Social Sciences & Humanities Open 13 (2026) 102647

Contents lists available at ScienceDirect

Social Sciences & Humanities Open

journal homepage: www.sciencedirect.com/journal/social-sciences-and-humanities-open

Regular Article

Artificial and convolutional neural networks applications to predict heating and cooling loads for residential buildings

Derrick Mirindi [a,b,*], Manisha Kandel [c,**], Abiodun Adedoyin [a], David Sinkhonde [d], Tajebe Bezabih [d], Ahana Bayongwa Samuel [e,f], Eberechi Cecilia Osuagwu [g], Frederic Mirindi [h]

[a] School of Architecture and Planning, Morgan State University, USA
[b] School of Engineering and Applied Science, University of Pennsylvania, USA
[c] Department of Civil, Construction and Environmental Engineering, University of Delaware, USA
[d] Department of Civil Engineering, Pan Africa University Institute for Basic Sciences and Technology, Kenya
[e] Institute of Water and Energy Sciences Including Climate Change, Pan Africa University, University of Tlecen, Algeria
[f] Institute of Climate and Energy Systems, Julich Systems Analysis, Forschungszentrum Julich, Germany
[g] Forestry and Environmental Management, Michael Okpara University of Agriculture, Nigeria
[h] Department of Economics, University of Manitoba, Canada

ARTICLE INFO

Keywords:
Artificial neural network
Building energy prediction
Convolutional neural network
Cooling load
Heating load

ABSTRACT

In the face of the rapid evolution of artificial intelligence (AI) and the increasing cost of energy in residential buildings, accurately predicting thermal loads has become crucial for sustainable construction practices. We present the use of artificial neural network (ANN) and convolutional neural network (CNN) models to predict the energy efficiency of residential buildings. The dataset comprises eight input parameters, namely surface area, relative compactness, wall area, overall height, roof area, glazing area, orientation, and glazing area distribution, with two output thermal loads (the heating load (HL) and the cooling load (CL)). Additionally, we split the data, comprising 768 observations, into training (70%), testing (15%), and validation sets (15%). Results based on the Pearson correlation matrix indicated that all input variables exhibit a positive correlation with the thermal loads, except the surface and roof areas of the building. In addition, the feature importance and Shapley Additive exPlanation (SHAP) analysis demonstrated that building geometry parameters, such as relative compactness, wall, surface, and glazing areas, dominate thermal load predictions. Furthermore, the ANN models showed high performance, with R^2 values ranging from 0.9618 to 0.9783 for HL and CL. However, the CNN models significantly outperformed ANN models. When comparing training, testing, and validation, CNN models achieve exceptional R^2 values exceeding 0.99 for all dataset splits, even in the presence of outliers. K-fold cross-validation analysis demonstrated the outstanding reliability of the CNN models, with coefficient of variation (CV) values of 0.26% for HL and 0.65% for CL, suitable for engineering applications and real-world deployment. However, the ablation study results identified the non-regularized CNN configuration as optimal for production deployment, having low gap metric values between training and validation HL (−0.0001) and CL (0.0040) models. Beyond technical achievement, this research demonstrates that building energy prediction serves as a tool for advancing household energy consumption. Community engagement and five ethical considerations are proposed for citizen science programs and scientific education initiatives focused on sustainable energy consumption.

1. Introduction

The contemporary construction sector is under greater pressure than ever to be more environmentally responsible and energy-efficient regarding residential infrastructure (Mirindi & Yazdandoust, 2024). The global nature of the challenge is demonstrated by countries that have adopted stringent building energy codes. For instance, this evidence can be illustrated in countries such as Germany (with the Passivhaus standard stating that an annual space cooling load (CL) of a building should not exceed 15 kWh/m^2/year to be certified) (Thinking,

* Corresponding author. School of Architecture and Planning, Morgan State University, USA.
** Corresponding author.
E-mail address: demir1@seas.upenn.edu (D. Mirindi).

https://doi.org/10.1016/j.ssaho.2026.102647
Received 15 September 2025; Received in revised form 2 March 2026; Accepted 3 March 2026
Available online 10 March 2026

2015), or the United States (ASHRAE 90.1 standards) (Goel et al., 2017). This urgency has sparked a high interest in optimizing thermal management systems, particularly heating, ventilation, and air conditioning (HVAC) technologies, which are the largest energy consumers in building operations. The United States Department of Energy (DOE) claims that heating and cooling-related consumption constitutes 35% of the energy (Energy, 2025). Some factors include misuse or improper use of equipment, building, occupancy, or design. The significance of energy use in buildings extends to other economic considerations; hence, energy use is a crucial aspect of international efforts to promote sustainability. The United Nations (UN) environmental program report indicates that buildings contribute to about 34% of the world's final energy consumption (Unep, 2024). This result has made it the priority to create superior predictive models that can effectively forecast energy demands and optimize the performance of systems to meet global net-zero goals by 2050 (Globalabc, 2025).

In the building industry, machine learning (ML) and deep learning (DL) algorithms have become transformative tools for tackling complex prediction tasks due to remarkable technological advancements in recent decades of artificial intelligence (AI) (Mirindi et al., 2024). The recent rise of ML and DL, subsets of AI, in building energy prediction, has been propelled by numerous converging factors (Ardabili et al., 2022). First, the rapid evolution in the availability of building performance data and sensor technologies has created rich datasets that can be used to run algorithms (Epochai, 2022). This development can be attributed to the rising frequency of research publications, as past data on building performance studies in North America, Europe, and the Asia Pacific regions can be utilized in model training and validation. Second, the computational power required to perform complex neural network operations has become significantly more affordable and widespread, with the cost of Graphics Processing Unit (GPU) hardware decreasing by 60 percent over the last five years and performance improving ten times (Epochai, 2022). Third, the growing focus on sustainable building processes and building regulatory standards has generated a large market base for precision in energy prediction equipment. The market for building energy management systems is estimated to reach $19.9 billion by 2030 (Strategicmarketresearch, 2025).

The building energy consumption includes various and interdependent thermal loads such as space heating load (HL), space cooling load (CL), domestic hot water heating, and auxiliary electrical loads due to lighting and equipment. The creation of precise predictive models of both HL and CL, in particular, can be considered among the most crucial steps in attaining energy efficiency in building operation and designing an HVAC system (Chaganti et al., 2022; Xu, 2024). Conventional modeling methods, including random forest (RF), decision tree (DT), support vector machine (SVM), and multiple linear regression, perform reasonably well when a single output is required; however, they fail to resolve the underlying correlations and shared physical motivations between HL and CL. Recent comparative studies indicate that the models inform design choices, optimize system size, and allow real-time control measures that reduce energy use (Billanes et al., 2025). Moreover, these novel predictive analytics support policy-making and code enhancements to yield a quantitative understanding of energy performance linkages and utilize evidence-grounded regulations (Long, 2023).

Although ML and DL techniques have become more widely applied to replace traditional statistical models in the construction of energy analysis modules, most previous studies have utilized linear or tree-based methods that are well-suited to address nonlinear and short-range interactions among key variables (e.g., relative compactness, surface area, roof area, etc.) of the building envelope (Ardabili et al., 2022). Furthermore, a significant portion of current research focuses on either HL or CL prediction alone, without considering the existing thermodynamic interrelations and structural regularities between these two thermal requirements, which exhibit non-zero correlation coefficients. Moreover, previous studies have mainly employed traditional ML algorithms, such as DT, RF (Tiwari et al., 2024), support vector regression (SVR), and light gradient boosting machine (LightGBM) (Guo et al., 2023) to forecast HL and CL without systematically comparing two DL algorithms, such as a convolutional neural network (CNN) and an artificial neural network (ANN).

This study addresses these essential gaps by constructing and validating strong predictive models with systematic cross-validation of two advanced DL architectures, shared-trunk ANN and CNN, using a k-fold cross-validation method. CNN and ANN hyperparameters were carefully selected to build predictive models that can predict HL and CL in residential buildings. This study demonstrates the high performance and accuracy of neural-network-based solutions, notably the CNN algorithm, as the foundation of next-generation building energy management systems. These systems can achieve prediction accuracy exceeding R^2, with values greater than 0.99, thereby optimizing energy usage and ensuring occupant comfort.

2. Data and methods

2.1. Data collection and processing

2.1.1. Dataset description

As detailed in Table 1, the variables collectively characterize the geometric, structural, and thermal properties of the building envelope that significantly influence its energy performance. The dataset represents simulated residential buildings located in Athens, Greece, over a year. These data were collected from 768 diverse residential buildings using Ecotect, with inputs including surface area, relative compactness, wall area, overall height, roof area, glazing area, orientation, and glazing area distribution, and outputs of HL and CL as reported by Tsanas and Xifara (Tsanas & Xifara, 2012). For instance, the relative compactness variable exhibits values ranging from 0.6 to 1.0 with a mean of 0.8 and a standard deviation of 0.1, indicating moderate variation in building form factors. Equally important, the surface area, a key component in residential buildings, ranges substantially from 514.5 to 808.5 m^2, with a mean and standard deviation of 671.7 m^2 and 88.1 m^2, respectively. These results reflect diverse building scales within the dataset. Wall area varies from 245.0 to 416.5 m^2 with a mean of 318.5 m^2 and a standard deviation of 43.6. In comparison, the roof area ranges from 110.3 to 220.5 m^2, implying a mean value of 176.6 m^2, with a standard deviation of 45.2.

Similarly, our statistical analysis reveals that the geometric characteristics span an overall height range of 3.5 to 7.0 m, with a mean of 5.3 m and a standard deviation of 1.8, representing typical residential building heights. Moreover, glazing area ranges from 0.0 to 0.4 with a mean of 0.2 and a standard deviation of 0.1. In contrast, glazing area distribution varies from 0.0 to 5.0 with a mean and standard deviation of 2.8 and 1.6, respectively, indicating diverse fenestration configurations. The average HL and CL are 22.3 and 24.6 kWh/m^2, respectively.

2.1.2. Data preprocessing and normalization

The dataset was first examined for data quality issues. Analysis confirmed that all 768 observations contained complete data with no missing values across the eight input features and two output variables, eliminating the need for imputation techniques. A preliminary outlier analysis was conducted using the interquartile range (IQR) method. Given that the dataset represents physically simulated buildings with controlled parameter ranges, all observations were retained as they represented legitimate combinations of building design parameters rather than measurement errors or data collection anomalies.

As illustrated in Table 1, substantial differences in scale exist across input features (e.g., relative compactness, ranging from 0.62 to 0.98, versus surface area, ranging from 514.5 to 808.5 m^2). To ensure that features with larger numeric ranges do not dominate the gradient descent optimization process during neural network training, min-max normalization was applied to transform all input features to a uniform scale of [0, 1] as illustrated in Equation (1).

Table 1
Statistical parameters of input and output variables.

Features	No. of possible	Minimum	Maximum	Standard deviation	Mean	Skewness	Unit
Relative compactness	12	0.6	1.0	0.1	0.8	0.5	dimensionless
Surface area	12	514.5	808.5	88.1	671.7	−0.1	m^2
Wall area	7	245.0	416.5	43.6	318.5	0.5	m^2
Roof area	4	110.3	220.5	45.2	176.6	−0.2	m^2
Overall height	2	3.5	7.0	1.8	5.3	0.0	m
Orientation	4	2.0	5.0	1.1	3.5	0.0	dimensionless (code) (2, 3, 4, 5)
Glazing area	4	0.0	0.5	0.1	0.2	−0.1	dimensionless (ratio)
Glazing area distribution	6	0.0	5.0	1.6	2.8	−0.1	dimensionless (code) (0-5)
Heating load	586	6.0	43.1	10.1	22.3	0.4	kWh/m^2
Cooling load	636	10.9	48.0	9.5	24.6	0.4	kWh/m^2

$$x'_{normalized} = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

where x' represents the normalized value, x is the original feature value, and x_{min} and x_{max} are the minimum and maximum values for each feature.

Two variables in the dataset (orientation and glazing area distribution) are categorical in nature, encoded as integer codes representing

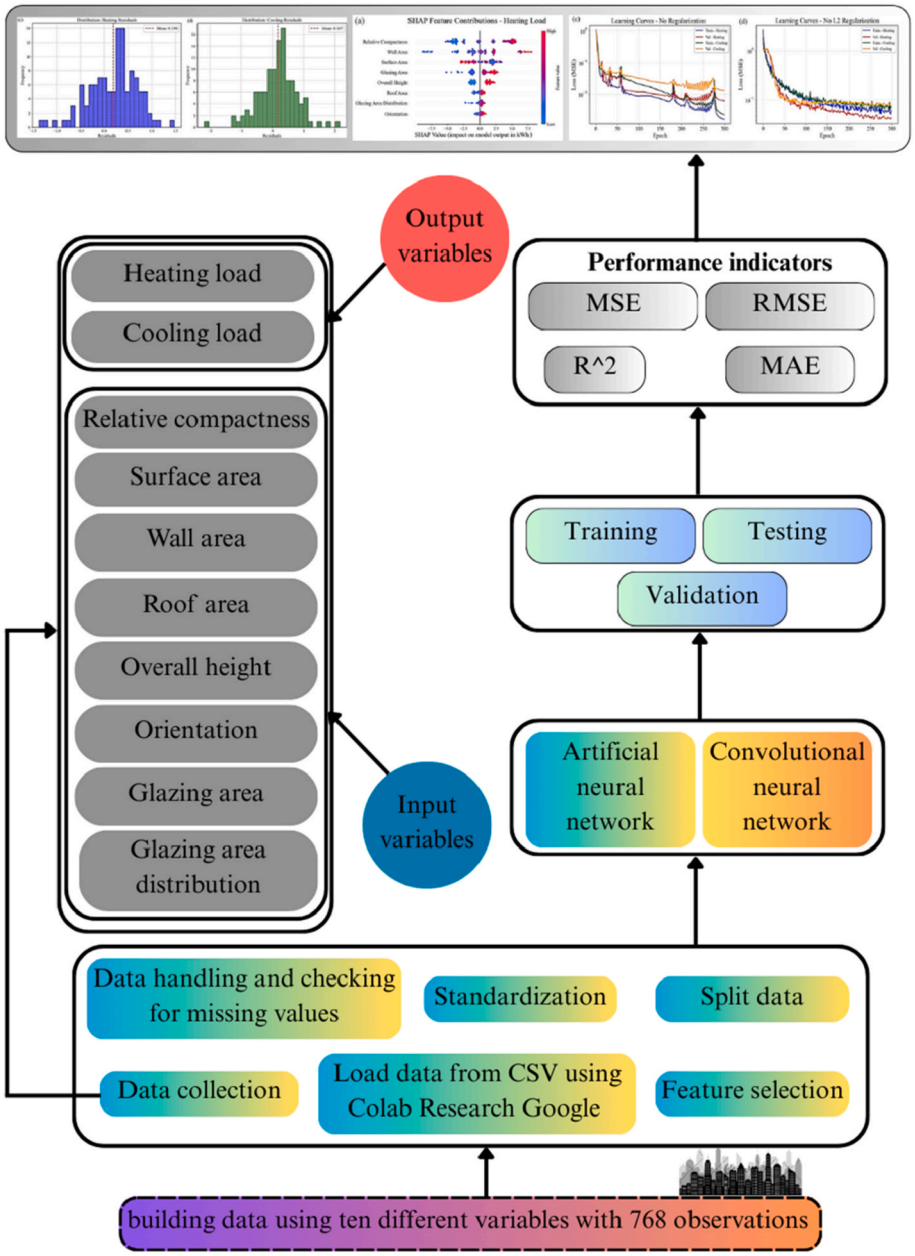

Fig. 1. Data collection and preprocessing flowchart.

discrete categories (orientation: (2, 3, 4, 5) representing North, East, South, West orientations; glazing area distribution: (0, 1, 2, 3, 4, 5) representing different glazing distribution patterns). Rather than applying one-hot encoding, which would expand the feature space from 8 to 14 dimensions, these categorical variables were treated as ordinal and included in the min-max normalization process. This approach maintains a compact input representation suitable for the ANN and CNN architecture, which processes features sequentially.

The output variables (HL, ranging from 6.01 to 43.10 kWh/m^2, and CL, ranging from 10.90 to 48.03 kWh/m^2) were not normalized. The models were trained to directly predict actual thermal load values in kWh/m^2, facilitating straightforward interpretation and practical deployment without requiring inverse transformation of predictions.

To develop and evaluate the predictive models, the dataset was split into training (70%, n = 537), validation (15%, n = 115), and test sets (15%, n = 116). Hyperparameter optimization was performed systematically to identify the optimal configuration for both ANN and CNN models. A two-stage approach was employed, namely, a coarse-grained random search to broadly explore the hyperparameter space, followed by a fine-grained grid search to refine promising configurations. The process and methodology used to develop ANN and CNN models are illustrated in Fig. 1.

2.2. Artificial intelligence

AI can be defined as the field of computer science focused on creating systems that perform tasks typically requiring human intelligence, such as reasoning, learning, perception, and decision-making (Mirindi et al., 2025; Mirindi et al., 2025; Mirindi & Mirindi, 2025). As illustrated in Fig. 2, this concept is not new; its evolution started in the 1950s (Giannini & Bowen, 2024). Indeed, AI enables machines to process data, recognize patterns, make predictions, and adapt to changing circumstances with unprecedented speed and accuracy (Mirindi et al., 2025). Within AI, machine learning (ML) algorithms such as linear regression, decision tree (DT) (Mirindi et al., 2025; Myles et al., 2004), random forest (RF), support vector machine (SVM) (Sinkhonde et al., 2025), k-nearest neighbors (k-NN) (Chirici et al., 2016), naive Bayes (Frank et al., 2000), and k-means clustering (Likas et al., 2003) are widely used

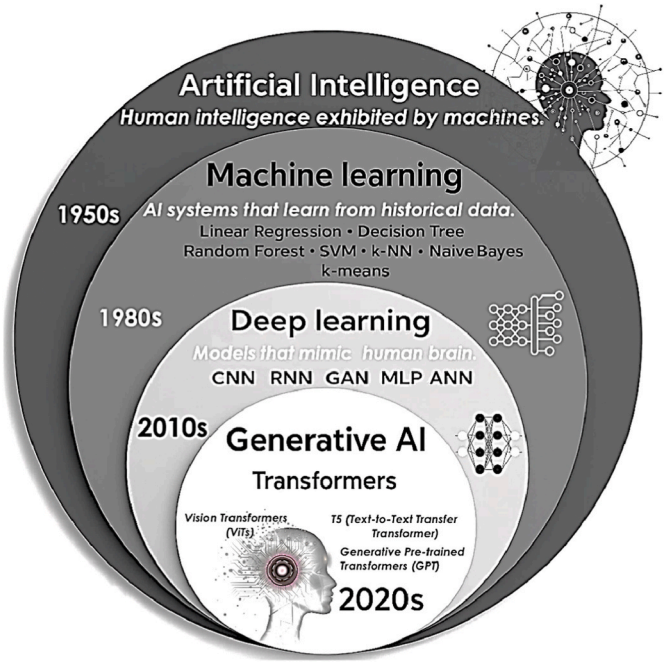

Fig. 2. Typical representation of artificial intelligence, machine learning, deep learning, and generative artificial intelligence.

in architecture and construction to allow computers to learn from data and make autonomous decisions. ML algorithms operate by being trained on historical data, so they can identify complex relationships and make predictions or classifications without explicit programming (Mirindi et al., 2025).

DL, a subset of ML, is characterized by the use of an artificial neural network (ANN) with hidden layers that learn intricate patterns automatically from large volumes of data (Mirindi et al., 2024). DL models are especially adept at uncovering abstract features and nonlinear relationships in high-dimensional data. The most popular DL algorithms include the convolutional neural networks (CNN) for image and pattern recognition (Traore et al., 2018), the recurrent neural networks (RNN) for sequential data (Weerakody et al., 2021), the generative adversarial network (GAN) for data synthesis (Liu et al., 2022), and the multilayer perceptron (MLP) for classification and regression tasks (Murtagh, 1991). Beyond that, these algorithms have advanced various fields ranging from natural language processing to autonomous medicine. Therefore, DL can be considered a powerful tool for modern AI applications. Furthermore, generative artificial intelligence (GenAI), which emerged since 2020 with the COVID-19 pandemic, a subset of DL, represents a transformative branch of AI that leverages advanced ML and DL models to autonomously generate new and original content based on learned data patterns(Mirindi et al., 2025). GenAI utilizes transformers, including generative pre-trained transformers (GPT) (Kamnis, 2023), vision transformers (ViT) (Hatamizadeh et al., 2023), and text-to-text transfer transformers (T5) (Rodriguez-Torrealba et al., 2022). Unlike traditional AI, which focuses primarily on tasks such as prediction or classification, GenAI utilizes generative models (typically trained on vast datasets) and transformers to create new content, including text, images, audio, code, and video, in response to user prompts.

2.2.1. Artificial neural network

An ANN is a computational model designed to simulate how the human brain analyzes and processes information. In the supplementary information (SI1), the diagram illustrates a typical representation of the human brain (Ghritlahre & Prasad, 2018). ANN consists of interconnected nodes (also called neurons), which are subdivided into layers. An ANN model has an input layer, one or more hidden layers, and an output layer.

Each neuron receives signals from other neurons or external sources, processes these signals using mathematical functions, and passes the output to the next layer, as demonstrated in SI2. The strength of connections between neurons (called weights) can be adjusted during training to improve the network's performance in tasks such as classification, regression, or pattern recognition. Each neuron computes an affine transform $z = W \cdot x + b$ followed by a nonlinearity, denoted as $= \varphi(z)$, with ReLU ($\varphi(z) = \max(0, z)$) being used in this instance. This process culminates in a task-specific output layer.

In this study, Fig. 3 illustrates the ANN model employed. The network comprises three hidden layers with sizes of 256→128→64 responsible for learning common thermal representations from the height input building features. Furthermore, the network branches into two linear heads, each dedicated to predicting HL or CL. The hidden layers have sizes of 32→16 neurons per task, culminating in individual output layers for the final predictions. Concurrently, to ensure stable optimization and improved generalization, it is essential to mention that several regularization and normalization techniques are employed. We applied batch normalization after each trunk dense layer to mitigate internal covariate shift. Also, L2 weight decay, with a regularization strength of $\lambda = 1e^{-3}$, constrains the model's weights. Additionally, dropout is applied at rates of 0.3, 0.3, and 0.2 in the trunk layers, and at a rate of 0.2 in each of the two heads.

Table 2 summarizes the ANN hyperparameters selected and the training settings, where the Adam optimizer is used with a learning rate of 1e^{-3} and training is conducted with mini-batches of size 32. The

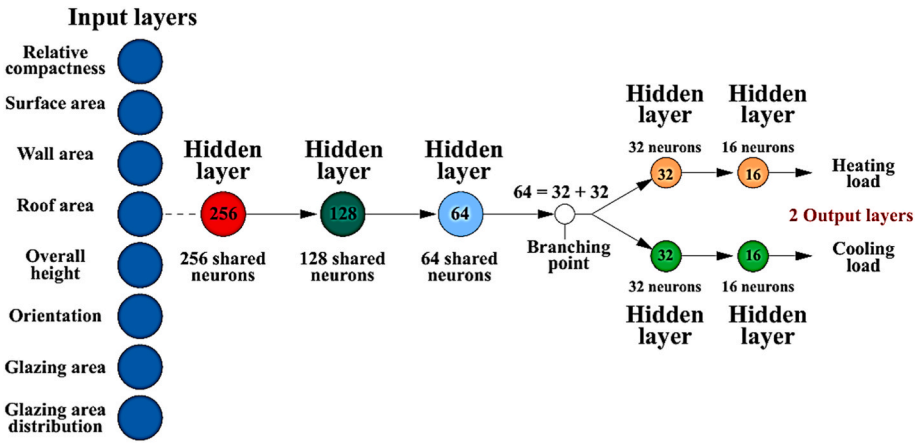

Fig. 3. Artificial neural network architecture model.

Table 2
Artificial neural network hyperparameters.

Hyperparameter	Value
Hidden layer sizes	Trunk: 256 → 128 → 64; Branching point: 32 → 16 per task Three hidden layers in shared trunk, then two Dense layers in each branch (heating/cooling)
Activation functions	ReLU in all hidden Dense layers; Linear for outputs activation = 'relu' in hidden; activation = 'linear' in final Dense (1) per task
Regularization and normalization	L2 weight decay = $1e^{-3}$ on trunk Dense layers; Batch Normalization after each trunk Dense; Dropout: 0.3, 0.3, 0.2 in trunk; 0.2 in each head
Optimizer & learning rate	Adam with learning rate = $1e^{-3}$
Training schedule and callbacks	Epochs = 300, batch size = 32; Early Stopping (patience = 50, restore best weights = True); Reduce LR On Plateau (factor = 0.5, patience = 25, min_lr = $1e^{-7}$); Model Check point (save best only = True)

model trains for up to 300 epochs, incorporating Early Stopping with a patience of 50 epochs and restoring the best weights found during training. A ReduceLROnPlateau scheduler is also used, reducing the learning rate by a factor of 0.5 if the validation loss plateaus for 25 epochs, with a minimum learning rate of $1e^{-7}$. Model checkpointing is also implemented. Formally, the forward pass for a layer l is expressed as:

$$a^{(l)} = \varphi\left(W^{(l)}a^{(l-1)} + b^{(l)}\right) \tag{2}$$

where $\varphi(x) = \max(0, x)$ for Rectified Linear Unit (ReLU).

The two output layers, designed for regression, utilize linear activation. The network's parameters $\theta = \{W, b\}$ are learned by minimizing a total loss $L = L_{HL} + L_{CL}$ which is the sum of the per-target losses (e.g., mean squared error (MSE)). This minimization is performed using gradient-based updates, $\theta \leftarrow \theta - \eta\nabla_\theta L$.

2.2.2. Convolutional neural network

A CNN is a DL model particularly suited for analyzing grid-structured data, such as images, which can represent aspects of building geometry or environmental conditions affecting energy loads. Visual Geometry Group (Yaseliani et al., 2022), VGG-16, a prominent 16-layer deep CNN architecture, exemplifies this approach through its systematic design. Indeed, we construct five convolutional blocks (Conv-1 through Conv-5) as depicted in SI3, utilizing small 3 × 3 filters. This strategy was followed by three fully connected layers (FC-6, FC-7, FC-8). By providing an illustrative example of a building facade from Morgan State University in the United States, we can see that these blocks progressively extract hierarchical features from the input image.

In parallel, CNNs effectively extract features by employing small, trainable filters (kernels) that scan across the input data, sharing weights to identify patterns regardless of their precise location, a property known as translation equivariance. These stacked convolutional layers can progressively identify features, ranging from basic edges to complex textures and objects, in the input data. Indeed, pooling layers then reduce the spatial resolution. Therefore, they improve invariance to minor input variations and manage computational cost. As a result, a final layer, typically a fully connected dense layer or global average pooling, maps these extracted features to the desired outputs, such as predictions for HL or CL loads.

The computation of a CNN proceeds layer by layer. For an input tensor $X \in \mathbb{R}^{H\times W\times C_{in}}$ representing the input data and a kernel $K \in \mathbb{R}^{h\times w\times C_{in}\times C_{out}}$ with stride s and padding p, each output activation y(i,j, c_{out}) is calculated as:

$$y(i,j,c_{out}) = \varphi\left(\sum_{c=1}^{C_{in}}\sum_{u=0}^{h-1}\sum_{v=0}^{w-1} K(u,v,c,c_{out})\cdot X(i\cdot s+u-p, j\cdot s+v-p, c) + b_{c_{out}}\right) \tag{3}$$

where φ is a nonlinearity like the Rectified Linear Unit (ReLU), defined as $\varphi(z) = \max(0, z)$. The spatial dimensions of the output feature map are:

$$H_{out} = [(H + 2p - h)/s] + 1 \tag{4}$$

$$W_{out} = [(W + 2p - w)/s] + 1 \tag{5}$$

Pooling operations, such as an $r \times r$ max-pool with stride r, further process the feature maps, for instance, by computing $Y(i, j, c) = \max_{0\le u,v<r} X(i\cdot r + u, j\cdot r + v, c)$. Our CNN parameters, comprising all kernel weights and biases, are learned by minimizing a loss function L through an optimization algorithm, such as gradient descent. The update rule is typically $\theta \leftarrow \theta - \eta\cdot\partial L/\partial\theta$, where θ represents the parameters and η indicates the learning rate. The necessary gradients $\partial L/\partial\theta$ are efficiently calculated using the backpropagation algorithm. Simultaneously, for a kernel, backpropagation involves convolving upstream errors with the input and accumulating the gradients across spatial positions due to the shared weights. Our structure ensures that the model's predictions for HL and CL improve over iterations.

As depicted in Fig. 4, the CNN architecture model begins with a stem Conv1D layer that projects the raw feature sequence into 64 channels, followed by three residual blocks that preserve sequence length with the same padding while increasing the network's effective receptive field through progressive dilation. Specifically, the dilation schedule of 1, 2, and 4 across the three residual stages yields effective receptive fields of approximately 3, 5, and 9 on a length-8 sequence. This structure enables

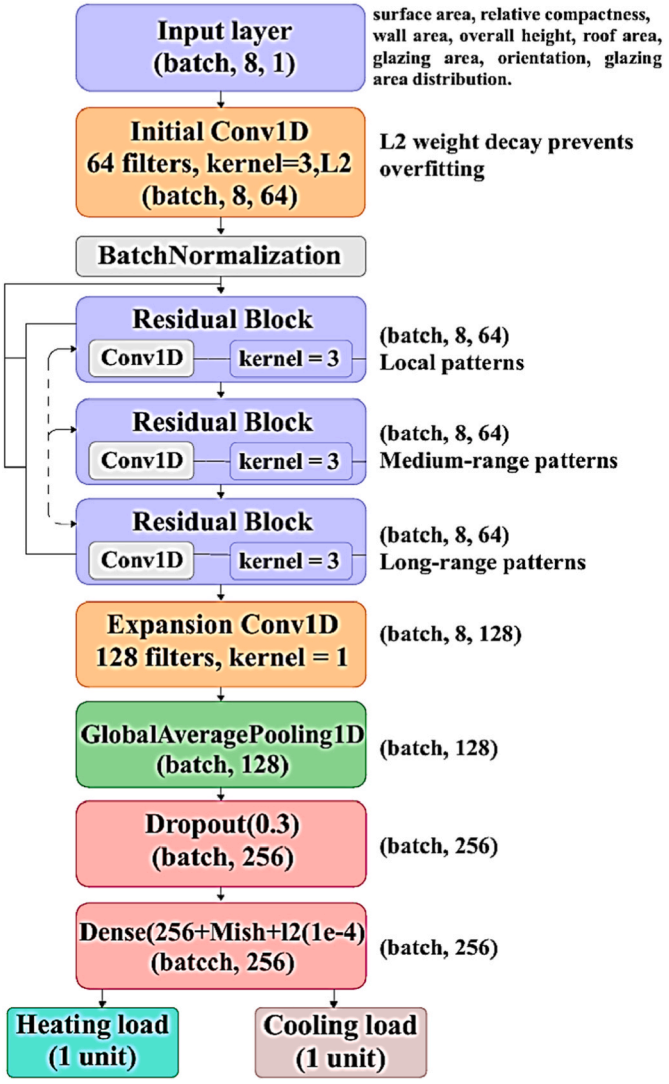

Fig. 4. Convolutional neural network architecture model.

the model to capture local and near-global dependencies among building geometry (e.g., relative compactness, wall and roof areas, etc.) and envelope parameters (e.g., glazing levels and distribution). Each convolutional operation is normalized by batch normalization and activated with Mish, improving gradient flow and representation smoothness. Here, residual skip connections help stabilize the extraction of deeper features and mitigate the vanishing gradient problem. Additionally, after the residual stack, expanding the convolutional block Conv1D with 1×1 kernels lifts the channel dimension to 128 for channel mixing, without altering the sequence length. GlobalAveragePooling1D then aggregates features across the sequence to produce a compact, order-invariant representation that feeds a dense prediction head. Regularization is applied via L2 weight decay ($1e^{-4}$) on Conv1D and hidden layers, and dropout (0.3) is used between the two hidden layers, each with 256 and 128 neurons, to reduce overfitting.

Table 3 summarizes the CNN hyperparameter configuration. This includes 64 filters in the stem and each residual block, kernel size of three throughout the residual path, Batch Normalization after every convolution, and Adam optimizer with a base learning rate of $1e^{-3}$.

Table 3
Convolutional neural network hyperparameter.

Hyperparameter	Value and details
Convolution filters per stage	Initial Conv1D: 64 filters; Residual Blocks: 64 filters each (3 stages); Expansion Conv1D: 128 filters (1×1 kernel)
Kernel sizes	3 for all Conv1D layers inside residual blocks and the initial Conv1D; 1 for the expansion Conv1D layer (channel mixing)
Dilation rates	Residual Block 1: dilation = 1; Residual Block 2: dilation = 2; Residual Block 3: dilation = 4 (effective receptive fields: 3, 5, 9 on a length-8 sequence)
Padding	same for all Conv1D layers (sequence length preserved)
Activations	Mish after each Batch Normalization in the stem and inside residual blocks; Dense (256) and Dense (128) also use Mish; final outputs use Linear
Normalization	Batch Normalization after every Conv1D operation (stem and inside residual blocks)
Regularization	L2 weight decay $1e^{-4}$ on all Conv1D and Dense layers that specify kernel_regularizer; Dropout (0.3) between Dense (256) and Dense (128)
Global pooling	GlobalAveragePooling1D (aggregates across sequence length before dense head)
Optimizer	Adam; base learning_rate = $1e^{-3}$ (compatible with optional cyclical LR callback if used)

2.3. Evaluation of model performance

Standard errors and fit measures can be considered to evaluate the predictive performance of the ANN and CNN models (Mirindi et al., 2025). This assessment includes mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and the coefficient of determination R^2. Indeed, R^2 measures the proportion of variance in the observed values explained by the model, and it can be computed following Equation (6). MSE computes the average squared difference between predictions and observations as given by Equation (7). By squaring the residuals, MSE places greater weight on larger errors, making it sensitive to outliers. It is advantageous when large deviations are more critical to penalize than smaller ones. RMSE, as computed in Equation (8), extends this concept by taking the square root of MSE. Subsequently, because the square root returns the error to the same unit as the original target variable, RMSE provides an interpretable measure of the model's typical prediction error magnitude. Additionally, a smaller RMSE indicates that predictions tend to be closer to the observed values on average. MAE, as highlighted in Equation (9), instead evaluates the mean of the absolute deviations between predictions and observations. Unlike MSE and RMSE, MAE treats all errors linearly and is therefore helpful in mitigating the influence of extreme outliers. Taking this further, it provides a straightforward interpretation: the average absolute difference between the predicted and observed values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{6}$$

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{7}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{8}$$

$$MAE = \frac{1}{n} \sum_{i=1}^{n} | y_i - \hat{y}_i | \tag{9}$$

where each of the error metrics captures how far the predicted values $\hat{y}_i$ deviate from the actual observations y_i over n samples, while $\bar{y}$ is the mean of the observed values.

2.4. Taylor diagram

The Taylor diagram was introduced by Karl E. Taylor (Sinkhonde et al., 2025). It serves as a comprehensive statistical visualization tool

that simultaneously illustrates multiple performance metric models on a single polar coordinate plot. It allows us to demonstrate efficient comparison of model predictions against observed data. Moreover, this diagram incorporates three key statistical measures, including the correlation coefficient, standard deviation, and RMSE. Utilizing this diagram, we can provide a holistic assessment of model performance through geometric relationships between R^2, RMSE, and the standard deviation. Therefore, this diagram relies on the fundamental relationship between statistical metrics. It is derived from the law of cosines. This implies that we can mathematically define this relationship for a model with predictions y_i' and observations y_i, the correlation coefficient r as:

$$r = \frac{\sum_{i=1}^{n}(y - y_i')(y_i - y')}{\sqrt{\sum_{i=1}^{n}(y - y_i')^2}\sqrt{\sum_{i=1}^{n}(y_i - y')^2}} \tag{10}$$

The normalized standard deviation of predictions σ_p and observations σ_o are defined as:

$$\sigma_p = \sqrt{\sum_{i=1}^{n}(y - y_i')^2} \tag{11}$$

$$\sigma_o = \sqrt{\sum_{i=1}^{n}(y_i - y')^2} \tag{12}$$

The key geometric relationship underlying the Taylor diagram is expressed through the centered root mean squared error (CRMSE) using Equation (13):

$$CRMSE^2 = \sigma_p^2 + \sigma_o^2 - 2\sigma_p\sigma_o r \tag{13}$$

In the Taylor diagram representation, the radial distance from the origin represents the standard deviation of predictions, the angular position indicates the correlation coefficient (with angles measured from the positive x-axis), and the distance from the reference point (representing perfect agreement) corresponds to the CRMSE. The reference point, typically located at coordinates $(\sigma_o, 0)$, represents the observed data statistics.

2.5. Cross-validation

Cross-validation represents a fundamental statistical resampling technique designed to assess model generalization capability and mitigate overfitting by systematically partitioning the available dataset into training and validation subsets (Jiang & Wang, 2017; Zhang & Liu, 2023). Specifically, we have utilized the k-fold cross-validation methodology, which divides the dataset into approximately equal-sized k folds. Indeed, each fold serves sequentially as a validation set while the remaining $k-1$ folds constitute the training set. In parallel, this process ensures that all data points participate in both training (k-1 folds) and validation (1 fold) phases across different iterations. It provides an important performance estimate.

The mathematical framework for k-fold cross-validation involves partitioning the dataset D with n samples into k disjoint subsets: $D = D_1 \cup D_2 \cup \ldots \cup D_k$, where $|D_i| \approx \frac{n}{k}$. For each iteration i, the model is trained on $D_{train}^{(i)} = D \backslash D_i$ and validated on $D_{val}^{(i)} = D \backslash D_i$. The overall performance metric M is computed as the average across all folds as follows:

$$M_{CV} = \frac{1}{k}\sum_{i=1}^{k} M^{(i)} \tag{14}$$

where $M^{(i)}$ represents the performance metric (R^2, RMSE, MAE) evaluated on fold i.

The coefficient of variation (CV) quantifies the relative variability of performance across folds and can be mathematically computed as illustrated in Equation (15).

$$CV = \frac{\mu_M}{\sigma_M} \times 100\% \tag{15}$$

where μ_M and σ_M are the standard deviation and mean of the performance metric across all folds, respectively.

3. Results and discussion

Fig. 5 illustrates the Pearson correlation matrix for the eight input variables and two target outputs (HL and CL). Our results show that inverse and linear dependencies emerge among several input variables, reflecting the inherent geometric constraints in the dataset. For instance, the relative compactness and surface area are nearly perfectly negatively correlated (r = −0.992), while roof area and overall height also demonstrate a strong negative relationship (r = −0.973). In addition, HL exhibits substantial positive correlation with overall height (r = 0.889) and strong negative correlation with the roof area (r = −0.862), the surface area (r = −0.673), and the building orientation (r = −0.003). On the other hand, CL displays a strong correlation with overall height (r = 0.896), relative compactness (r = 0.634), and wall area (r = 0.427), as well as an inverse correlation with roof area (r = −0.863) and surface area (r = −0.673). Fenestration variables, including glazing area and distribution, exhibit only weak associations with HL and CL, with correlation coefficients of 0.270 for glazing area and 0.237 for glazing area distribution, respectively. These results indicate modest solar-gain effects but far less influence than geometry-related parameters. Importantly, HL and CL are highly correlated (r = 0.976), with the exception of surface and roof area, which exhibit negative correlations, as well as HL and building orientation.

Moreover, as illustrated in Fig. 6, the ANN's predicted versus actual analysis for both HL and CL clusters tightly around the 45° line. These results show evidence of low bias and strong calibration. Additionally, the dataset points are distributed across the entire load range, with only mild dispersion at the extremes, a common phenomenon that occurs when data density decreases, and physical nonlinearities increase. This behavior is logical, given the reported training metrics, R^2 values of 0.9783 and 0.9768 for HL and CL, respectively. These results indicate that the shared-trunk, dual-head ANN extracts a proper latent representation that generalizes beyond the training set. Notably, CL prediction exhibits a slightly larger spread than heating at higher loads. This performance differential is due to the stronger dependence of CL on interactions between glazing, orientation, solar gain, and nonlinear effects, whereas HL is primarily driven by simpler conductive heat transfer through the building envelope.

Besides, the analysis shows that the ANN predictions remain stable across training, validation, and test splits, with no evident indication of overfitting. This consistency reflects the benefits of the regularization strategy (batch normalization, dropout, and weight decay), which help mitigate variance and ensure that the learned representation is not overly tuned to the training data. The marginally reduced accuracy for CL, relative to HL, echoes prior studies where CL is found to be more challenging to capture due to its dependence on fenestration and orientation variables, which often exhibit nonlinear and interaction-heavy effects (Amasyali & El-Gohary, 2018; Ekici & Aksoy, 2009).

It is essential to note that, despite these challenges, the ANN significantly enhances model performance compared to traditional regression approaches. They often fail to capture nonlinear dependencies in building load prediction (Neto & Fiorelli, 2008). The tight clustering of the ANN shows that DL can recover meaningful mappings from building geometry to thermal loads even with a small number of input descriptors.

Correspondingly, the CNN's scatter plots depicted in Fig. 7 are even tighter along the perfect line, with visibly reduced variance relative to the ANN plots. This result aligns with the substantially high R^2 values (>0.99) for the training, testing, and validation CNN models in

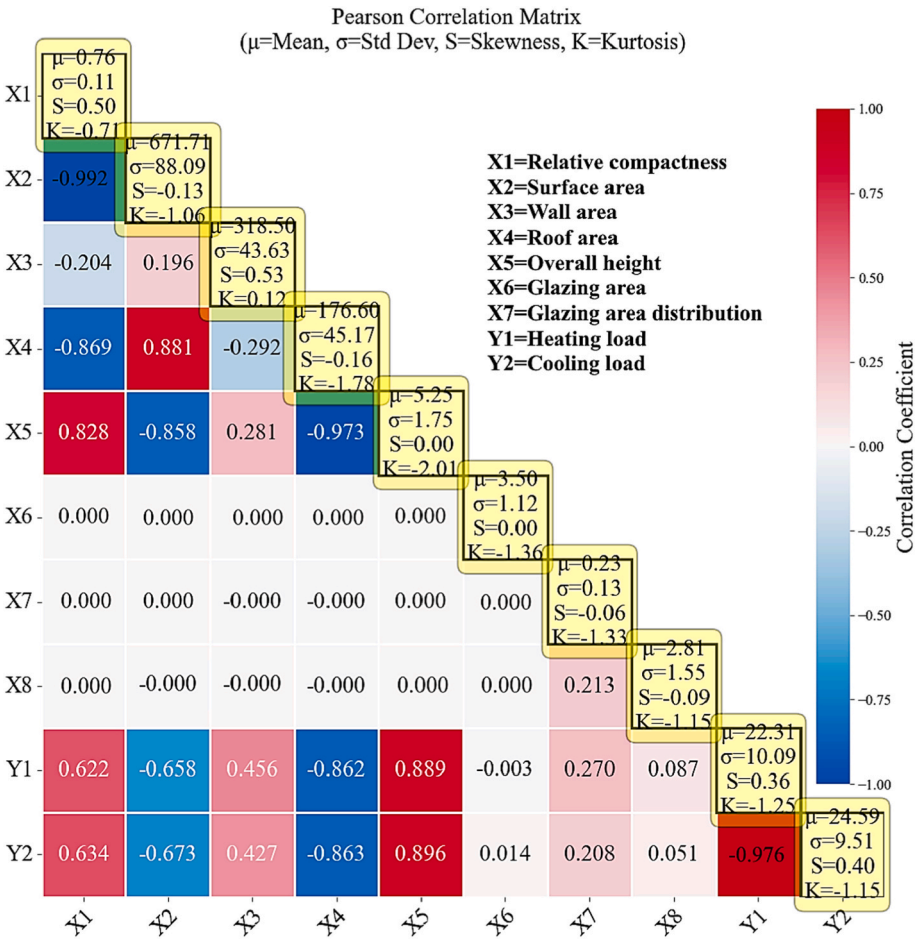

Fig. 5. Pearson correlation matrix.

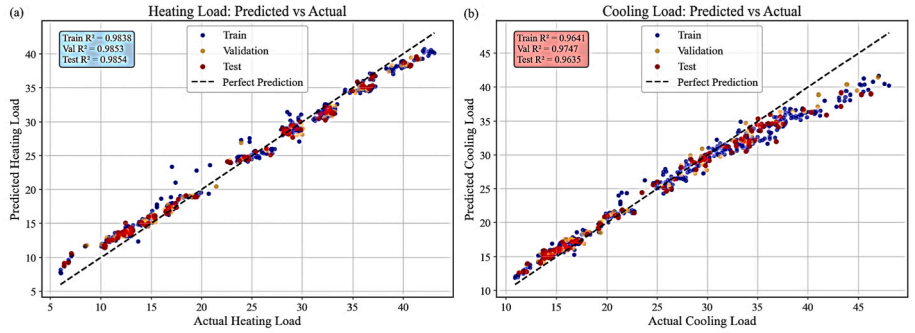

Fig. 6. Predicted versus actual model of ANN: (a) heating load, (b) cooling load.

predicting CL and HL, compared to those of the ANN models. Indeed, the CNN1D convolutions and dilated residual blocks capture localized feature interactions (e.g., coordinated variations of surface, wall, and roof areas with glazing) and multi-scale patterns that the ANN may not regularize as efficiently. Therefore, the result is improved fidelity across the entire operating range, particularly where coupled envelope effects predominate.

Both HL and CL predictions demonstrate near-perfect clustering along the perfect line. Moreover, these results show the CNN's ability to generalize without significant overfitting. On the other hand, we have noticed that the close overlap of points across splits suggests that weight sharing and residual connections effectively constrain variance while enabling flexible representation learning. The slightly larger spread in CL compared to HL remains visible, although it is reduced relative to the ANN. This approach supports findings that CL prediction remains inherently more challenging due to fenestration- and solar-gain-driven variability (Amasyali & El-Gohary, 2018; Ekici & Aksoy, 2009).

Compared with the ANN, the CNN improves accuracy by embedding inductive biases through convolutional filters. These filters exploit feature ordering and local dependencies in the building descriptors, enabling the network to learn short- and medium-range couplings (e.g., compactness with wall and roof area, glazing with orientation) more effectively. Similar improvements from CNN architectures have been reported in building load forecasting and energy time-series prediction, where convolution layers reduce residual variance and improve stability under unseen test data (Wang et al., 2020). In summary, the scatter plots

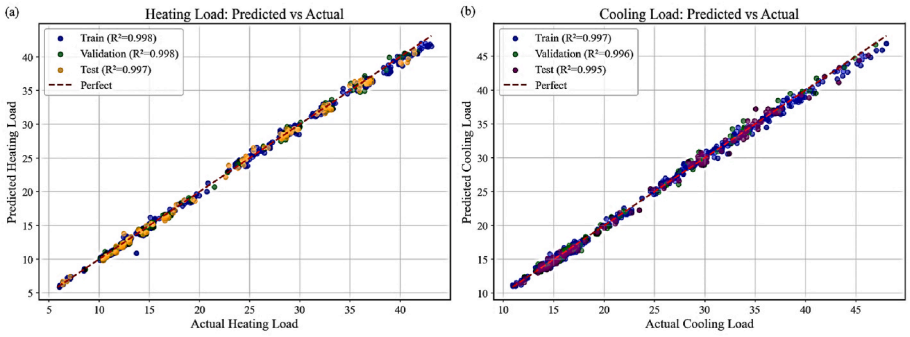

Fig. 7. Predicted versus actual model of CNN: (a) heating load, (b) cooling load.

in Fig. 7 demonstrate that CNNs achieve a superior fit relative to ANNs, providing near-linear predicted–actual alignment and lower error variance across datasets. As a result, this stability is especially valuable for practical applications, such as HVAC system sizing and compliance assessment, where even minor improvements in predictive accuracy can reduce design risks and enhance operational efficiency.

Notably, CNN models demonstrate superior predictive accuracy and residual behavior compared to ANN models (SI4). Fig. 8 provides evidence of this by revealing significantly improved homoscedastic residual patterns. Both HL and CL residuals (see Fig. 8(a) and (b)) exhibit substantially reduced variance across the prediction range. Correspondingly, the residuals maintain consistent scatter around 0 with minimal heteroscedasticity. Therefore, these results indicate that CNN's architecture can capture nonlinear relationships without systematic bias. Notably, the residual distributions, as depicted in Fig. 8(a) and (b), approach ideal Gaussian characteristics with near-zero means, approximately 0.190 and 0.167 for HL and CL, respectively. They demonstrate unbiased predictions and validate the model's statistical assumptions.

The training dynamics presented in Fig. 9 illustrate exceptional convergence stability of the CNN compared to ANN models (see SI5). We can substantiate this evidence by considering, for instance, the model loss in Fig. 9(a), which exhibits a rapid initial decay with perfect alignment between the training and validation curves in the middle, between 0 and 50 epochs. This result indicates minimal overfitting despite the model's complexity. Expanding on this, the MAE trajectories for both HL (Fig. 9(b)) and CL (Fig. 9(c)) demonstrate smooth behavior lines, but with a remarkably high level of turbulence in amplitude occurring between 250 epochs and 300 epochs. At the same time, it presents a monotonic convergence to remarkably low error values. Furthermore, the per-target MSE analysis, as illustrated in Fig. 9(d), confirms balanced learning across both outputs, with convergence occurring within the first 50 epochs.

These findings are reflected in the metrics data presented in Table 4, which compares the performance of both ANN and CNN models. The R^2 values for the ANN models range from 0.96 to 0.98 across all dataset splits, while those for the CNN models exceed 0.99 in all cases. Particularly noteworthy are the dramatically reduced error magnitudes of CNN models with the test RMSE values of 0.5319 and 0.6476 for HL and CL, respectively, representing an improvement in predictive precision. Substantially, the consistent performance across validation (R^2 values of 0.9975 for HL and 0.9963 for CL) confirms the model's reliability for practical deployment in building energy management applications. These performance metrics results can be compared to a study by Guo et al. (Guo et al., 2023), in which algorithms such as SVR, RF, and LightGBM were used to predict the energy performance of buildings. For instance, as indicated in Table 5, SVR achieved R^2 values of only 0.8710

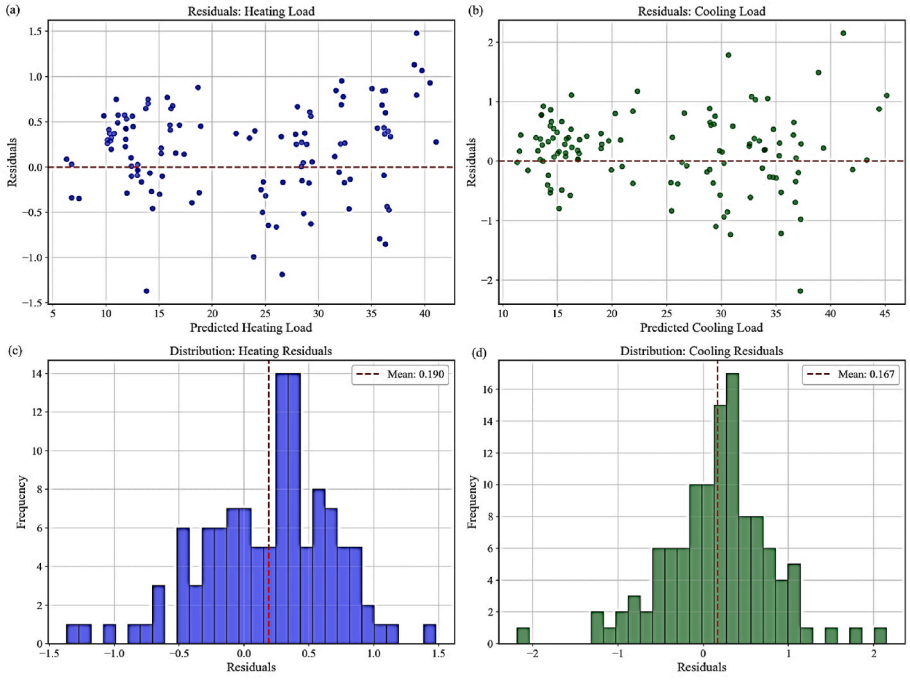

Fig. 8. CNN residuals versus predicted model of (a) heating load, and (b) cooling load; with residuals distribution of (c) heating load, and (d) cooling load.

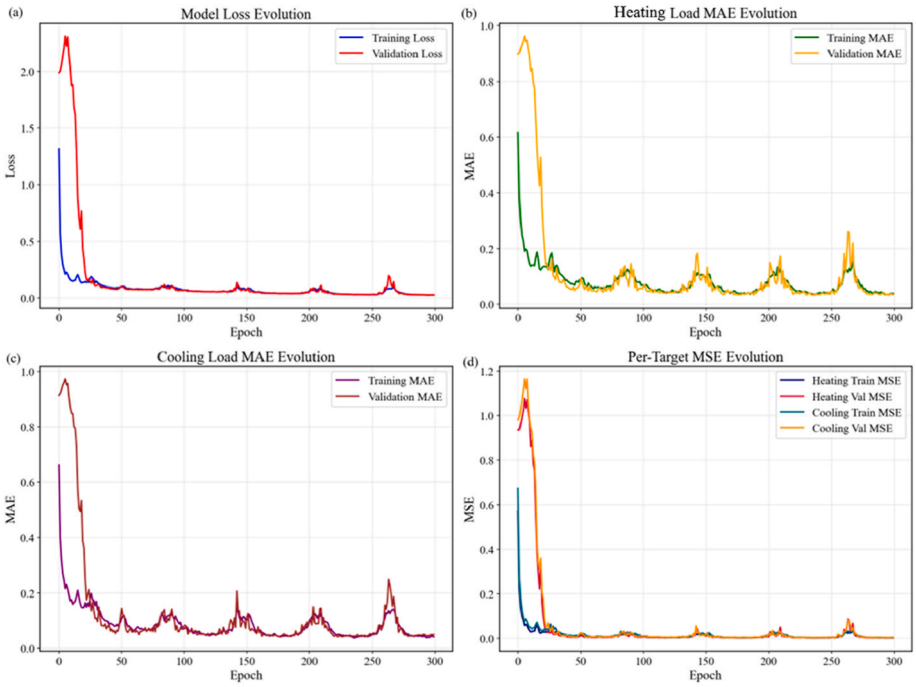

Fig. 9. CNN (a) model loss, (b) MAE heating load, (c) MAE cooling load, and (d) per-target MSE.

Table 4
Performance metric of the CNN and ANN models.

Model	Dataset	Target	MSE	RMSE	MAE	R^2
CNN	Training	Heating Load	0.1898	0.4357	0.3182	0.9982
CNN	Training	Cooling Load	0.2302	0.4797	0.3659	0.9975
CNN	Validation	Heating Load	0.2211	0.4737	0.3659	0.9975
CNN	Validation	Cooling Load	0.3160	0.5621	0.4494	0.9963
CNN	Test	Heating Load	0.2830	0.5319	0.4389	0.9973
CNN	Test	Cooling Load	0.4194	0.6476	0.4974	0.9953
ANN	Training	Heating Load	2.1203	1.4561	1.2052	0.9783
ANN	Training	Cooling Load	2.0478	1.4310	1.1016	0.9768
ANN	Validation	Heating Load	2.5669	1.6022	1.3441	0.9766
ANN	Validation	Cooling Load	3.0684	1.7517	1.2065	0.9670
ANN	Test	Heating Load	2.5265	1.5895	1.3331	0.9758
ANN	Test	Cooling Load	3.5380	1.8809	1.3384	0.9618

for HL and 0.6498 for CL, while RF improved these metrics to 0.9835 and 0.7941, respectively. These algorithms demonstrate low accuracy in predicting energy building compared to the CNN algorithm. However, when compared to the four hybrid models derived from the LightGBM developed in their study, the tree-structured parzen estimator TPE-LightGBM model, for instance, proved to be the best-performing algorithm in their study, with R^2 values of 0.9981 for HL and 0.9924 for CL. Meanwhile, these results remain comparable to our CNN performance but are still higher (R^2 values of 0.9982 for HL and 0.9975 for CL).

To further illustrate the performance metrics of ANN and CNN models, the Taylor diagram (Fig. 10) provides a comprehensive statistical comparison of the performance of both the ANN and CNN models across all dataset splits. By further comparing the metric performances of the two models, the diagram reveals that the CNN models consistently cluster closer to the reference point (the perfect model) compared to the ANN models. On a similar note, we can definitely conclude that CNNs exhibit higher correlation coefficients and lower normalized standard deviations, with RMSE contours confirming substantially reduced prediction errors compared to those of ANN. Therefore, the next step is to analyze the CNN model's performance through k-fold cross-validation, after presenting the feature importance and the SHAP analysis.

Fig. 11 reveals differential error sensitivity patterns between HL and CL predictions. For HL (Fig. 11 (a)), the roof area exhibits the strongest correlation with residuals (approximately 0.25), followed by overall height and wall area. These results suggest that the geometric parameters dominate prediction errors. On the other hand, CL errors (Fig. 11 (b)) indicate that the glazing area is the primary contributor to the error (approximately 0.12), consistent with the increased complexity of modeling solar gain effects. It is useful to mention that the reduced overall correlation magnitudes in CL errors indicate better model performance, while the distinct error patterns confirm that HL are more sensitive to building geometry, whereas cooling loads are more influenced by fenestration characteristics.

Fig. 12 shows that the relative compactness, wall area, surface area, and glazing area have the most significant impact on HL and CL. These results could be explained by the fact that relative compactness has the potential to predict the annual building energy consumption. In fact, a

Table 5
Performance metric of the SVR, RF, and LightGBM.

Model	RMSE - HL	MAE - HL	R^2 - HL	MAPE - HL	RMSE - CL	MAE - CL	R^2 - CL	MAPE - CL	Ref.
SVR	1.9231	1.6624	0.8710	5.0079	1.1140	0.9207	0.6498	17.3530	Guo et al. (2023)
RF	0.7731	0.600	0.9835	1.8968	0.9383	0.7640	0.7941	13.4980	
LightGBM	0.7162	0.5634	0.9852	1.7176	0.3909	0.3093	0.9645	5.1995	
Random-LightGBM	0.4938	0.3894	0.9934	1.2038	0.2568	0.1807	0.9859	3.1249	
Grid-LightGBM	0.3610	0.2562	0.9965	0.8025	0.2247	0.1538	0.9893	2.6707	
CMA-ES-LightGBM	0.4023	0.2996	0.9957	0.9348	0.2272	0.1772	0.9888	2.9920	
TPE-LightGBM	0.2714	0.1416	0.9981	0.4699	0.1901	0.1394	0.9924	2.3509	

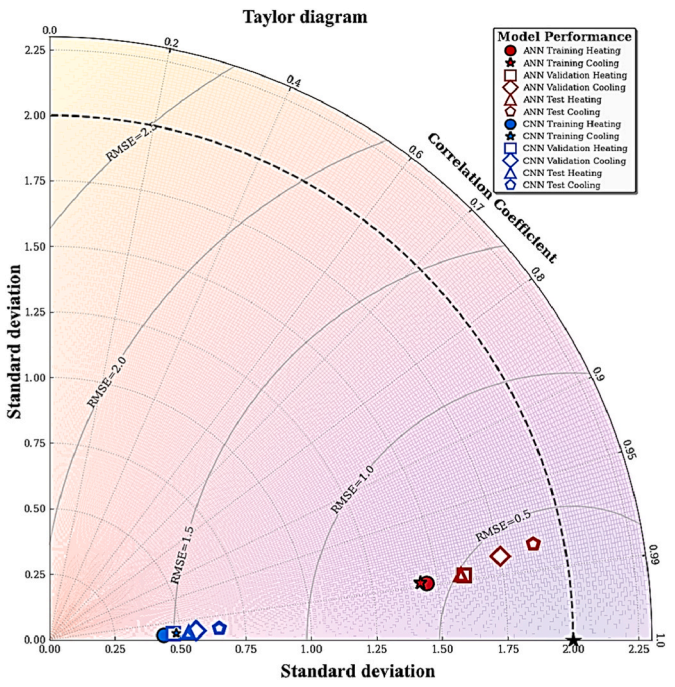

Fig. 10. Taylor diagram analysis.

higher ratio reduces the HL and CL (Kavaklioglu, 2018). In addition, the wall area, especially when residential buildings are not adequately insulated, causes heat transfer (Cui et al., 2024). Wall materials, such as concrete, wood, or particleboard, may have varying impacts on predicting HL and CL. For instance, clay brick, plywood, and concrete (normal weight) exhibit ρ_value of 0.2, 0.14, and 0.15, respectively (Mehta et al., 2013).

In the same vein, surface and glazing areas have an impact on HL and CL because window openings create heat transfer between the interior and exterior environments (Cherier et al., 2024). Conservatively, building orientation has the lowest significance on HL and CL (Ruiz & Bandera, 2014) because other geometric factors often overshadow its effects and are highly dependent on specific climatic conditions and building design.

Fig. 13 presents the Shapley Additive exPlanations (SHAP) feature construction analysis of the HL and CL. The relative compactness and wall area indicate the most dominant influence variable on the HL and CL due to the widest distribution of SHAP values. The wider the distribution of SHAP values for these variables, the greater their influence on the building's thermal predictions. Most importantly, the surface area variable shows a distinct negative correlation pattern; higher surface area values cluster on the negative side of the SHAP axis, indicating reduced HL and CL. Glazing area emerges as a critical factor for HL and CL, with high glazing values predicting strongly positive CL and HL. Overall height, roof area, glazing area distribution, and building orientation exhibit a low influence on the HL and CL.

The building envelope components, including walls, roofs, and windows, are fundamentally defined by their thermal resistance (R-value) and thermal transmittance (U-value). These two fundamental variables govern the rate of heat conduction across these assemblies, with higher R-values and lower U-values providing superior resistance to unwanted heat transfer. Additionally, building orientation varies substantially across different climate zones. In the USA, the ASHRAE building energy standards recognize eight primary climate zones (Openai, 2025). The USA climate zone includes Zone 1 (Warm-Humid, including Hawaii, Guam, Puerto Rico, and the Virgin Islands), Zone 2 (hot-dry and mixed-dry), Zone 3A and 3B (Hot-Humid and Hot-Dry), Zone 4A and 4B (mixed-humid and mixed-dry), Zone 5A and 5B (cold), Zone 6A and 6B (cold and very-cold), Zone 7 (very-cold), and

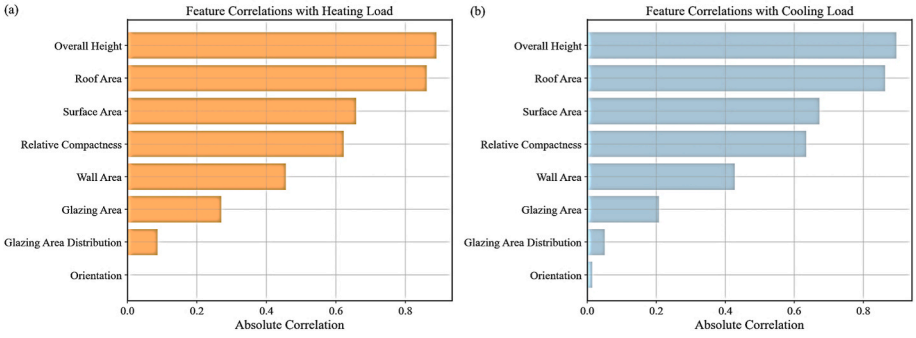

Fig. 11. Impacts on (a) heating and (b) cooling errors based on CNN models.

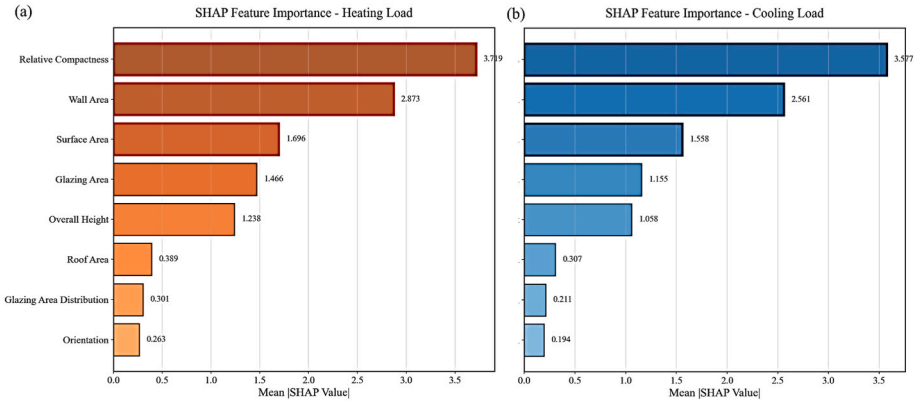

Fig. 12. Feature importance on (a) HL and (b) CL based on CNN models.

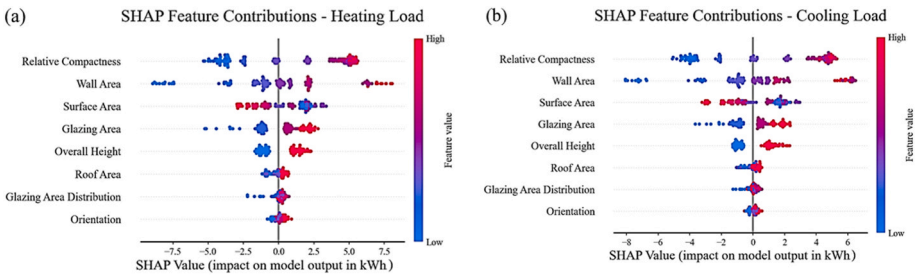

Fig. 13. SHAP Feature contributions of (a) heating load, and (b) cooling load.

Zone 8 (subarctic, primarily Alaska).

4. Cross-validation analysis

Our comprehensive assessment employs a 5-fold cross-validation strategy. Also, it helps to ensure that the CNN model's performance is tested across diverse data partitions while maintaining statistical validity in performance estimation. Taking this further, the R^2 score analysis, as depicted in Fig. 14(a), demonstrates exceptional consistency in HL predictions across all five validation folds. The validation R^2 values exhibit remarkable stability, ranging from 0.9609 (Fold 5) to 0.9686 (Fold 2), representing a narrow performance band of only 0.0077. This minimal variance indicates that the CNN architecture successfully captures generalizable thermal patterns rather than memorizing fold-specific artifacts. Specifically, the fold performance shows fold 1 (0.9641), fold 2 (0.9686), fold 3 (0.9637), fold 4 (0.9653), and fold 5 (0.9609), with a mean validation R^2 of 0.9645 $\pm$ 0.0026. On the other hand, CL predictions, while maintaining high overall performance, exhibit slightly greater variability. The validation R^2 scores span from 0.9253 to 0.9419. As a result, CV provides critical insights into model stability: HL demonstrates exceptional consistency with a CV of 0.26%, while CL maintains acceptable stability with a CV of 0.65%. Thus, both values fall well below the 2% threshold typically considered acceptable for engineering applications by indicating superior model reliability. In the same vein, the RMSE evaluation, as shown in Fig. 14(b), demonstrates high consistency across validation folds, ranging from 1.7837 (fold 2) to 1.9752 (fold 3) for HL with a mean validation RMSE of 1.8889 $\pm$ 0.0735. In addition, CL RMSE values, while exhibiting higher absolute errors, maintain reasonable consistency, ranging from 2.2525 (fold 5) to 2.5559 (Fold 2). It yields a mean validation RMSE of 2.3943 $\pm$ 0.1486. Therefore, we can conclude that the elevated RMSE values for CL reflect the increased complexity of modeling solar-driven thermal effects and their nonlinear interactions with building envelope characteristics compared to HL.

Fig. 15(a) presents the MAE analysis, revealing consistent absolute error patterns across all validation folds. The MAE values provide a useful measure of prediction accuracy that is less sensitive to outliers compared to RMSE. HL MAE values cluster tightly, indicating consistent prediction quality across all data partitions. On the other hand, CL MAE values, while higher in magnitude, demonstrate acceptable consistency across folds, reinforcing the model's robust error characteristics under diverse data conditions. Therefore, the relationship between RMSE and MAE values suggests that the model exhibits minimal sensitivity to extreme prediction errors, as evidenced by the relatively modest RMSE-to-MAE ratios. This characteristic is particularly valuable for practical applications where consistent prediction accuracy is more important than occasional exceptional performance. Furthermore, the training history validation loss curves (Fig. 15(b)) reveal uniform convergence patterns across all five folds. It illustrates consistent optimization behavior regardless of data partitioning. The curves exhibit rapid initial decay within the first 25 epochs, followed by gradual stabilization around 150 epochs. As a result, this consistent convergence behavior indicates that the CNN architecture and hyperparameter configuration are well-suited to the underlying data distribution across all folds. The training curves demonstrate that the 300-epoch training duration provides sufficient optimization time while avoiding overtraining. The stabilization phase, which begins around epoch 50, suggests that early stopping mechanisms can optimize computational efficiency without compromising performance quality.

The training versus validation performance analysis (Fig. 16(a)) reveals excellent alignment between training and validation metrics across all folds, with data points clustering located around the perfect agreement line. For HL, the training R^2 values (ranging from 0.9651 to 0.9696) closely match validation performance (0.9609 to 0.9686), indicating minimal overfitting across all folds. The training and validation gap CV analysis provides critical insights into model

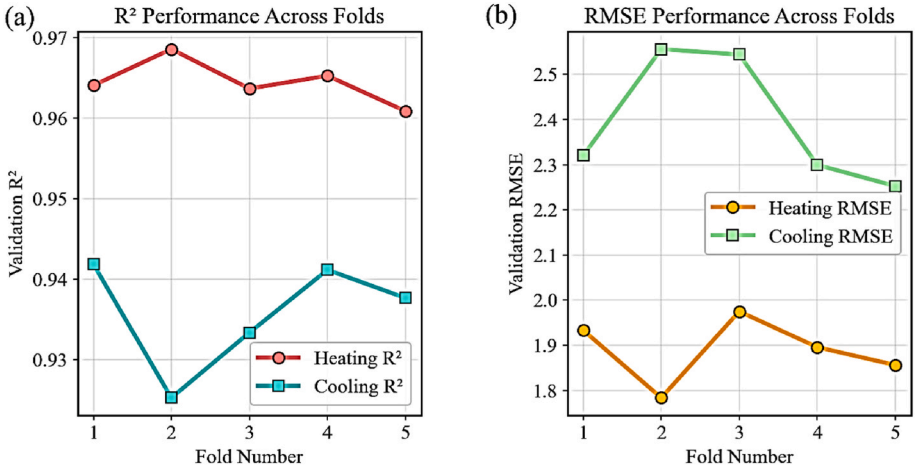

Fig. 14. Heating and cooling load k-fold cross-validation of (a) R^2 scores and (b) RMSE.

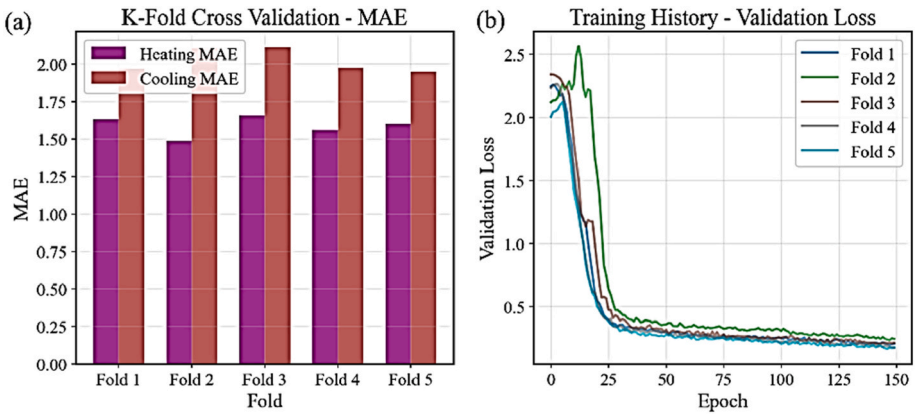

Fig. 15. (a) Heating and cooling cross-validation of MAE, (b) training history validation loss.

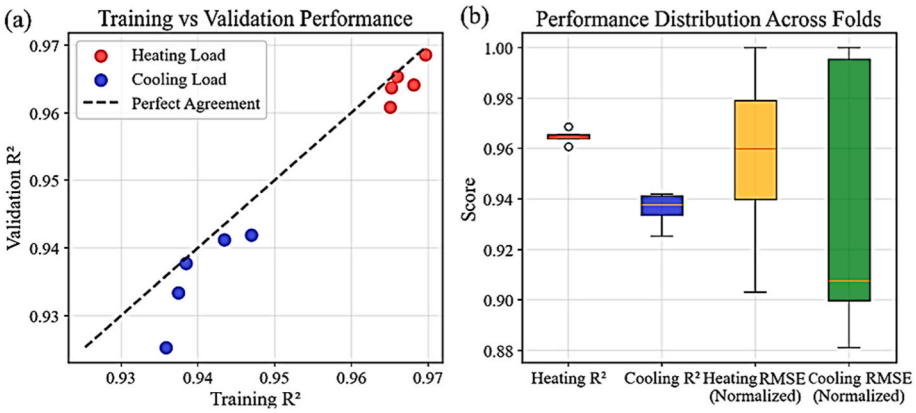

Fig. 16. Heating and cooling loads with (a) training versus validation performance, (b) performance distribution across folds.

generalization, with the following results:

- HL R^2 training is 0.9668 ± 0.0018, versus R^2 validation is 0.9645 ± 0.0026,
- CL R^2 training is 0.9405 ± 0.0042, versus R^2 validation is 0.9359 ± 0.0065.

The performance distribution analysis (Fig. 16(b)) provides a comprehensive statistical assessment through box plots. It reveals performance quartiles and outlier detection. Apart from that, HL R^2 scores exhibit tight interquartile ranges with median values near 0.965, demonstrating exceptional stability across all data partitions. The minimal presence of outliers confirms consistent performance regardless of data composition. In continuation, CL distributions exhibit broader interquartile ranges but maintain acceptable central tendencies, with median R^2 values of around 0.937. Therefore, the normalized RMSE distributions reinforce these patterns, with HL showing compact distributions and CL displaying wider but still acceptable distributions.

Fig. 17(a) quantifies CV stability through mean performance metrics with standard deviation indicators. The analysis reveals that:

- HL reliability, the mean validation R^2 is 0.9645 ± 0.026 (CV = 0.26%),

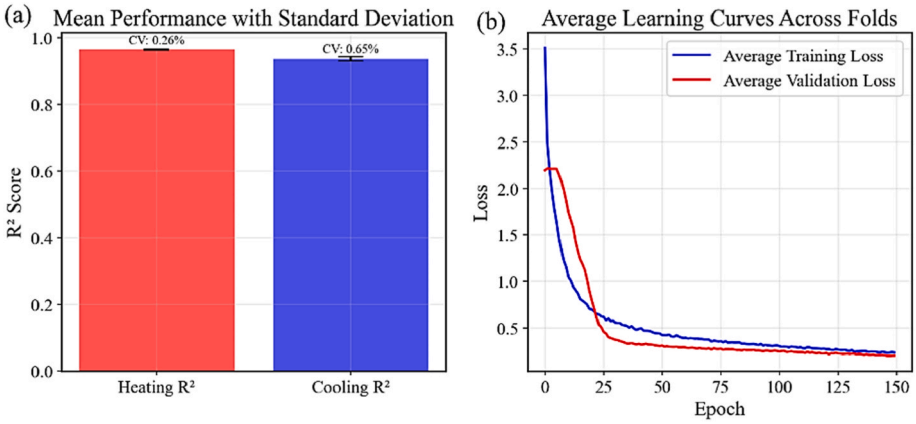

Fig. 17. (a) Mean performance with standard deviation, (b) average learning curves across folds.

- CL reliability, the mean validation R^2 is 0.9359 ± 0.065 (CV = 0.65%).

The 5-fold cross-validation analysis demonstrated outstanding model reliability with CV values of 0.26% for HL and 0.65% for CL, both well below the 2% (Pontoriero et al., 2021). Indeed, these low coefficients of variation provide strong statistical evidence of model robustness and reliability.

The average learning curves across folds (Fig. 17(b)) demonstrate optimal training dynamics with consistent convergence patterns. As a result, both the average training and validation loss curves exhibit a rapid initial decay followed by gradual stabilization, with close alignment between the training and validation curves throughout the optimization process. Therefore, this behavior indicates effective regularization and balanced model complexity.

Considering the ablation study presented in Table 6 after k-fold cross-validation analysis, the results demonstrate the importance of regularization mechanisms for model generalization. Table 6 highlights the metric performance of the CNN algorithm configuration with no dropout, batch normalization, L2 regularization, and no regularization.

As depicted in Fig. 18(a), the no-dropout configuration exhibits training instability, characterized by high-frequency oscillations between 150 epochs and 200 epochs, with an MSE loss magnitude of approximately 10^{-2}. Despite minimal train and validation R^2 gaps (HL: 0.0019; CL: 0.0011), this oscillatory behavior suggests marginal convergence stability. Conversely, the no batch normalization configuration achieves a stable learning curve, with the HL and CL learning curves converging to a range of 10^{-1} and 10^{-2} MSE loss values at 300 epochs (Fig. 18(b)). Most notably, the no regularization configuration achieves both the lowest terminal loss values and the fastest convergence to stability, with the training HL loss reaching approximately 10^{-3} at 300 epochs, along with train and validation R^2 gaps of −0.0001 for HL and 0.0040 for CL, respectively (Fig. 18(c)). The no L2 regularization variant, as illustrated in Fig. 18(d), exhibits intermediate convergence characteristics, positioning itself between the stable but slower configuration with no batch normalization and the rapid no regularization configuration. Critically, the no regularization configuration emerges as the optimal choice for production deployment.

This configuration reduces inference latency by 8% to 12% (on Jetson Nano: 32 ms versus 35 to 39 ms for regularized variants). It potentially decreases the memory footprint by approximately 15 kilobytes (KB) due to the elimination of memory associated with batch normalization. Furthermore, this configuration implies a monthly operational cost reduction for cloud infrastructure serving, for example, more than 1000 buildings. We recommend deploying the non-regularization configuration of the CNN model for real-time HVAC control, including climate zone parameters, where prioritizing inference speed, computational efficiency, and operational simplicity remain essential.

Therefore, as a proof of concept and educational tool, with the assumption of not considering parameters such as climate variations, for instance, using the HL and CL prediction application (see interface in Fig. 19) can be forecasted using the following link repository: https://derrickmirindi.github.io/building-energy-prediction/

5. Community engagement and ethical considerations

In any part of the world, the successful implementation of CNN-driven building energy prediction models within affordable housing and energy equity contexts necessitates robust community engagement strategies and adherence to rigorous ethical principles. Indeed, we propose establishing a comprehensive citizen science program and science education program that actively engages community members, students, and researchers as collaborative partners in developing the building energy sector. We expect that this participatory approach recognizes that sustainable building energy consumption solutions require not only technical innovation but also the capacity-building of communities to understand, utilize, and equitably benefit from energy prediction technologies.

5.1. Community engagement implementation

Our proposed citizen energy consumption science program aims to foster trust and long-lasting relationships among academic researchers, housing professionals, and community members through intentional collaboration and multi-level community engagement. To successfully implement the CNN application, we recommend establishing an equitable pathways committee comprising representatives from municipal housing authorities, non-governmental organizations serving affordable housing communities, neighborhood-based justice organizations, and academic institutions. Meanwhile, this committee would convene quarterly to translate technical building energy data and AI predictions into the lived experiences of community members, develop equitable pathways for technology deployment, and establish transdisciplinary research to ensure the appropriateness of the study and the benefit to the community. Notably, the steering committee will guide research priorities, community outreach strategies, and the development of research questions that address actual community needs, rather than remaining academic interests.

The science education program will train community members, including youth residents of affordable housing, community organizers, students, and building maintenance personnel, to understand building thermal performance, collect real-time data, and participate meaningfully in energy retrofit planning and HVAC system optimization decisions. The education science program will focus on educating community members on fundamental building physics principles (e.g., heat transfer, thermal envelope performance, and HVAC systems), data literacy skills (e.g., interpreting energy consumption data and understanding prediction confidence intervals), and policy applications (building codes, retrofit incentive programs, and energy labeling systems). Engaging communities will create opportunities for students and researchers to collaborate on research, promoting mutual respect between professional scientists and the public while enhancing the learning experiences of all parties involved. Universities serve as an essential intersection between scholarly research and community interests. They ensure that AI-driven building energy initiatives remain ethically grounded and focused on achieving lasting societal benefits.

5.2. Ethical considerations for responsible artificial intelligence deployment

Despite the substantial benefits of implementing citizen science and

Table 6
Performance metric of CNN with no Dropout, Batch Normalization, L2 regularization, and regulation.

Model configuration	Training HL R^2	Validation HL R^2	Training CL R^2	Validation CL R^2	HL gap	CL gap	Overfitting status
No Dropout	0.9944	0.9924	0.9908	0.9897	0.0019	0.0011	Low
No BatchNorm	0.9898	0.9890	0.9729	0.9815	0.0008	−0.0087	Low
No L2 Reg	0.9755	0.9735	0.9425	0.9463	−0.0010	−0.0038	Low
No Regularization	0.9959	0.9960	0.9908	0.9868	−0.0001	0.0040	Optimal

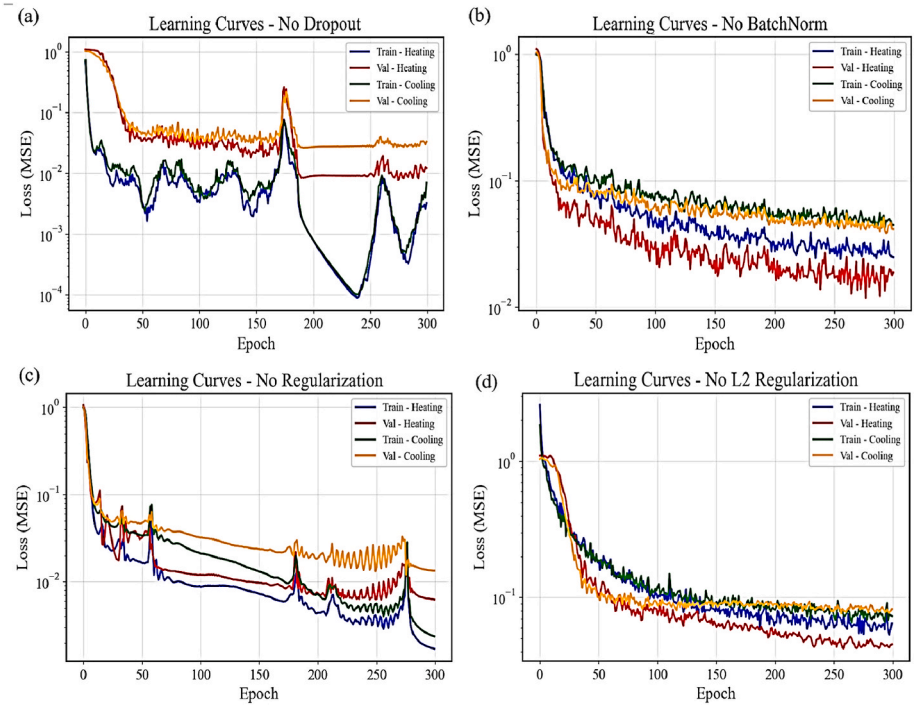

Fig. 18. Learning curves of CNN models with no Dropout, Batch Normalization, L2 regularization, and regulation.

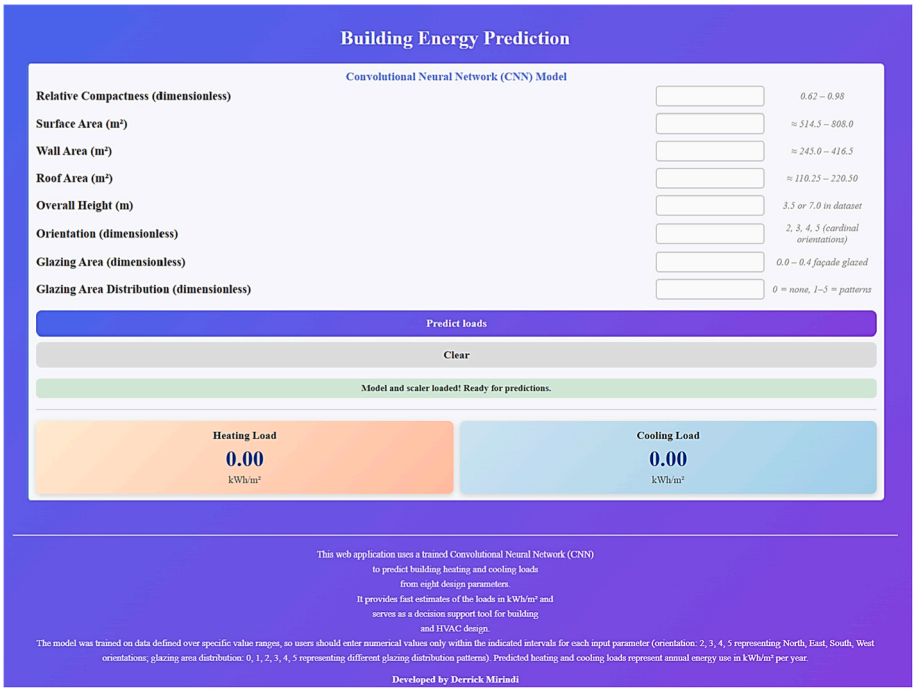

Fig. 19. Building an energy prediction interface based on a convolutional neural network (access link: https://derrickmirindi.github.io/building-energy-prediction/).

education science programs, they introduce critical ethical considerations that should be rigorously addressed throughout the research lifecycle, from initial design through data collection, analysis, and dissemination of findings. Therefore, we identify five essential ethical principles that should be taken into consideration to implement AI-driven building energy successfully:

i. Data quality, integrity, and scientific rigor: The reliability and validity of building energy data collected through citizen science should adhere to scientific standards to ensure AI model predictions remain trustworthy and actionable for policy applications. In fact, the implementation of comprehensive training programs for citizen scientists will cover standardized data collection protocols, quality assurance procedures, and validation techniques. Following established frameworks for

information quality in citizen science, and four elements should be addressed: scientific quality (adherence to rigorous methodologies ensuring data validity), product quality (usability and relevance of data for building energy applications), stewardship quality (long-term data management, accessibility, and preservation), and service quality (ongoing support, training, and guidance for citizen scientists) (Downs et al., 2021). Professional oversight by building scientists and engineers will ensure data accuracy while respecting community contributions and input. In addition, transparent documentation of data sources, collection methods, potential limitations, and uncertainty quantification should maintain scientific integrity while acknowledging the valuable role of non-professional participants.

ii. Data ownership, intellectual property, and equitable benefit-sharing: Citizen scientists who contribute substantial time, effort, and local knowledge to building energy data collection deserve recognition, appropriate compensation, and meaningful authority in determining data usage and distribution. It is essential to establish transparent agreements at project initiation defining: (a) data ownership rights and responsibilities; (b) intellectual property arrangements for research findings, publications, and potential commercialization; (c) mechanisms for equitable benefit-sharing ensuring communities contributing data receive tangible benefits such as prioritized retrofit program access, reduced energy costs, or infrastructure improvements; and (d) privacy protections preventing misuse of household energy consumption data for discriminatory purposes.

iii. Informed consent, voluntary participation, and right to withdraw: Ethical citizen science requires that all participants provide informed consent based on a comprehensive understanding of project purposes, methodologies, potential risks and benefits, time commitments, data usage, and privacy protections. It is crucial to develop accessible consent materials in multiple languages, clearly explaining these dimensions, utilizing visual aids, and employing plain language to ensure comprehension across diverse literacy levels. Participants should understand their voluntary status and unconditional right to withdraw at any time without penalty or negative consequences. Consent processes will address specific concerns relevant to building energy research including: (a) how household energy consumption data will be used, stored, and protected; (b) whether individual building addresses or household identities will be disclosed; (c) potential risks such as data breaches or inadvertent revelation of occupancy patterns; and (d) mechanisms for participants to access, review, and request correction or deletion of their contributed data.

iv. Privacy protection, confidentiality, and prevention of discriminatory use: Building energy data inherently reveals sensitive information about household occupancy patterns, economic status, and daily routines, creating substantial privacy risks if inadequately protected. It is essential to implement rigorous data security protocols, including encryption, secure storage systems, restricted access controls, and regular security audits, to ensure the protection of sensitive information. Household-level energy data-related information should be aggregated or anonymized whenever possible to prevent identification of individual households, with geographic resolution sufficient for neighborhood-scale analysis while protecting household privacy. Critically, these strategies should be committed to preventing the discriminatory use of building energy data, for example, its use by landlords to justify rent increases, by insurers to deny coverage, or by municipalities to target enforcement actions that disproportionately affect low-income neighborhoods. Legal agreements and ethical oversight committees will establish enforceable prohibitions against such discriminatory applications. Transparent data governance frameworks will specify legitimate uses (such as research, policy development, and community benefit) and explicitly prohibit illegitimate uses that threaten participant privacy or perpetuate inequities.

v. Equity, inclusivity, and prevention of exploitation: Citizen science must prioritize equitable participation opportunities, ensuring that underrepresented populations, including low-income households, racialized communities, indigenous peoples, persons with disabilities, elderly residents, and recent immigrants, have meaningful access to participation and receive proportional benefits from research outcomes. To meet these expectations, it will be important to: (a) proactive outreach to marginalized communities often excluded from research; (b) removal of participation barriers through provision of necessary equipment, technology access, transportation, childcare support, and language interpretation services; (c) accommodation of diverse abilities through accessible materials, flexible participation formats, and assistive technologies; (d) fair compensation recognizing citizen scientists' time and expertise through stipends, reduced utility bills, prioritized retrofit program access, or other tangible benefits; and (e) transparent communication preventing exploitation wherein participants contribute labor without receiving proportional benefits or influence over research direction. Exploitation prevention requires clearly defining roles, responsibilities, time commitments, and compensation at project initiation, ensuring participants understand their actual influence on research design versus data collection roles, and providing regular opportunities for feedback on project operations and benefit distribution.

By adhering to these five ethical principles and establishing transparent governance structures, including ethics review committees with community representation, our proposed programs, facilitated through community engagement, can democratize access to building energy information and ensure that AI-driven technologies serve equitable development rather than exacerbating existing inequities. Also, these ethics positions community members not as passive data sources, but as active collaborators who shape research priorities, interpret findings, and translate technical knowledge into community-led climate resilience strategies.

6. Conclusions

We have used DL algorithms, such as ANN and CNN, to predict the HL and CL of residential buildings. This study results in the following conclusion.

- The correlation analysis and the feature importance analysis revealed that the relative compactness, wall, surface, and glazing areas dominate thermal load predictions, with HL and CL. Additionally, all input variables exhibit a significant positive correlation with HL and CL, except for the building surface and roof areas, as well as the glazing area, which has a significant positive correlation with HL.
- CNN algorithm significantly outperformed the ANN architecture. CNN models achieve the highest R^2 values, exceeding 0.99 for all dataset splits (training, testing, and validation datasets), with dramatically reduced error; however, the models suggest potential overfitting.
- The 5-fold cross-validation trained models analysis demonstrated outstanding model reliability with CV values of 0.26% for HL and 0.65% for CL, acceptable for engineering applications.
- The no regularization configuration of the CNN model emerges as optimal for real-world deployment applications, prioritizing computational efficiency and operational cost reduction. This configuration reduces inference latency by 8% to 12% (from 32 ms to 29-33 ms for regularized configurations) and decreases memory

storage by approximately 15 KB by eliminating the batch normalization configuration. The results show low gap metric values, typically −0.0001 and 0.0040, between the training and validation HL and CL models.

- The implementation of citizen science and education science programs through community engagement will democratize access to building energy information, support capacity-building among vulnerable populations, and ensure equitable benefits from AI-driven technologies. Five ethical considerations guide this implementation, namely data quality and scientific integrity, data ownership and equitable need-sharing, informed consent and voluntary participation, privacy protection and prevention of discriminatory use, and equity, inclusivity, and prevention of exploitation.

For further research, we recommend:

- To incorporate occupants, meteorological conditions (e.g., temperature, solar radiation, wind speed, humidity), and occupancy schedules to accurately improve the thermal load prediction by varying climate conditions, which differ by country or region and occupancy parameters.
- To conduct a detailed cost implication analysis by quantifying capital cost reductions, operational savings, health benefits, and return on investment using an AI-assisted thermal load optimization.
- Participatory co-design: conducting collaborative research with housing residents and community organizations to ensure AI tools reflect user needs and democratize energy information access.
- To develop a responsible AI governance approach addressing data privacy, algorithmic fairness, community benefit, and accountability.
- To incorporate probabilistic prediction intervals, enabling stakeholders to understand model confidence levels and associated risks.

CRediT authorship contribution statement

Derrick Mirindi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Manisha Kandel:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis. **Abiodun Adedoyin:** Writing – review & editing, Writing – original draft. **David Sinkhonde:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation, Conceptualization. **Tajebe Bezabih:** Writing – review & editing, Writing – original draft. **Ahana Bayongwa Samuel:** Writing – review & editing, Writing – original draft. **Eberechi Cecilia Osuagwu:** Writing – review & editing, Visualization. **Frederic Mirindi:** Writing – review & editing, Writing – original draft, Data curation, Conceptualization.

Data availability statement

The dataset used in this study is publicly available and was originally published by Tsanas & Xifara, 2012. The dataset can be accessed through the UC Irvine Machine Learning Repository at https://archive.ics.uci.edu/ml/datasets/Energy+efficiency.

Declaration of the use of AI assisted technologies

AI tools were used to assist in coding language, but the authors are fully responsible for the content.

Ethics statement

Ethical approval is not applicable to this manuscript.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or non-profit sectors.

Declaration of competing interest

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ssaho.2026.102647.

References

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

Ardabili, S., Abdolalizadeh, L., Mako, C., Torok, B., & Mosavi, A. (2022). Systematic review of deep learning and machine learning for building energy. *Frontiers in Energy Research, 10*, Article 786027. https://doi.org/10.3389/fenrg.2022.786027.

Billanes, J. D., Ma, Z. G., & Jørgensen, B. N. (2025). Data-driven technologies for energy optimization in smart buildings: A scoping review. *Energies, 18*(2), 1–49. https://doi.org/10.3390/en18020290.

Chaganti, R., Rustam, F., Daghriri, T., Díez, I.d. l. T., Mazón, J. L. V., Rodríguez, C. L., & Ashraf, I. (2022). Building heating and cooling load prediction using ensemble machine learning model. *Sensors, 22*(19), 7692. https://doi.org/10.3390/s22197692.

Cherier, M. K., Hamdani, M., Kamel, E., Guermoui, M., El Amine Bekkouche, S. M., Al-Saadi, S., Djeffal, R., Bashir, M. O., Elshekh, A. E. A., Drozdova, L., Kanan, M., & Flah, A. (2024). Impact of glazing type, window-to-wall ratio, and orientation on building energy savings quality: A parametric analysis in Algerian climatic conditions. *Case Studies in Thermal Engineering, 61*, Article 104902. https://doi.org/10.1016/j.csite.2024.104902

Chirici, G., Mura, M., McInerney, D., Py, N., Tomppo, E. O., Waser, L. T., Travaglini, D., & McRoberts, R. E. (2016). A meta-analysis and review of the literature on the k-Nearest Neighbors technique for forestry applications that use remotely sensed data. *Remote Sensing of Environment, 176*, 282–294. https://doi.org/10.1016/j.rse.2016.02.001.

Cui, X., Lee, M., Koo, C., & Hong, T. (2024). Energy consumption prediction and household feature analysis for different residential building types using machine learning and SHAP: Toward energy-efficient buildings. *Energy and Buildings, 309*, Article 113997. https://doi.org/10.1016/j.enbuild.2024.113997.

Downs, R. R., Ramapriyan, H. K., Peng, G., & Wei, Y. (2021). Perspectives on citizen science data quality [Perspective]. *Frontiers in Climate, 3*. https://doi.org/10.3389/fclim.2021.615032.

Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software, 40*(5), 356–362. https://doi.org/10.1016/j.advengsoft.2008.05.003.

Energy. (2025). *Heating, ventilation, air conditioning, refrigeration, and water heating.* https://www.energy.gov/eere/buildings/heating-ventilation-air-conditioning-refrigeration-and-water-heating#:~:text=and%20Water%20Heating-,Heating%2C%20Ventilation%2C%20Air%20Conditioning%2C%20Refrigeration%2C%20and%20Water%20Heating,by%2050%25%20within%20a%20decade.

Epochai. (2022). Trends in GPU price-performance. https://epoch.ai/blog/trends-in-gpu-price-performance.

Frank, E., Trigg, L., Holmes, G., & Witten, I. H. (2000). Technical note: Naive bayes for regression. *Machine Learning, 41*(1), 5–25. https://doi.org/10.1023/A:1007670802811.

Ghritlahre, H. K., & Prasad, R. K. (2018). Application of ANN technique to predict the performance of solar collector systems-A review. *Renewable and Sustainable Energy Reviews, 84*, 75–88. https://doi.org/10.1016/j.rser.2018.01.001.

Giannini, T., & Bowen, J. P. (2024). The digital road to the Age of AI: 1950s to the present. In T. Giannini, & J. P. Bowen (Eds.), *The arts and computational culture: Real and virtual worlds* (pp. 95–133). Nature Switzerland: Springer. https://doi.org/10.1007/978-3-031-53865-0_4.

Globalabc. (2025). Global status report for buildings and construction 2024/2025. https://globalabc.org/sites/default/files/2025-03/Global-Status-Report-2024_2025.pdf.

Goel, S., Rosenberg, M. I., & Eley, C. (2017). *ANSI/ASHRAE/IES standard 90.1-2016 performance rating method reference manual.*

Guo, J., Yun, S., Meng, Y., He, N., Ye, D., Zhao, Z., Jia, L., & Yang, L. (2023). Prediction of heating and cooling loads based on light gradient boosting machine algorithms. *Building and Environment, 236*, Article 110252. https://doi.org/10.1016/j.buildenv.2023.110252.

Hatamizadeh, A., Yin, H., Heinrich, G., Kautz, J., & Molchanov, P. (2023). *Global context vision transformers* proceedings of the 40th international conference on machine learning. *Proceedings of Machine Learning Research*. https://proceedings.mlr.press/v202/hatamizadeh23a.html.

Jiang, G., & Wang, W. (2017). Error estimation based on variance analysis of k-fold cross-validation. *Pattern Recognition, 69*, 94–106. https://doi.org/10.1016/j.patcog.2017.03.025.

Kamnis, S. (2023). Generative pre-trained transformers (GPT) for surface engineering. *Surface and Coatings Technology, 466*, Article 129680. https://doi.org/10.1016/j.surfcoat.2023.129680.

Kavaklioglu, K. (2018). Robust modeling of heating and cooling loads using partial least squares towards efficient residential building design. *Journal of Building Engineering, 18*, 467–475. https://doi.org/10.1016/j.jobe.2018.04.018.

Likas, A., Vlassis, N., & Verbeek, J. J. (2003). The global k-means clustering algorithm. *Pattern Recognition, 36*(2), 451–461. https://doi.org/10.1016/S0031-3203(02)00060-2.

Liu, S., Jiang, H., Wu, Z., & Li, X. (2022). Data synthesis using deep feature enhanced generative adversarial networks for rolling bearing imbalanced fault diagnosis. *Mechanical Systems and Signal Processing, 163*, Article 108139. https://doi.org/10.1016/j.ymssp.2021.108139.

Long, L. D. (2023). An AI-driven model for predicting and optimizing energy-efficient building envelopes. *Alexandria Engineering Journal, 79*, 480–501. https://doi.org/10.1016/j.aej.2023.08.041.

Mehta, M., Scarborough, W., & Armpriest, D. (2013). *Building construction: Principles, materials, and systems*.

Mirindi, D., Hunter, J., Sinkhonde, D., Bezabih, T., & Mirindi, F. (2025). Performance of machine learning algorithms to evaluate the physico-mechanical properties of nanoparticle panels. *Green Technologies and Sustainability, 3*(4), Article 100235. https://doi.org/10.1016/j.grets.2025.100235.

Mirindi, D., Hunter, J., Sinkhonde, D., & Mirindi, F. (2025). Machine learning-driven analysis of nanoparticle performance on concrete mechanical properties. *Manufacturing Letters, 44*, 1657–1668. https://doi.org/10.1016/j.mfglet.2025.07.001.

Mirindi, D., Khang, A., & Mirindi, F. (2025). *Artificial intelligence (AI) and automation for driving green transportation systems: A comprehensive review. Driving Green Transportation System Through Artificial Intelligence and Automation: Approaches, Technologies and Applications* (pp. 1–19). https://doi.org/10.1007/978-3-031-72617-0_1.

Mirindi, D., Mirindi, F., Bezabih, T., Sinkhonde, D., & Kiarie, W.. Review: The role of artificial intelligence in building information modeling conference. https://doi.org/10.1145/3716489.3728433.

Mirindi, D., Sanders, T. N., & Hunter, J. (2024). Integration of artificial intelligence and smart technologies in offsite construction: A comprehensive review. *Transforming Construction with Off-site Methods and Technologies*. https://doi.org/10.57922/6pbaw583.

Mirindi, D., Sinkhonde, D., Mirindi, F., & Bezabih, T. (2025). A review on Aerospace-AI, with ethics and implications. https://doi.org/10.11648/j.jccee.20251002.12.

Mirindi, D., Sinkhonde, D., Mirindi, F., & Bezabith, T.. Advanced evaluation of BIM-GenAI using OpenAI o1 and ethical considerations. https://doi.org/10.1145/3716489.3728431.

Mirindi, D., Sinkhonde, D., Mirindi, F., & Bezabith, T. (2025). Advance toward artificial superintelligence with openAI's O1 reinforcement learning and ethics. *2025 6th international conference on artificial intelligence* (pp. 475–481). https://doi.org/10.1109/AIRC64931.2025.11077494.

Mirindi, D., & Yazdandoust, F. (2024). Historical foundations to modern innovations of urban village. *Proceedings of the international conference of contemporary affairs in architecture and urbanism-ICCAUA, 7*(1), 605–613. https://doi.org/10.38027/ICCAUA2024EN0191.

Murtagh, F. (1991). Multilayer perceptrons for classification and regression. *Neurocomputing, 2*(5), 183–197. https://doi.org/10.1016/0925-2312(91)90023-5.

Myles, A. J., Feudale, R. N., Liu, Y., Woody, N. A., & Brown, S. D. (2004). An introduction to decision tree modeling. *Journal of Chemometrics: A Journal of the Chemometrics Society, 18*(6), 275–285. https://doi.org/10.1002/cem.873.

Neto, A. H., & Fiorelli, F. A. S. (2008). Comparison between detailed model simulation and artificial neural network for forecasting building energy consumption. *Energy and Buildings, 40*(12), 2169–2176. https://doi.org/10.1016/j.enbuild.2008.06.013.

Openai. (2025). *ASHARE climate zones*. https://openei.org/wiki/ASHRAE_Climate_Zones.

Mirindi, F., & Mirindi, D. (2025). Forecasting energy prices using machine learning algorithms: A comparative analysis. In M. Z. Abedin, & W. Yong (Eds.), *Machine learning technologies on energy economics and finance: Energy and sustainable analytics, 1* pp. 135–146). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-94862-6_6.

Pontoriero, A. D., Nordio, G., Easmin, R., Giacomel, A., Santangelo, B., Jahuar, S., Bonoldi, I., Rogdaki, M., Turkheimer, F., & Howes, O. (2021). Automated data quality control in FDOPA brain PET imaging using deep learning. *Computer Methods and Programs in Biomedicine, 208*, Article 106239. https://doi.org/10.1016/j.cmpb.2021.106239.

Rodriguez-Torrealba, R., Garcia-Lopez, E., & Garcia-Cabot, A. (2022). End-to-End generation of multiple-choice questions using text-to-text transfer transformer models. *Expert Systems with Applications, 208*, Article 118258. https://doi.org/10.1016/j.eswa.2022.118258.

Ruiz, R., & Bandera, F. (2014). Importance of orientation in building energy-savings. *Proceedings of the World Sustainable Building Conference, 14*, 327–334.

Sinkhonde, D., Mirindi, D., Dabakuyo, I., Bezabih, T., Mashava, D., & Mirindi, F. (2025). Applications of machine learning algorithms on the compressive strength of laterite blocks made with metakaolin-based geopolymer and sugarcane molasses. *Waste Management Bulletin, 3*(3), Article 100212. https://doi.org/10.1016/j.wmb.2025.100212.

Strategicmarketresearch. (2025). Building energy management system market report 2030. https://www.strategicmarketresearch.com/market-report/building-energy-management-system-market.

Thinking. (2015). There's nothing passive about passive house. https://hcma.ca/thinking/theres-nothing-passive-about-passive-house/#:~:text=The%20concept%20of%20Passive%20House,then%20satisfy%20overall%20energy%20demand.

Tiwari, S. K., Kaur, J., & Kaur, R. (2024). Predictive modelling for heating and cooling load systems of residential building. *2024 IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI), 2*, 1–6. https://doi.org/10.1109/IATMSI60426.2024.10503016.

Traore, B. B., Kamsu-Foguem, B., & Tangara, F. (2018). Deep convolution neural network for image recognition. *Ecological Informatics, 48*, 257–268. https://doi.org/10.1016/j.ecoinf.2018.10.002.

Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings, 49*, 560–567. https://doi.org/10.1016/j.enbuild.2012.01.024.

Unep. (2024). Why making buildings greener is crucial to countering climate change. https://www.unep.org/news-and-stories/story/why-making-buildings-greener-crucial-countering-climate-change#:~:text=What%20is%20the%20link%20between,Report%20for%20Buildings%20and%20Construction.

Wang, Z., Hong, T., & Piette, M. A. (2020). Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy, 263*, Article 114683. https://doi.org/10.1016/j.apenergy.2020.114683.

Weerakody, P. B., Wong, K. W., Wang, G., & Ela, W. (2021). A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing, 441*, 161–178. https://doi.org/10.1016/j.neucom.2021.02.046.

Xu, Y. (2024). Research on cooling load estimation through optimal hybrid models based on naive bayes. *Journal of Engineering and Applied Science, 71*(1), 75. https://doi.org/10.1186/s44147-024-00396-9.

Yaseliani, M., Hamadani, A. Z., Maghsoodi, A. I., & Mosavi, A. (2022). Pneumonia detection proposing a hybrid deep convolutional neural network based on two parallel visual geometry group architectures and machine learning classifiers. *IEEE Access, 10*, 62110–62128. https://doi.org/10.1109/ACCESS.2022.3182498.

Zhang, X., & Liu, C.-A. (2023). Model averaging prediction by K-fold cross-validation. *Journal of Econometrics, 235*(1), 280–301. https://doi.org/10.1016/j.jeconom.2022.04.007.