban Environment;
[3] Spitzer School of Architecture;
[4] Department of Sociology;
[5] Building Performance Laboratory; and
[6] Department of Computer Science, The CUNY City College, 160 Convent Avenue, New York, NY 10031

**Abstract.** This paper presents the results of a research that created and analyzed a Multimedia dataset for building energy efficiency estimation. First a new Multimedia Building Energy Efficiency (MMBEE) dataset was created from publicly available data. This work then explored the use of the window-to-wall ratio (WWR) information from building facade images and integrated it with traditional tabular data to create new training data, in order to predict building energy efficiency measures. Finally, we discuss potential applications and future research directions in using the MMBEE dataset for building energy efficiency prediction. Throughout the paper, a number of important processes and analyses were performed, which include feature selection, data correlation analysis, WWR extraction, and comparison of deep network and random forest models in building energy efficiency estimation. From this first attempt at using the Multimedia dataset for building energy efficiency estimation, we found the performances of deep models were better than traditional models such as random forest. We also found that there was an optimal point of what features shall be used for the prediction. Nonetheless, the incorporation of the current WWR estimation results did not yield the anticipated enhancement in estimation performance. Subsequently, a comprehensive investigation was conducted to ascertain potential contributing factors, and several avenues for future research were identified to enhance the predictive utility of the WWR feature.

* The work is supported by the College-wide Research Vision (CRV) Initiative from the CCNY Provost's Office. The work is also supported by NSF through Awards #2131186 (CISE-MSI), #1827505 (PFI), and #1737533 (S&CC), the US Air Force Office of Scientific Research (AFOSR) via Award #FA9550-21-1-0082, and the ODNI Intelligence Community Center for Academic Excellence (IC CAE) at Rutgers University (#HHM402-19-1-0003 and #HHM402-18-1-0007).

**Keywords:** Building energy efficiency · window-to-wall ratio · computer vision · machine learning, · Multimedia databases

## 1 Introduction

In New York City, buildings cause nearly 70% of carbon emissions because of fossil fuel usage for heating, cooling, and powering purposes [4]. More than one million buildings in NYC consume approximately 85% of the city's electricity. A considerable portion of these structures were erected prior to the establishment of contemporary energy efficiency regulations. Consequently, they frequently demand a higher energy input for heating, cooling, and powering when compared to more recently constructed edifices. To achieve a carbon-neutral NYC by 2050, the New York City Council enacted Local Law 97 in 2019, which covers most buildings over 25,000 square feet [7]. The legislation mandates building proprietors to undertake building renovations aimed at capping carbon emissions, thereby resulting in the dual outcomes of bolstering energy efficiency and constraining energy utilization. In addition, buildings are graded based on ENERGY STAR efficiency scores that evaluate a building's energy efficiency by comparing its performance to similar building types in comparable environments [5].

In our previous work[8], NYC buildings' historical energy consumption data was used in machine learning models to determine their ENERGY STAR Scores for time series analysis and future prediction. In this work, we define this historical energy consumption data including numerical and text data as traditional tabular data. The building's administrative matrices serve as a prevalent traditional data source for estimating building energy efficiency performance. These matrices provide direct information that differentiates each building, such as water usage, electric consumption and gas usage.

Recent years have seen a surge of interest in utilizing non-traditional building information for estimating energy efficiency thanks to the success of computer vision approaches with deep learning models. One of the focus areas is building facade images. Research has shown that building facade images contain valuable information regarding building energy efficiency. Factors such as structure, facade material and Window-to-Wall Ratio (WWR) exert a significant influence on heat reflection and absorption. Also, the orientation and the location of the building impact the optimal WWR [12].

This paper proposes a method that utilizes machine learning to estimate the buildings' energy efficiency grades using traditional energy consumption data and facade images obtained from the Google Street View API. The contributions of this project include the following:

- Creation of a new Multimedia Building Energy Efficiency (MMBEE) dataset from publicly available data, and analysis of the importance of each of the Multimedia features.
- Extraction of a building's window-to-wall ratio (WWR) data from facade images and integration with traditional data to create new training data.

- Utilization of neural network models to predict energy efficiency grades, examining the WWR's impact on determining building grades.
- Discussions of potential issues and future research directions in using the MMBEE dataset for building energy efficiency prediction.

Throughout the paper, a number of important processes and analyses were performed including feature selection, data correlation, WWR extraction, and deep network and random forest models in building energy efficiency estimation. From our first attempt at using the dataset for building energy efficiency estimation, we found the performances of deep models were better than traditional models such as random forest. We also discovered that there exists an ideal threshold for selecting the features to use in our predictions. At this juncture, including the current WWR results would not enhance the performance of our estimation as we have hypothesized. Nevertheless, the dataset would be valuable for research and applications, and the analyses offer some insights into the data and the models. In particular, to make WWR a good predictor for estimating building energy efficiency, we propose the following considerations: enhancing data quality, encompassing actions like obstacle removal (such as trees, vehicles and pedestrians), performing image rectification, collecting more relevant image data, and including supplementary data to building images. These factors warrant further investigation in our future research endeavors.

## 2 Related Work

The escalating climate concerns have led to a surge in research on predicting and improving building energy efficiency using techniques that integrate neural networks. Mena et al. (2014) deployed artificial neural networks (ANN) to estimate electricity consumption of Spain's Solar Energy Research Center (CIESOL) bioclimatic buildings and revealed a positive correlation between electricity usage and outdoor temperature and sunlight [9]. Yalcintas et al. (2007) [16] employed ANN in conjunction with the Commercial Buildings Energy Consumption Survey to estimate electricity consumption per square meter (EUI) and discovered that grouping buildings based on property types yielded more accurate predictions.

Besides studies using traditional numerical and text data, the increasing availability of image data and advancements in computational power have allowed researchers to use images as another vital data source for estimating energy efficiency. Visual attributes like window-to-wall ratio (WWR) can be extracted using computer vision techniques as a crucial factor in estimating building energy consumption. Li et al. (2020) [3] introduced an approach for window detection in facade images, diverging from traditional computer vision methods of predicting bounding boxes or segmenting the facade. While the method has the capability to identify windows efficiently and precisely, its effectiveness heavily relies on accurate keypoint detection, which is usually a hard problem under various conditions, such as illumination changes and occlusions.

Szcześniak et al. (2021) [11] introduced an automated methodology that utilizes computer vision techniques to extract WWRs for buildings on a large urban scale. This methodology can be applied universally in cities worldwide, given the availability of geotagged street view imagery such as Google Street View (GSV) or similar datasets. The approach still employs a traditional method with a Sobel filter to detect both horizontal and vertical edge lines, which is hard to generalize due to the need for experimentally setting thresholds for the detection. Sun et al. (2022) [10] conducted research in Glasgow to estimate building energy efficiency using administrative and emerging urban big data through deep learning techniques. Their study demonstrated that building facade images provide valuable information for estimating energy efficiency. This research showcased the potential of Multimedia models to outperform single-source models. However, there was no such study being done for New York City buildings.

Phillip et al. (2023) were working on predicting future ENERGY STAR Scores using current year building data in NYC [8]. They extracted 12 important features from The NYC Mayor's Office of Climate and Environmental Justice, which collects data annually through the EPA ENERGY STAR Portfolio Manager[6]. This data collection includes over 29,000 building metrics related to water and energy consumption. However, only traditional data were utilized for this study. Through the integration of image data into the model in this paper, such as the utilization of the (WWR) information, it is posited that the model's predictive capabilities in assessing building energy efficiency are likely to experience enhancements. Should these enhancements not materialize as anticipated, a rigorous examination of the primary contributing factors will be conducted, with a view to devising strategies for optimizing the efficacy of WWR as a predictive variable.

## 3 Creation of the MMBEE Dataset

In this study, a Multimedia Building Energy Efficiency (MMBEE) dataset is constructed by combining Multimedia data from the NYC Energy and Water Data Disclosure and Google Street View (GSV) images for estimating building energy efficiency.

### 3.1 Data from NYC Energy and Water Data Disclosure

The NYC Mayor's Office of Climate and Environmental Justice annually collects more than 29,000 buildings' water and energy consumption data through the EPA ENERGY STAR Portfolio Manager. This data encompasses privately owned buildings exceeding 25,000 sq ft and City-owned buildings exceeding 10,000 sq ft [2]. It enables building owners to assess and compare their energy and water consumption with similar building types in comparable settings, improving the assessment of efficiency and sustainability. The dataset includes water electric, and gas consumption, greenhouse gas (GHG) emissions, and other essential building characteristics such as locations and primary usage types. Each building has an ENERGY STAR Score (A, B, C, or D), which indicates a building's energy efficiency. The score is computed by comparing the building's energy usage to other similar buildings within the same category and comparable settings, such as the physical attributes, its operations and how people use it[2].

**Table 1.** Dataset from NYC energy and water data disclosure per usage types.

| Usage Types | Total | Rank A | Rank B | Rank C | Rank D |
|---|---|---|---|---|---|
| Office | 1,000 | 185 | 287 | 208 | 320 |
| Multifamily Housing | 1,000 | 181 | 148 | 159 | 512 |
| K-12 School | 1,000 | 206 | 267 | 188 | 339 |

**Table 2.** Dataset from NYC energy and water data disclosure per borough.

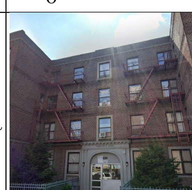
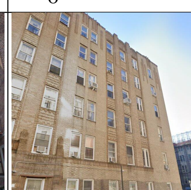
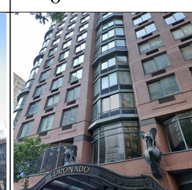
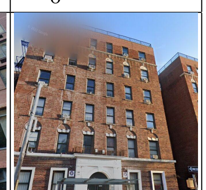
| Borough | Total | Rank A | Rank B | Rank C | Rank D |
|---|---|---|---|---|---|
| Bronx | 476 | 91 | 91 | 66 | 228 |
| Brooklyn | 715 | 141 | 167 | 134 | 273 |
| Manhattan | 1,264 | 243 | 312 | 242 | 467 |
| Queens | 464 | 74 | 110 | 99 | 181 |
| Staten Island | 81 | 23 | 22 | 14 | 22 |

To generate a balanced dataset for machine learning experiments, we randomly sampled 1,000 buildings from the pool of each of three major building usage types: multifamily housing, office, and K-12 school. We chose these three specific building types because of the availability of these data within the NYC Energy and Water Data Disclosure, which allows a more robust comparative analysis compared to the other building types. These three building types collectively represent more than 81% of the primary building types found in the Disclosure obtained from the NYC Mayor's Office of Climate and Sustainability. The ENERGY STAR rank distribution based on the three building usage types is presented in Table 1 and the distribution based on the five boroughs in NYC is presented in Table 2.

The original dataset comprised over 250 features for both numerical and categorical data. A surplus of features can lead to model overfitting. In addition, we aimed to narrow down the number of features to focus on the building attributes that are most relevant to energy efficiency. Therefore, we undertook a feature selection procedure guided by the analysis of important features in Phillip et al. study(2023) [8]. The 12 attributes include the following (the digit at the end of each attribute type shows the number of measures): electricity use (3), gross floor area (1), latitude (1), longitude(1), year built (1), occupancy (2), natural gas use (kBtu) (1), eGRID output emissions rate (kgCO2e/MBtu) (1), and green power - offsite (kWh) (1). A full list of the 12 attributes is shown in Table 4 of Section 4.1, where we will explain a correlation analysis for the importance of the 12 features to see if the selection is appropriate.

In addition to the 12 numerical features, our dataset also included 2 categorical features, the primary usage types and the boroughs, which we utilized in our

**Table 3.** Examples of Multimedia features of **multifamily housing**. (EU, WNEU, EURS, GFA, GFAP, LA, LO, OCC, YB, NGU, EGR, GPO) representing *electricity use, weather normalized site electricity (kWh), electricity Use – generated from on-site renewable Systems (kWh), gross floor area, gross floor area (Parking), latitude, longitude, year built, occupancy, natural gas use (kBtu), eGRID output emissions rate (kgCO2e/MBtu), and green power - Offsite (kWh) electricity use, gross floor area, latitude, longitude, year built, occupancy, natural gas use (kBtu), eGRID output emissions rate (kgCO2e/MBtu), and green power - Offsite (kWh).*

| Energy Score | A | B | C | D |
|---|---|---|---|---|
| EU: | 58887.3 | 516312 | 1391565.5 | 271552.6 |
| WNEU: | 58887.3 | 502495.5 | 1368769.6 | 266471.2 |
| EURS: | 0 | 0 | 0 | 0 |
| GFA: | 38800 | 147660 | 204762 | 78167 |
| GFAP: | 0 | 0 | 11110 | 0 |
| LA: | 40.639208 | 40.840864 | 40.776824 | 40.84902 |
| LO: | -73.956953 | -73.911364 | -73.981431 | -73.917491 |
| YB: | 1935 | 1941 | 1990 | 1939 |
| OCC: | 100 | 100 | 100 | 100 |
| NGU: | 2218977.6 | 8210472.6 | 8523620.1 | 6526143.7 |
| EGR: | 84.7 | 84.7 | 84.7 | 84.7 |
| GPO: | 0 | 0 | 0 | 0 |
| Facade & WWR |  WWR = 0.1278 |  WWR = 0.2413 |  WWR = 0.07603 |  WWR = 0.1526 |

prediction. However, neural network models require numerical input, therefore it was necessary to apply the one-hot encoding technique to convert these categorical features into a format suitable for our model. One-hot encoding transforms each category into binary vectors and creates new binary columns in the dataset. In these new columns, the corresponding column value will be 1 if the building belongs to that category and 0 otherwise.

### 3.2 Data from Google Street View (GSV) Images

Street view imagery offers rich visual data that is more intuitive and human-oriented compared to other forms of data. We acquire images from the Google Street View (GSV) service utilizing its dedicated Application Programming Interface (API). We requested building images from GSV using the address information from the 3,000 buildings sampled from the NYC Energy and Water Data Disclosure database. We successfully acquired 2,972 images out of those 3,000 buildings; the 28 images were not acquired probably because of the typos or differences in the names of the addresses.

Table 3 lists examples of Multimedia features for one of the three building usage types: multifamily housing. The data show a glimpse of the 12 attributes from traditional data, and the corresponding building facade images and their WWRs. Note that some of the WWR measures are more accurate than others. This can also serve as a Multimedia interface for understanding the building energy efficiency measures. This work is part of a larger effort to establish a web-based platform with consolidated data, visualizations and simulations for ongoing public dialogue, program evaluation, and decision-making to facilitate collaborations among academic, community, and policymakers [1].

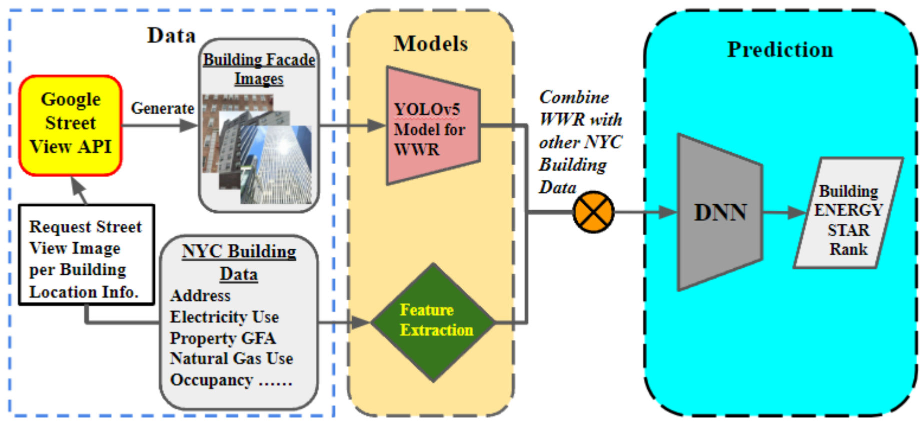



**Fig. 1.** Multimedia deep learning model overview

## 4 Analysis of the MMBEE Dataset

The analysis of the MMBEE dataset contains two stages (Figure 1). Firstly, a feature section module and a YOLOv5 detection model [15] are employed in a preprocessing step to extract important features from traditional data and obtain WWR information from building facade images, respectively. Secondly, a Multimedia deep learning model will be constructed by integrating a deep neural network (DNN) model to predict energy efficiency grades.

### 4.1 Feature Analysis from Traditional Data

Per the analysis of important features in Phillip et al. (2023) [8], we have chosen 12 numerical attributes from the NYC Energy and Water Data Disclosure. The attribute collection includes electricity use, gross floor area, latitude, longitude, year built, occupancy, natural gas use (kBtu), eGRID output emissions rate (kgCO2e/MBtu), and green power - Offsite (kWh). We computed the correlation between all 12 features and the ENERGY STAR Score, and the magnitude of this correlation signifies the strength of the relationship between each feature and the ENERGY STAR Score (Table 4). A positive correlation indicates that the feature

**Table 4.** Selected features arranged by the absolute correlation to ENERGY STAR Score.

| Features | Correlation to ENERGY STAR Score |
|---|---|
| Natural Gas Use (kBtu) | -0.1261 |
| Electricity Use - Grid Purchase (kWh) | -0.0529 |
| Weather Normalized Site Electricity (kWh) | -0.0526 |
| Property GFA - Calculated (Buildings) ($ft^2$) | 0.0502 |
| Occupancy | -0.0406 |
| Latitude | -0.0327 |
| eGRID Output Emissions Rate (kgCO2e/MBtu) | -0.0203 |
| Longitude | -0.0192 |
| Electricity Use – Generated from Onsite Renewable Systems (kWh) | 0.0186 |
| Year Built | -0.0079 |
| Property GFA - Calculated (Parking) ($ft^2$) | -0.0047 |
| Green Power - Offsite (kWh) | 0.0041 |

positively influences the ENERGY STAR Score, while a negative correlation implies a negative impact on the ENERGY STAR Score. Notably, natural gas usage and electricity consumption exhibit the highest absolute correlation values. This aligns with common intuition, as increased energy consumption typically corresponds to reduced energy efficiency.

However, some of the features do not have a strong correlation with the ENERGY STAR Score, for example, Year Built, Property GFA - Calculated (Parking) ($ft^2$), and Green Power - Offsite (kWh) have a minimum correlation with the ENERGY STAR Score (Table 4). One of the reasons for the low correlation to Year Built data is that the data may not be able to accurately reflect the energy use, since the date of construction does not indicate what significant renovations may have been done (i.e., new windows, new boiler, even gut renovation with insulation). The green power-offsite probably applies to very few buildings and would be a contract/purchase decision by the owner and not related to the energy use.

### 4.2 WWR Estimation Based on a YOLOv5 Model

The building facade image data is processed separately by a YOLOv5, due to its straightforward architecture with a balance between efficiency and accuracy. The YOLOv5 detects the position of building windows and facades, which are then used to calculate the window-to-wall ratio for each building. The calculated ratio becomes an additional numerical feature for the building in the DNN model. Regrettably, there is no existing pre-trained model specifically designed for detecting facades and windows. As a result, we trained the model using the facade database from the Center for Machine Perception (CMP) [13], which comprises 605 building images annotated with 11 label classes (facade, molding, cornice, pillar, window, door, sill, blind, balcony, shop, deco) [14].

Furthermore, we conducted another training by exclusively using facade and window labels of the same dataset, in order to assess whether performance improves when distractions from other labels are minimized. The purpose of this model is to detect windows on the facade image. Once the model is trained, it predicts the coordinates of the identified windows and facades. These coordinates serve as the basis for performing image segmentation on the windows and facades, effectively isolating them from the rest of the image. By quantifying the number of pixels within these segmented areas, we are able to calculate the window-to-wall ratio (WWR). To illustrate, in Figure 2, we observed that the window mask encompassed 119,606 pixels, while the facade mask contained 349,980 pixels. Consequently, the window-to-wall ratio (WWR) for the given image is calculated as 0.3418.

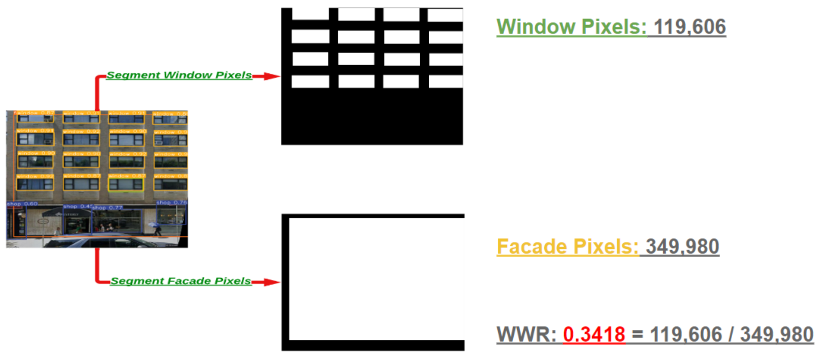



**Fig. 2.** Example for demonstrating the processes for calculating the Window-to-Wall Ratio (WWR) of a building facade based on the areas of windows (upper) and facade (bottom) shown in white.

### 4.3 DNN Model for Estimating ENERGY STAR

The DNN model for predicting the ENERGY STAR rating of buildings is a conventional multiple-layer neural network model 1. The model comprises one input layer, two hidden layers and one output layer. The input layer took in up to 21 inputs, 12 numerical features that we selected based on their importance (Table 4) and 8 one-hot encoded categorical features (5 for boroughs and 3 for property usage types), and one from the WWR precitor. In the 2 hidden layers, we employed 128 nodes and 32 nodes respectively and used the Rectified Linear Unit (ReLU) activation function. To address potential model overfitting, we introduced a 40% dropout between the first hidden layer and the second hidden layer, followed by another 50% dropout between the second hidden layer and the output layer. At the very end, the output layer utilized the softmax activation function to generate a vector consisting of four binary values, each representing one of the four energy efficiency ranks (A, B, C and D). For training, the model utilized the Categorical Cross-Entropy loss function and the Adam (adaptive moment estimation) optimization algorithm. This combination is widely utilized in multi-class classification studies and aligns well with the objectives of our research.

# 5 Experiments and Results

## 5.1 Window and Facade Detection

When the YOLOv5 model is trained (fine-tuned) using 11 labels, we observed that it performs well in detecting facades (82%) and windows (78%) in the CMP dataset compared to the other label classes. However, when specifically examining the "window" label, it exhibited misclassifications of 1% and 3% as "sill" and "blind," respectively. Therefore, we enhanced the accuracy of window detection by fine-tuning the model with only 2 labels (windows and facades). This results in an increase in window detection accuracy from 78% to 80%, although the accuracy for facades detection remains the same at 82%. Additionally, it requires less time to achieve optimal performance.

## 5.2 Multimedia Classification

**Using Traditional Data** In conducting our DNN model evaluation with the 12 numerical features (Table 4) and 8 one-hot encoded categorical features (for building location and usage type information), our analysis revealed a significant influence of building usage types on the accuracy of energy efficiency ranking predictions (Table 5). The model's performance saw notable improvement when tested separately on each building type, from 3.40% for Office Buildings to 8.09% for Multifamily Housing, as opposed to aggregating them into a single dataset (All Types). To further assess the impact of building usage types on the prediction of energy efficiency ranks, we conducted an additional evaluation. This time, we excluded the attributes for primary building usage type (3 of the 8 one-hot encoded categorical features) from the dataset. Our findings demonstrated a marked improvement in model performance when these usage types were omitted (Table 5), particularly boosting up the performance for Office Buildings, to a 2.64% improvement.

**Table 5.** DNN model estimation results with the traditional data.

| Building Types | With Types | Without Types |
|---|---|---|
| All Types | 45.68% | 47.81% |
| Multifamily Housing | 53.77% | 53.85% |
| Office | 49.08% | 51.72% |
| K-12 School | 53.48% | 53.59% |

**Using Window-To-Wall Ratio** We used the YOLOv5 model for 2 classes (facade and window) fine-tuned by the CMP dataset to extract the window-to-wall ratio (WWR) from building images, which we then incorporated into our tabular dataset. Owing to the varying quality of the images and constraints of our model, WWR extraction was feasible for approximately 75% of the 2,972 acquired images for the 3000 building tabular samples. Of the 2,213 WWR values extracted from the 2,972 acquired images, 844 pertained to multifamily housing, 713 to K-12 schools, and 656 to office buildings. Without WWR labeling, we are unable to provide an exact estimation of the precision of WWR determination when employing the fine-tuned YOLOv5 model trained on the CMP dataset. Nevertheless, it is reasonable to anticipate that the model's performance is unlikely to surpass the established accuracy levels of 80% and 82% for façade and window recognition in the CMP test dataset.

Also as shown in Figure 3, we also observed that buildings categorized as multifamily housing and office spaces with the highest energy efficiency ranks exhibited the largest average WWR values. In contrast, K-12 schools with an energy efficiency rank of 'A' displayed the lowest average WWR values. Furthermore, the upper limit of WWR for multifamily housing buildings was notably smaller compared to the other two building types. This phenomenon may be attributed to the prevalent traditional façade style of multifamily housing buildings, in contrast to K-12 schools and offices, which often feature façade styles with higher WWR, such as curtain walls.

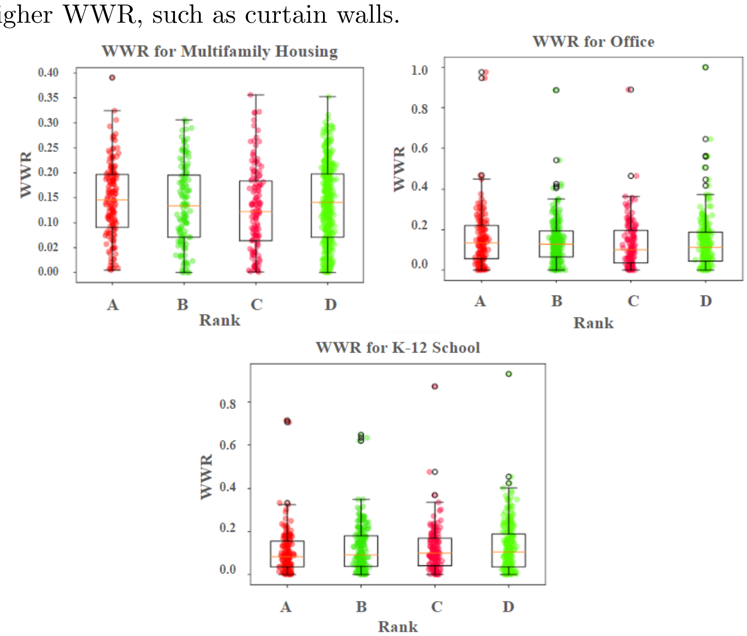



**Fig. 3.** Window-to-Wall Ratio distribution per building usage and energy efficiency ranks for multifamily housing, office and K-12 school buildings. The majority of WWR were clustered between 0.10 to 0.20. Nevertheless, it's worth noting that there were outliers among office and K-12 school buildings.

We assessed our DNN model's performance using the updated tabular dataset, now encompassing WWR values for the buildings. Not surprisingly, our experimentation revealed that the addition of WWR did not yield an improvement in our model's performance (Table 6). It is important to note that the training and testing samples lacking WWR information were derived exclusively from data instances possessing facade images. Consequently, the performance exhibited in this context deviates slightly from that observed when considering the complete dataset. (Table 5).

Initially, we had anticipated that merging WWR into our dataset would enhance the model performance. However, contrary to our expectations, its inclusion resulted in a minor decline in overall model performance. There are multiple reasons for this. First, as we have analyzed above, the accuracy in WWR estimation using the fine-tuned YOLOv5 model is not as high as we expected. Second, the facade images only cover one side of the buildings, and they do not cover the whole side at all (as we can see clearly in Table 3. Third, the orientations of the building would also be a factor in WWR for energy efficiency.

**Table 6.** DNN model estimation results with and without including WWR for the MMBEE samples.

| Building Types | With WWR | Without WWR | Difference |
|---|---|---|---|
| All Types | 55.00% | 53.13% | -0.13% |
| Multifamily Housing | 60.47% | 61.54% | -1.07% |
| Office | 41.89% | 44.14% | -2.25% |
| K-12 School | 49.03% | 49.12% | -0.09% |

**Excluding Low Correlation Features** Our experiments have demonstrated that improving model performance involves a delicate balance in feature selection. We found that removing features with low correlations enhanced the model by reducing distractions. However, relying excessively on a limited set of highly correlated features can also undermine performance. In experiments, we used the tabular dataset with WWR that we generated in the previous section for the WWR test (all types included) the result is shown in Table 7. Setting a correlation threshold of 0.03 for feature selection resulted in a substantial enhancement in estimation accuracy (a 2.24$ improvement than No Exclusion). The new estimation accuracy achieved a z-score of 1.43, approximately corresponding to a 7.55% significance level. Upon setting the threshold at 0.02, a discernible enhancement was still evident; however, it did not represent a significant deviation when contrasted with the inclusion of the entire set of variables. Conversely, opting for a threshold of 0.04 led to the selection of too few features, significantly impairing the model's performance. The z-score for the estimation accuracy plummeted to -5.97, indicating an almost 0% significance level and emphasizing the importance of a balanced approach to feature selection.

This experiment also explained why adding the WWR results as of now cannot improve the performance of estimation. Our analysis showed a correlation of -0.0135 between WWR and a building's ENERGY STAR Score. This correlation falls below the correlation threshold (0.03) that we found to optimize the model's performance and the correlation threshold (0.02) that does not hurt the performance (Table 7).

**Table 7.** DNN model estimation results with different correlation thresholds for the MMBEE samples with all building types.

| Excluding Features | Accuracy | STD |
|---|---|---|
| No Exclusion | 55.00% | 1.56% |
| < abs(0.04) Correlation | 45.68% | 1.63% |
| < abs(0.03) Correlation | 57.24% | 1.71% |
| < abs(0.02) Correlation | 55.85% | 1.46% |

**RF-Based Model** In addition to employing DNN models, we also constructed a random forest (RF) model for our data analysis, as a baseline model to show if the DNN models have some advantages. To determine the optimal parameters, we iteratively explored various combinations, ultimately identifying the most effective configuration as having a tree depth of 16 levels and employing 100 trees in the forest. In comparison to our DNN models, the random forest model yielded worse results (Table 8, for all building types). However, random forest models are less influenced by the correlation coefficient of features. More features do not negatively impact their performance; in fact, having a greater number of features often enhances their effectiveness. On the other hand, our DNN models needed feature selection to mitigate distractions by excluding less correlated features, reflecting the contrasting feature handling approaches between the two kinds of models.

**Table 8.** Comparison of results using RF and DNN models on different correlation thresholds for the MMBEE samples with all building types.

| Excluding Features | Accuracy (RF) | Accuracy (DNN) |
|---|---|---|
| Full Feature Set | 53.77% | 55.00% |
| < abs(0.04) Correlation | 44.13% | 45.68% |
| < abs(0.03) Correlation | 48.95% | 57.24% |
| < abs(0.02) Correlation | 51.51% | 55.85% |

## 6 Conclusion and Discussion

In this paper, we introduced a new Multimedia Building Energy Efficiency dataset - MMBEE, which includes both tabular and image data extracted from publicly available sources. This dataset can not only serve as a Multimedia interface for understanding building energy efficiency but also provide a benchmark for testing machine learning algorithms and computer vision algorithms. We performed a number of important processes and analyses, including feature selection, data correlation, WWR extraction, and various ways of building energy efficiency estimation.

While the dataset would be valuable for research and applications, and the analyses offer some insights into the data and the models, our preliminary experimental results yielded unexpected outcomes, as our initial hypothesis as well as the results from other research groups posited that the inclusion of image data would enhance model performance by providing valuable visual information for predicting building energy efficiency. With the multiple factors we analyzed in the experiment section, our combined model not only failed to improve performance but actually exhibited a slight decrease in accuracy. It is important to note, however, that we cannot definitively conclude that image data lacks utility for energy efficiency prediction.

For one thing, we suspect that the WWR estimation model's limitations, particularly its inability to accurately capture all windows and facade elements in the images, played a role in this outcome. The fact that WWR extraction was successful for only 75% of the images underscores the imperfections in our YOLOv5 model. To address this, we plan to refine the building facade segmentation and obstacle removal processes, which should enhance the YOLOv5 model's ability to accurately capture WWR from images.

In order to acquire the necessary building images for our system, we utilized the GSV API to scrape the data. By providing specific address details, including the street name, city, state, and zip code, the API returns the closest camera facing the building. Additionally, we can flexibly adjust the camera's field of view (FOV) and pitch to optimize the coverage of a building's facade images from the API. However, it is important to note that images obtained from the Google Street View API often exhibit distortion due to varying camera angles, the camera not facing the front side of the building directly. This distortion alters the geometric properties of the building structure, posing challenges for our model to accurately detect facades and windows. In addition, trees, vehicles and pedestrians often occlude the building facades. Therefore, we need to first remove those occluded regions from consideration, for example, using a semantic segmentation model, and then rectify the images of building facades after the removal.

Furthermore, to maximize the utility of WWR in our prediction model, we intend to complement it with additional information, such as building window directions, the orientations of the building facades in the images, and the specific building facade types and building materials. These aspects will be key focal points in our future research efforts to improve the overall accuracy and efficacy of our energy efficiency predictions. In addition, the integration of Window-to-Wall Ratio (WWR) into predictive models necessitates consideration of building typologies, architectural configurations, and material selections as influential variables. For instance, a structure employing a curtain wall design, equipped with high-performance glass (or triple/quadruple-paned windows), might accommodate a more generous WWR while simultaneously upholding energy efficiency and indoor comfort.

In summary, we have worked out a pipeline to automatically extract a Multimedia Building Energy Efficiency (MMBEE) dataset from publicly available databases. As of now, our model's performance did not align with our initial expectations and contradicted existing research findings due to the limitations of the collected image data. Nevertheless, the pipeline and the insights from analyzing the data enable us to bring several aspects to our attention that we believe can help reconcile the disparities between our experimental results and those of other researchers. A notable portion of these considerations revolves around enhancing data quality, encompassing actions like obstacle removal (such as trees, vehicles and pedestrians), image rectification, collection of more relevant image data, and inclusion of supplementary data to building images.

## References

1. Climate Solidarity: Reimagine the future of new york city via co-created scalable urban resilience projects. Available at https://climatesolidarity.nyc/(accessed

2023/10/25))
2. Energy Star: How the 1–100 energy star score is calculated. Available at https://www.energystar.gov/buildings/benchmark/understand_metrics/how_score_calculated (accessed 2023/10/18)
3. Li, C.K., Zhang, H., Liu, J.X., Zhang, Y.Q., Zou, S.C., Fang, Y.T.: Window detection in facades using heatmap fusion. Journal of Computer Science and Technology **35** (2020). https://doi.org/10.1007/s11390-020-0253-4
4. Mayor's Office of Climate and Environmental Justice: NYC 2021 Energy Benchmarking - Mayor's Office of Sustainability. Available at https://www.nyc.gov/site/sustainability/codes/energy-benchmarking.page (accessed 2023/10/21)
5. Mayor's Office of Climate and Sustainability: 2023 Energy grades. Available at https://www.nyc.gov/site/buildings/property-or-business-owner/energy-grades.page (accessed 2023/05/01)
6. Mayor's Office of Climate and Sustainability: Energy and water data disclosure for local law 84 2021: Nyc open data. Available at https://data.cityofnewyork.us/Environment/Energyand-Water-Data-Disclosure-for-Local-Law-84-/usc3-8zwd (accessed 2023/10/20)
7. New York City: Local law 97 - sustainable buildings. Available at https://www.nyc.gov/site/sustainablebuildings/ll97/local-law-97.page (accessed 2023/10/18)
8. Phillip, D., Chen, J., Maksakuli, F., Ruci, A., Sturdivant, E., Zhu, Z.: Improving building energy efficiency through data analysis. Companion Proceedings of the 14th ACM International Conference on Future Energy Systems (2023). https://doi.org/10.1145/3599733.3600244
9. R. Mena, F. Rodríguez, M. Castilla, and M.R. Arahal: A prediction model based on neural networks for the energy consumption of a bioclimatic building. Energy and Buildings **82** (2014). https://doi.org/10.1016/j.enbuild.2014.06.052
10. Sun, M., Han, C., Nie, Q., Xu, J., Zhang, F., Zhao, Q.: Understanding building energy efficiency with administrative and emerging urban big data by deep learning in glasgow. Energy and Buildings **273** (2022). https://doi.org/10.1016/j.enbuild.2022.112331
11. Szcześniak, J.T., Ang, Y.Q., Letellier Duchesne, S., Reinhart, C.F.: A method for using street view imagery to auto-extract window-to-wall ratios and its relevance for urban-level daylighting and energy simulations. Building and Environment **207, Part B** (2022). https://doi.org/10.1016/j.buildenv.2021.108108
12. Troup, L., Phillips, R., Eckelman, M.J., Fannon, D.: Effect of window-to-wall ratio on measured energy consumption in us office buildings. Energy and Buildings **203** (2019). https://doi.org/10.1016/j.enbuild.2019.109434
13. Tyleˇcek, R.: The cmp facade database. CTU Publishing House, Czech Technical University in Prague (September 2012), http://cmp.felk.cvut.cz/ tylecr1/facade/CMP_facade_DB_2013.pdf
14. Ultralytics: Train Custom Data. https://github.com/ultralytics/yolov5/wiki/Train-Custom-Data (accessed 2023/10/21)
15. Ultralytics: YoloV5 Source Code. https://github.com/ultralytics/yolov5 (accessed 2023/10/21)
16. Yalcintas, M., Ozturk, U.A.: An energy benchmarking model based on artificial neural network method utilizing us commercial buildings energy consumption survey (cbecs) database. International Journal of Energy Research **31** (2007). https://doi.org/10.1002/er.1232