

Educational building retrofit: a data-driven modelling approach for optimized energy efficiency

Laura Carnieletto^{1*}, Nicolò Gregori², Francesco Viero³, Mattia D'Inca³, Michele De Carli², Wilmer Pasut¹

¹Dpt. Of Environmental Sciences, Informatics and Statistics, Ca' Foscari University of Venice, Mestre (Venice), Italy

²Dpt. Of Industrial Engineering, University of Padova, Padova, Italy

³ Prisma Engineering S.r.l., Saonara (Padova), Italy

**Corresponding author: laura.carnieletto@unive.it*

Abstract

Energy efficiency and energy retrofit actions have been essential requirements to satisfy national and international directives and regulations that aim at reducing energy consumption and mitigating the impacts of climate change. In Italy, the highest number of actions in the past years were carried out to improve the energy efficiency of residential buildings, which represent around the 80% of the Italian and European building stock. However, the needs of addressing climate change issues shifted also towards non-residential buildings such as schools, which represent a critical target for energy-saving interventions due to their high energy usage and the large number of occupants. School buildings represent a significant part of the public building stock and tend to be less energy-efficient, as many were constructed in the post-World War era when energy efficiency was not a priority. However, the significant initial investments needed to address the challenges of these buildings demand optimized strategies that balance energy efficiency with sustainable costs.

The research developed a predictive model using an Artificial Neural Network to estimate the current space heating energy demand in buildings and to plan future retrofit actions. A dataset was created by parametrically generating artificial buildings with varying geometries and envelope properties, simulating them in EnergyPlus, and recording their energy demand. The trained ANN identified relationships between input parameters and heating demand, enabling accurate predictions. Validation against EnergyPlus simulations confirmed the model's effectiveness, demonstrating its potential for building energy performance assessment. The findings are expected to inform future policy decisions and guide investments in energy retrofitting for educational institutions across the country, while considering the limitations due to the limited timeframes available for the retrofit installation, the operational needs of these buildings and the type of actions applicable.

The identification of the most effective strategies can support the development of policies that not only enhance energy efficiency but also contribute to the broader goals of sustainability and climate resilience in the built environment. Furthermore, the need of verifying that the expected savings are realized in practice highlight the requirement of ongoing monitoring and evaluation,

thereby closing the gap between predicted and actual performance. Concluding, educational buildings play a key role as example of sustainable practices, educating future generations about the impact of energy-related emissions.

Key Innovations

- Artificial Neural Network (ANN) Model: Development of a predictive model to estimate space heating energy demand and support retrofit planning.
- Parametric dataset creation: artificial buildings with varying geometries and envelope properties were generated and simulated in EnergyPlus to build a comprehensive dataset from a set of real building properties and characteristics
- Model validation: the trained ANN was tested against EnergyPlus simulations, confirming its ability to accurately predict energy demand.

Practical Implications

- Expansion of energy efficiency efforts: shift from a primary focus on residential buildings to include non-residential structures, particularly schools, due to their high energy usage and public role.
- Educational and policy impact: school buildings serve as educational tools for future generations on energy efficiency and climate impact.
- Guiding future policy and investments: findings can support informed decision-making on energy retrofitting strategies.
- Consideration of operational constraints: acknowledgment of limited installation timeframes, operational requirements, and applicable retrofit actions in school buildings.
- Sustainability and climate resilience: policies should align energy efficiency improvements with broader sustainability goals.
- Performance monitoring: emphasis on the need for ongoing evaluation to bridge the gap between predicted and actual energy savings.

Introduction

The rise in global population and the urgent need to reduce emissions from building energy consumption make predictive models essential for developing reliable and efficient energy efficiency strategies. While most efforts have focused on residential buildings, which represent about 80% of the building stock, there is now a shift towards energy-saving measures in non residential

constructions. Educational buildings, with high energy consumption and large occupant numbers, are often less energy-efficient due to their post-World War construction; therefore, they require the planning of optimized strategies that balance energy savings with sustainable costs.

Several attempts have been made in literature to predict buildings' energy demand and potential retrofit strategies, ranging from high-detail single-building models using EnergyPlus (U.S. DOE, 2022), TRNSYS (Beckman et al., 1994), and IDA-ICE (Equa simulation, 2022) to urban building energy modeling (UBEM) tools like EURECA (Prataviera et al., 2021), CityBES (Hong et al., 2016), and UMI (Reinhart et al., 2013), as well as deep learning approaches based on artificial neural networks (ANNs). In particular, the ANN-based approach is particularly promising for forecasting building energy demand because it is possible to capture complex nonlinear relationships between the factors influencing the buildings' behaviour and performance, adapting to large and diverse datasets. Unlike traditional physics-based models, which require extensive input data and calibration, ANN models can be trained on diverse building datasets and then applied to different building types and climates with relatively minimal adjustments.

Mohan and Pachauri (2024) tried to develop a model based on multiple machine learning (ML) algorithms capable of accurately predict the heating and cooling load of a building. The learning algorithms chosen were the extreme gradient boosting (XGB), decision trees (DT) and random forest (RF), which were used to identify and learn the input-output relationships from a specific dataset of buildings. The dataset considered was generated by choosing 12 real building floor plans and systematically modifying different parameters such as height, orientation and glazed area. Overall, the combinations of buildings generated is equal to 768, with various structural parameters. Afzal et al. (2023) used several ANN to evaluate the energy demand of a series of residential buildings. Relative compactness, surface area, roof area, wall area, orientation, overall height, glazing area (comprising glazing, frame, and sash components), and glazing area distribution were used as input parameters, investigating the heating and cooling load of the building.

Along with the different network implemented for the definition of the predictive model, several optimizing algorithms have been tried to find the perfect configurations of hyperparameters that maximizes the network performance. Olu-Ajayi et al. (2023) predicted the annual residential building energy usage with ML algorithms such as Gradient Boosting (GB), Deep Neural Networks (DNN), Random Forest (RF), Stacking, K Nearest Neighbors (KNN), Support Vector Machines (SVM) and Decision Tree (DT). Similarly, Karijadi and Chou, (2022) introduced a hybrid RF-LSTM (Long Short-Term Memory) method, augmented by Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) process to decompose data into frequency components. Their approach showed superior

results when tested on actual building energy records. Cao et al. (2023) proposed a Particle Swarm Optimization-Stacking Improved Ensemble (PSIE) that integrated 11 ML algorithms, applying a priority feature selection approach to enhance predictive accuracy.

Several researchers have continued to refine hybrid ensembles. Zheng et al., 2023b presented an architecture that fuses pattern recognition (DECPR) with temporal fusion transformers (TFT), reducing the mean average percentual error (MAPE) to 6.11%, which is a common statistical metric used to measure the accuracy of a predictive model. It calculates the average percentage difference between predicted and actual values that is particularly useful to evaluate forecasting models. Khan et al., (2023) developed a framework merging optimal data preprocessing with ConvLSTM and BDGRU for feature extraction, followed by Multi-Layer Perceptron (MLP) forecasting. By optimizing data flow, they achieved substantial MSE reductions compared to more conventional models. Y. Li et al. (2022) built an interval forecasting model based on fuzzy information granulation, attention mechanisms, and LSTM, demonstrating the value of uncertainty estimation in future energy predictions.

This work aims to develop an advanced predictive model for estimating space heating energy demand in educational buildings using an Artificial Neural Network. Unlike conventional approaches, which often rely on simplified methods, this study builds upon a dataset generated through parametric variations of real building geometries and envelope properties, ensuring a more accurate representation of diverse building typologies. By integrating dynamic EnergyPlus simulations, the model effectively captures the complex relationships between architectural and thermal parameters, enabling reliable energy demand predictions. This innovative approach enhances energy performance assessment and provides a robust tool for optimizing retrofit strategies and sustainable building design.

Methods

Description of the dataset of real buildings

The dataset considered consists of 41 school buildings located in northern Italy, corresponding to climatic zone E. They have been constructed between year 1950 and year 2000, and include nursery schools, elementary schools, and high schools, meaning that the dataset is characterized by buildings whose sizes are within a relatively large range (Table 1). The heated surface values follows a normal distribution, with most values ranging between 328 m² and 6000 m². However, the presence of larger buildings results in a high standard deviation and a relatively elevated mean. The mean and median values of the surface-to-volume ratio (S/V) are very close, indicating a nearly symmetric distribution. Most buildings fall within the 0.28 to 0.55 range, a relatively broad interval due to the diverse building shapes in the dataset. This variability in shape also contributes to significant differences in heating energy demand. The thermal

transmittance values of external walls are nearly evenly distributed between 0.39 and 1.34, showing a symmetric and homogeneous distribution. This is also reflected in the mean (0.84) and median (0.81) values, which are closely aligned. The wide range of transmittance values can be attributed to the construction periods of the buildings, spanning from the 1940s to the late 2000s. Consequently, variations in construction materials, element thicknesses, and building technologies have led to significant differences in thermal performance. From this database, 15 buildings with specific characteristics have been simulated with detailed single-building EnergyPlus models and used for the validation of the ANN algorithm, as will be presented in the Results section.

Table 1. Main statistical indicators of the real dataset.

Parameter	Mean	Median	STD	Min	Max
Heated floor area [m ²]	5264	4711	4088	328	17976
Heated Volume [m ³]	15979	13618	11682	1312	53832
S/V [m ⁻¹]	0.44	0.42	0.10	0.28	0.73
S _{opaque} [m ²]	7904	5693	6049	1054	26870
S _{windows} [m ²]	496	400	387	38	1405
U _{EXT WALL} [W/(m ² K)]	0.84	0.81	0.26	0.39	1.34
U _{FLOOR} [W/(m ² K)]	0.48	0.50	0.07	0.25	0.60
U _{ROOF} [W/(m ² K)]	0.93	0.64	0.51	0.20	2.02
U _{WINDOWS} [W/(m ² K)]	2.32	2.30	0.50	1.50	4.00

Description of the training dataset

To train the Artificial Neural Network, a dataset of buildings representing a significant variety of possible educational buildings in northern Italy has been artificially created. Python scripts were applied to modify the IDF files corresponding to the detailed single-building energy models by varying the model inputs to obtain different building configurations. The four base geometries included rectangular shape, L-shaped, U-shaped and square shape with internal cloister. Simulations have been carried out with Energy Plus to calculate the space heating energy demand. Space cooling has not been considered because it is not commonly installed in Italian educational buildings. Geometry scaling, thermal transmittance and windows' size have been modified to create the artificial dataset. As can be seen in Table 2, most constructions fall within the range of 104 m² to 5000 m², with fewer buildings at larger sizes. This trend is due to the modification process, where building dimensions were scaled by linearly increasing their length and width. However, this linear scaling resulted in a quadratic increase of the ground surface area, leading to a parabolic growth trend. Additionally, the maximum heated volume in the real building dataset is 53832 m³, which falls within the first quarter of the volume range in the artificial dataset. This suggests that most real buildings have geometrical parameters

concentrated in this region of the distribution, where the artificial dataset is particularly dense. Observing the trend of thermal transmittance for external walls and roofs, the values become more closely spaced as they increase. This occurs because transmittance is modified in each IDF file by varying thermal conductivity over a uniformly distributed range while keeping the thickness constant. However, since the calculation also incorporates fixed internal and external convective heat transfer coefficients, the resulting transmittance values are not evenly distributed but tend to cluster toward the upper end of the range. The size of the windows was varied with a scaling factor of 0,025, which represents the fraction of the initial distance used as the incremental step in each stage of the scaling process. The final dataset included a total of about 2.7 million of generated buildings.

Table 2. Main statistical indicators of the artificial dataset.

Parameter	Mean	Median	STD	Min	Max
Heated floor area [m ²]	4905	2795	4987	104	25251
Heated Volume [m ³]	44574	25251	47502	416	196608
S/V [m ⁻¹]	0.41	0.34	0.18	0.17	1.08
Building Height [m]	10.00	8.00	4.40	4.00	16.00
S _{opaque} [m ²]	6985	4504	6324	116	28872
S _{windows} [m ²]	1425	946	1432	12	9728
U _{EXT WALL} [W/(m ² K)]	0.83	0.73	0.50	0.05	1.50
U _{WINDOWS} [W/(m ² K)]	3.25	2.90	1.20	1.50	5.00
U _{ROOF} [W/(m ² K)]	1.44	1.35	0.84	0.04	2.50
Heating operation [h]	8.00	8.00	1.63	6.00	10.00

Multi-Layer Perceptron algorithm

Neural networks are designed to mimic the structure and information-processing methods of the biological nervous system, enabling them to identify patterns, analyze data, and draw conclusions. Their accuracy improves over time through training on relevant datasets. Once optimized, they serve as powerful tools for data classification and clustering. Feedforward neural networks, also known as multi-layer perceptrons (MLPs), are among the earliest types and rely on sigmoid neurons, as most problems are nonlinear. During the training process, the network analyzed the dataset provided to learn the relationships between the building parameters and their corresponding energy demand, obtained through simulations.

The training script preprocesses and normalizes the dataset while defining the neural network's architecture and hyperparameters. A significant portion of the dataset (99.5%) is used for training, allowing the network to understand variable dependencies, while the remaining 0.5% is reserved for testing. In case of large datasets with a significant number of input variables, a large number of

epochs is required to effectively learn complex patterns and relationships within the data. Each epoch represents a complete pass through the entire dataset, allowing the network to iteratively adjust its internal weights and improve prediction accuracy, helping the model to gradually refine its learning process and avoid underfitting. At the same time, a good depth of the network (i.e., the number of neurons available in each layer of network) is essential to capture intricate dependencies between variables. However, to prevent overfitting, techniques such as early stopping or validation monitoring should be applied to determine an optimal number of epochs. A neural network with a significant quantity of neurons in its first layers represents a good trade-off between quick learning and computational effort required, in particular for relatively large amount of input data. For these reasons, the number of epochs chosen for the training of this network has been fixed equal to 1000. Considering the size of the dataset, containing approximately 2,7 million of buildings, the amount of time required to perform the training process was equal to 20 hours with the equipment used. Once the training of the network has been properly described, several trials have been performed either by varying some hyperparameters or modifying the dataset (i.e. variation of the number of samples in the dataset), to find the ANN with the best performance. A summary is reported in Table 3. N6-N10 implemented different combinations of variables, while N11, N12 and N13 modified the size of the dataset. In particular, for N11 and N12 the number of buildings for the training was halved and then reduced to a quarter, respectively, to observe the behavior of the network with a smaller sample. The reduction of the dataset for N11 and N12 was performed acting on the windows' scaling, while N13, the reduction process has also involved the variation of size of buildings.

Table 3. Input variables of the networks.

Parameter	N6	N9	N10	N11	N12	N13
Heated floor area [m ²]	•	•	•	•	•	•
Heated Volume [m ³]	•	•	•	•	•	•
S/V [m ⁻¹]	•	•		•	•	•
Building Height [m]	•		•	•	•	•
Ext wall Area [m ²]			•			
S _{opaque} [m ²]	•			•	•	•
S _{window} [m ²]	•		•	•	•	•
S _{window} /S _{opaque} [-]		•				
U _{EXT WALL} [W/(m ² K)]	•	•	•	•	•	•
U _{WINDOWS} [W/(m ² K)]	•	•	•	•	•	•
U _{ROOF} [W/(m ² K)]	•	•	•	•	•	•
Heating operation [h]	•	•	•	•	•	•

For each building of the test set, a percentual error is calculated between the EnergyPlus simulation outcome

and the value predicted by the network. The maximum value, the mean value and the standard deviation have been also calculated (Table 4). The average percentage error is low across all trained networks, indicating strong predictive performance on the test set buildings. This is further supported by the low standard deviation values, which shows that the percentage errors for different buildings are closely clustered around the average. Among the networks, N11 demonstrates the best performance, with an average error of just 0.3%. In contrast, networks with a larger number of samples, such as N06 and N09, may suffer from overfitting, ultimately reducing overall performance.

Table 4. Performances of the different networks.

	N6	N9	N10	N11	N12	N13
Max error	12.9%	43.1%	12.7%	20.4%	50.5%	14.8%
Mean error	0.3%	0.4%	0.4%	0.3%	0.4%	0.4%
STD	0.6%	0.8%	0.6%	0.5%	1.1%	0.6%

Along with these indicators, the performance of the best network obtained after the training process (N11) is shown if Figure 1, representing two curves:

- The training loss function, calculated for all batches during training;
- The validation loss function, computed only for the buildings in the validation split.

As observed in the chart, the blue curve (training loss) and the orange curve (validation loss) exhibit the same downward trend, which should ideally follow a similar parabolically descending trend. To improve readability across different orders of magnitude, a logarithmic scale is applied to the y-axis. However, the validation loss curve shows noticeable oscillations. This behavior can be attributed to several factors:

1. heterogeneous data distribution: the dataset includes a wide range of building sizes and structural parameters (e.g., varying thermal transmittance values), making it more challenging for the network to generalize effectively.
2. mini-batch size: when small batches are used, the validation set includes only a limited number of samples, which may not fully represent the dataset's variability. This can result in instability in the loss function evaluation.

Despite these oscillations, the overall downward trend continues, stabilizing around 1000 epochs, indicating that the training process has effectively converged.

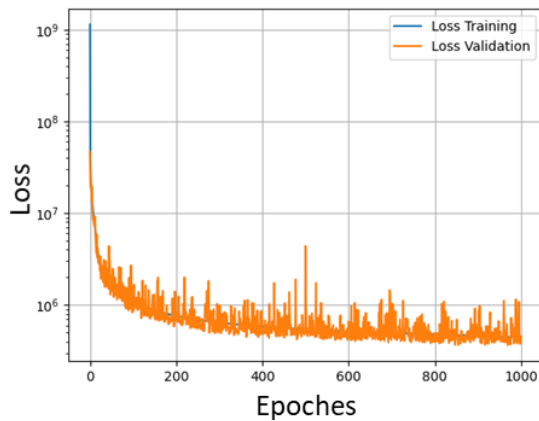


Figure 1: Loss functions of N11.

Figure 2 allows to observe the accuracy of the network when predicting the output. This is performed by considering the last 100 buildings of the test set comparing the predicted energy demand and the outcome obtained from the detailed energy simulation that almost overlap, meaning that the network correctly predicts the values of energy demand. Since the scale of the graph is limited, the differences between the points are not considerable, so in order to make more precise considerations on the prediction performances of the network, the numerical indicators have to be considered.

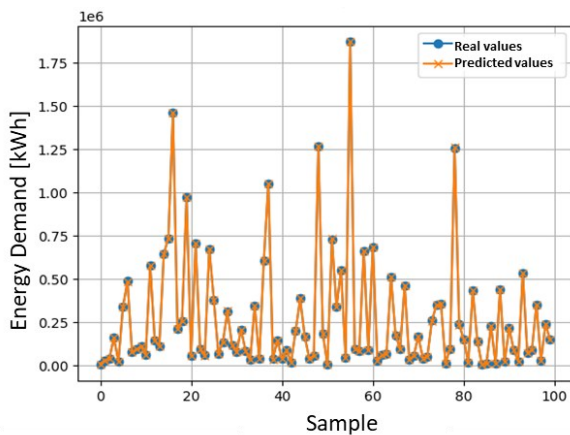


Figure 2: Comparison between predicted and simulated values (N11).

Results

To enhance the model's ability to predict building energy demand, it was tested on a selection of 15 buildings with diverse and complex shapes from the dataset of real buildings previously presented. The buildings were selected with irregular geometries and specific structural parameters, evaluating the model's performance within different conditions than those used for training. Each building was simulated using EnergyPlus, and the predicted energy demand from the trained networks was compared with the simulation results. The relative percentage error between the two values was calculated, considering a threshold of 15% above which predictions were considered unacceptable. This value was chosen according to ASHRAE Guideline 14-2014, which

establishes accuracy thresholds for energy simulation based on the mean bias error (MBE) and the coefficient of variation of the root mean square error (CV(RMSE)). Additionally, literature commonly considers an error range between 10% and 15% to be acceptable. The highest errors were observed for buildings S02, N11, and S18, characterized by highly complex geometries with multiple narrowings or separate sections. Since these features were not well-represented in the training dataset, the model struggled to accurately predict their energy demand. However, as shown in Table 5, the average percentage error for most networks remained significantly below 15%, indicating that the models provide accurate predictions for real buildings. The exception lies in cases where the geometry is significantly complex, leading to a higher prediction error. Among the tested networks, N11 demonstrated the best performance, which is consistent with the results obtained during model validation on the test set. At the end of each training attempt, statistical indicators were computed for the test set, confirming that Network 11 (N11) consistently outperformed compared to the others. This conclusion was further validated through testing on real-world buildings.

Table 5: Indicators for the different networks tested on real buildings.

Network	Mean percentual error	STD	RMSE
N06	7.16%	5.00	40553
N09	7.77%	5.42	37221
N10	7.76%	5.11	39877
N11	6.10%	3.04	29862
N12	6.51%	3.45	32876
N13	6.70%	5.17	40197

Figure 3 reports the energy demand of a set of real buildings calculated with EnergyPlus (blue), the best ANN (N11 - orange) and the worst ANN (N09 - grey). For the majority of the considered buildings, the outcomes obtained with N11 are closer to the values obtained from the single building detailed dynamic simulations.

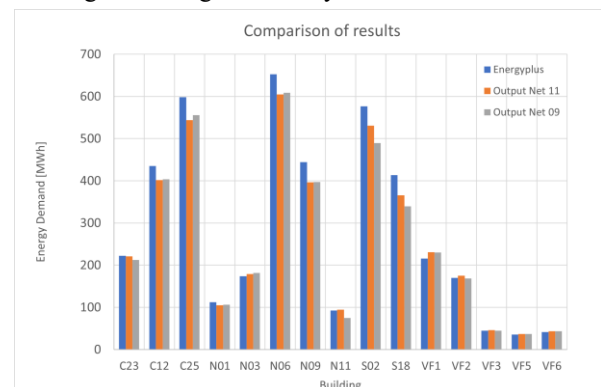


Figure 3: Comparison between predicted and simulated values.

After training and evaluating different neural networks based on statistical indicators, additional tests assessed the model's ability to predict energy demand for real buildings. Before this comparison, the dataset was analyzed examining the error trend between predicted and actual energy demand in relation to roof thermal

transmittance. This was done for different building sizes and levels of external wall insulation. For small buildings with highly insulated external walls, error trends were analyzed based on varying roof thermal transmittance and different daily heating periods (Figure 4).

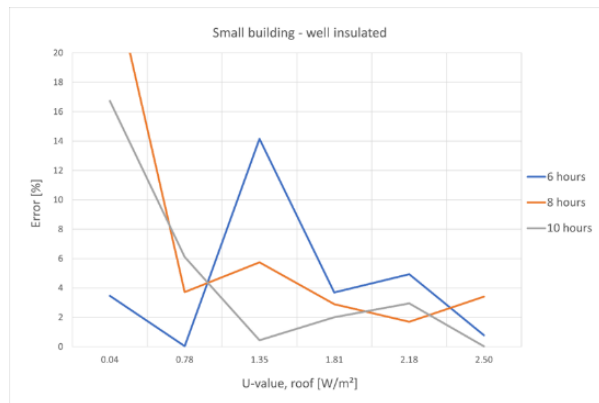


Figure 4: Error curves for a small and well insulated building.

The results showed significant variability, with three distinct error profiles emerging. Notably, prediction errors were higher for lower thermal transmittance values due to the dataset's construction hypotheses. The roof thermal transmittance values were denser at higher values, while lower values had larger gaps, leading to poor generalization and increased error in that range. A similar issue affected external walls, where thermal transmittance values were unevenly distributed, further complicating the network's ability to generalize. Since the dataset contained a higher proportion of small buildings, the network had limited exposure to varied insulation levels, making it difficult to learn the impact of thermal transmittance effectively. This resulted in inconsistencies in the predictions, with sharp variations in error trends. For example, a high percentage error was observed for the lowest roof transmittance value, followed by a steep decrease at the next transmittance level. These findings highlight the challenges in dataset construction and emphasize the need for a more balanced distribution of input variables to improve the model's predictive accuracy. Additional error profiles were examined for buildings with small dimensions and low external wall insulation. Figure 5 shows three curves with a more stable trend, without significant spikes but still displaying some variability. In this case, the thermal transmittance values of the external walls fall within the higher-density region, where more variations are present in the dataset. This increased variability allows the network to better capture the relationship between external wall transmittance and energy demand, resulting in lower prediction errors compared to the previous case.

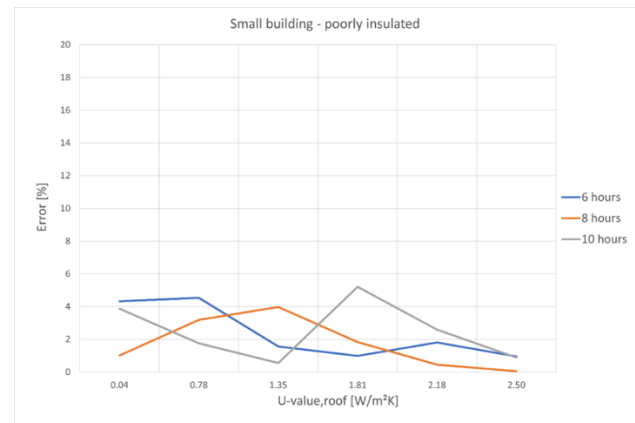


Figure 5: Error curves for a small and poorly insulated building.

Comparison between EnergyPlus and ENEA method

To further assess the prediction capabilities of the model created, the outcomes of the neural network related to the real buildings were compared to the value of energy demand calculated by means of a basic relationship provided by ENEA (2012). This simplified formula (Eq.1), is commonly used by professionals to provide an estimation of the energy demand for space heating to work on draft energy strategies for stakeholders. It is based on some structural and climatic parameters:

$$Q_a = GG \cdot 24 \cdot f \cdot R \cdot U \cdot \frac{S}{1000} \quad [kWh] \quad [11]$$

Where:

- GG = Heating Degree Days for the location
- f = Correction factor to consider the heating operation schedule
- R = Correction factor of the temperature difference of the boundary environment
- S = total surface of the building element considered.

The previous relationship should be adapted considering each structural element of the buildings and the related properties. The energy demand of the 15 real buildings considered to test the network has been estimated also with the ENEA equation, and then compared with the outputs of the EnergyPlus simulations, calculating the percentual error of the estimated value. The results are reported in Table 6. The ENEA equation leads to a significant overestimation compared to the outcomes simulated with EnergyPlus, giving an error higher than 15% for more than half of the buildings meaning that this simplified relationship often tends to overestimate or underestimate of the heating energy demand, thus leading for example to an oversizing of the heating system required that will translates into higher investment costs.

Table 6: Comparison between EnergyPlus and ENEA outcomes

Energy Demand			
Building	EnergyPlus	ENEA	Deviation
[-]	[kWh]	[kWh]	[-]
C23	222405	224210	0.8%
C12	435173	335130	23.0%
C25	598147	536761	10.3%
N01	112304	107867	3.9%
N03	173728	121401	30.1%
N06	652489	468962	28.1%
N09	444326	348682	21.5%
N11	92139	76244	17.3%
S02	576733	578201	0.3%
S18	413456	351820	14.9%
VF1	215602	190705	11.6%
VF2	169966	192896	13.5%
VF3	44301	51438	16.1%
VF5	35217	43731	24.2%
VF6	41167	51824	25.9%

The comparison is reported in Table 7, showing that N11 has the best performance. The relative percentual error was calculated between the simulated value and the output value from the model (Deviation EPlus-N11), or between the simulated value and the ENEA output (Deviation ENEA-N11). The difference between the two deviations (ΔE) is useful to understand which is the more precise method to predict the value of energy demand obtained from the EnergyPlus simulation. It can be clearly seen that, except for one single case, the value of ΔE is negative, meaning that the percentual error of the MLP model is lower compared to the error of the simplified method. Furthermore, the values of ΔE are generally lower than zero, meaning that the network provides values of energy demand which are much closer to the EnergyPlus output. The only exceptions involve the building N01 and S02, where the results of the ENEA relationship return a value of energy demand which is closer to the actual value from the simulation compared to the MLP model outcome. This may be due to the fact that the ENEA method primarily overlooks two key factors, solar gains and air infiltration, which have opposite effects and tend to compensate for each other. Although this is the reason why the formula still produces reasonably accurate results, it is difficult to determine why ENEA appears more precise in specific cases. Accuracy should be assessed on a large scale rather than on individual buildings, where ENEA's performance can be more inconsistent.

Table 7: Comparison between EnergyPlus and ENEA outcomes

Building	N11	Deviation EPlus-N11	Deviation ENEA-N11	ΔE
[-]	[kWh]	[-]	[%]	[%]
C23	221114	0.6%	0.8	-0.2
C12	401880	7.7%	23.0	-15.3
C25	543906	9.1%	10.3	-1.2
N01	104871	6.6%	4.0	2.7
N03	178877	3.0%	30.1	-27.2
N06	604816	7.3%	28.1	-20.8
N09	396433	10.8%	21.5	-10.8
N11	94620	2.7%	17.3	-14.6
S02	530570	8.0%	0.3	7.8
S18	365776	11.5%	14.9	-3.4
VF1	230834	7.1%	11.6	-4.5
VF2	174745	2.8%	13.5	-10.7
VF3	45944	3.7%	16.1	-12.4
VF5	37081	5.3%	24.2	-18.9
VF6	43396	5.4%	25.9	-20.5

Conclusion

The main goal of this work was to develop a predictive model able to estimate the annual energy demand of school buildings based on input variables that describe their geometry and structural characteristics, in order to provide a tool for accurate and practical energy savings strategies. This was achieved by training a Multi-Layer Perceptron (MLP) neural network using a large dataset of artificially generated buildings from a database of real constructions, each paired with its corresponding energy demand for space heating. The aim was to enable the model to accurately learn the relationships between input parameters and energy consumption. To evaluate the model's predictive accuracy, it was tested on a set of 15 real buildings. The predicted energy demand was compared with actual values obtained through detailed single building dynamic simulations. The model's predictions closely aligned with actual energy demands, with an average percentage error of 6.1% that can be considered a promising result, because it estimates the energy demand of generic buildings with a relatively high degree of accuracy. Additionally, the MLP model significantly outperformed the simplified method developed by ENEA, which is widely used for rough energy demand estimations and only accounts for heat losses through the building envelope. The ENEA method showed an average percentage error of 16.1%, which is considerably higher than the error of the MLP model. Despite its strong performance, the predictive model has some limitations that can affect its accuracy. The main challenges stem from the dataset used for training, because of the higher density of small buildings compared to those with larger dimensions, which leads to a non-

uniform data distribution, thus to model overfitting and reduced predictive accuracy for larger structures. Furthermore, the dataset was generated from four base geometries, chosen to represent typical school buildings in northern Italy. However, extending the analysis to other end uses or real-world buildings with irregular shapes, the training data should be updated to increase the representativeness. Since building geometry directly influences energy demand, this limitation affects prediction accuracy for irregularly shaped buildings like S18, where the model produced an error exceeding 10%. To enhance the model's accuracy and applicability, the scaling process can be improved for a more uniform distribution of building sizes and a broader and more balanced representation of envelope thermal transmittance. A refined approach would reduce the number of samples by decreasing the resolution of geometric scaling while incorporating a greater variety of other influential parameters. Future versions of the model can also integrate factors such as end uses, building orientation and climatic zones, which significantly impact energy demand. This would enhance the model's ability to generalize across different conditions. With these upgrades, the predictive model can become a more robust tool for energy performance assessment, supporting better-informed retrofit strategies and policy decisions.

Acknowledgement

This study was funded by the European Union - NextGenerationEU, in the framework of the GRINS - Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 - CUP H73C22000930001), National Recovery and Resilience Plan (NRRP) - PE9 - Mission 4, C2, Intervention 1.3. The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

References

- Pereira, L.D., Raimondo, D., Corgnati, S. P., Gameiro da Silva, M., (2014) Energy consumption in schools – a review paper, *Renewable and Sustainable Energy Reviews* 40, 911-922.
- U.S. Department of Energy. (2022). *EnergyPlus Engineering Reference: The Reference to EnergyPlus Calculations* (Version 22.1.0).
- Beckman, W. A., Broman, L., Fiksel, A., Klein, S. A., Lindberg, E., Schuler, M., & Thornton, J. (1994). TRNSYS—the most complete solar energy system modeling and simulation software. *Renewable Energy*, 5(1-4), 486-488.
- EQUA Simulation AB. (2022). *IDA Indoor Climate and Energy (IDA-ICE) Version 4.8*.
- Hong, T., Chen, Y., Luo, X., Luo, N., & Lee, S. H. (2016). CityBES: A Web-based Platform to Support City-Scale Building Energy Efficiency. *Urban Computing 2016: 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1-2.
- Reinhart, C. F., Dogan, T., Jakubiec, J. A., Rakha, T., & Sang, A. (2013). UMI—An urban simulation environment for building energy use, daylighting and walkability. *13th Conference of International Building Performance Simulation Association*, 476-483.
- Prataviera, E., Romano, P., Carnieletto, L., Pirotti, F., Vivian, J., Zarrella, A., (2021). EURECA: An open-source urban building energy modelling tool for the efficient evaluation of cities energy demand, *Renewable Energy*, 173, 544-560.
- Mohan R., Pachauri, N., (2024). An ensemble model for the energy consumption prediction of residential buildings, *Energy* 314.
- Afzal, S., Shokri, A., Ziapour, B. M., Shakibi, H., Sobhani, B. (2024). Building energy consumption prediction and optimization using different neural network-assisted models; comparison of different networks and optimization algorithms, *Engineering Applications of Artificial Intelligence*, 127, Part B.
- Olu-Ajayi, R., Alaka, H., Sulaimon, I., Balogun, H., Wusu, G., Yusuf, W., Adegoke, M. (2023). Building energy performance prediction: A reliability analysis and evaluation of feature selection methods, *Expert Systems with Applications*, 225.
- Kharjadi I., Chou, S.-Y. (2022). A hybrid RF-LSTM based on CEEMDAN for improving the accuracy of building energy consumption prediction, *Energy and Buildings*, 259.
- Cao, Y., Liu, G., Sun, J., Bavirisetti, D. P., Xiao, G., (2023). PSO-Stacking improved ensemble model for campus building energy consumption forecasting based on priority feature selection, *Journal of Building Engineering*, 72.
- Zheng, P., Zhou, H., Liu, J., Nakanishi, Y., (2023). Interpretable building energy consumption forecasting using spectral clustering algorithm and temporal fusion transformers architecture, *Applied Energy*, 349.
- Khan, S. U., Khan, N., Ullah, F. U. M., Kim, M. J., Lee, M. Y., Baik, S. W., (2023). Towards intelligent building energy management: AI-based framework for power consumption and generation forecasting, *Energy and Buildings*, 279.
- 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.
- American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). (2014). *ASHRAE Guideline 14-2014: Measurement of Energy, Demand, and Water Savings*. ASHRAE.
- Agenzia Nazionale per le Nuove Tecnologie, l'Energia e lo Sviluppo Economico Sostenibile, ENEA, 2012. *Methodologies and tools for the evaluation of building energy performance*