Energy Reports 13 (2025) 4247–4260

Contents lists available at ScienceDirect

Energy Reports

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

Research Paper

Energy efficiency in building: Entropy-based Grey Wolf Optimization for improved MLP performance

İhsan Tuğal

Muş Alparslan University, Department of Software Engineering, Muş 49250, Türkiye

A R T I C L E I N F O

Keywords:
Building HVAC systems
Energy efficiency
Multi-layer perceptron
Hyperparameter optimization
Meta-heuristic algorithms
Entropy

A B S T R A C T

Energy efficiency in building HVAC systems is essential for sustainable development, requiring accurate predictions of heating and cooling loads. Traditional hyperparameter tuning methods often lead to suboptimal performance in artificial neural network models due to local optima and premature convergence issues. This study proposes an Entropy-Based Grey Wolf Optimization to enhance the hyperparameter tuning of Multi-Layer Perceptron models for improved load prediction. Unlike traditional Grey Wolf Optimization, which relies on a fixed linear decay for search balancing, the Entropy-Based Grey Wolf Optimization dynamically adjusts the search strategy based on entropy levels. This prevents premature convergence, enhances global search capabilities, and improves fine-tuning of Multi-Layer Perceptron hyperparameters for HVAC load prediction. The proposed optimization approach is compared with ten different optimization techniques, and results indicate that Entropy-Based Grey Wolf Optimization achieves superior accuracy with lower mean squared error and faster convergence. The study contributes to both the improvement of energy efficiency in HVAC systems and the development of a generalizable entropy-based optimization framework that can be applied to various machine learning tasks.

1. Introduction

Energy efficiency is defined as using less energy to perform a task. We need to design tools, devices, buildings, and production facilities that consume less energy. With the continuous rise in global energy demand, energy efficiency in buildings has become a crucial issue. Increasing concerns about energy waste and its long-term negative impact on the environment have led to significant research on energy performance of buildings (EPB) in recent years. Achieving energy efficiency in residential buildings is crucial both economically and environmentally. Today's smart buildings need to monitor, predict, and make decisions based on consumption data to optimize energy use. While modern residential buildings often meet energy-saving standards, various measures are still needed to enhance the efficiency of heating, ventilation, and air conditioning (HVAC) systems, which are the most energy-consuming systems in buildings (Ming et al., 2020). For instance, methods such as developing heat recovery heat pumps (Khanlari et al., 2020), utilizing renewable energy sources (Dermentzis et al., 2021), and optimizing the parameters of system equipment (Peng et al., 2023) are employed (Zhao et al., 2023). However, in practice, there are still many issues in the operation of building energy systems. Considering the high level of global energy consumption, energy shortages are likely to become a significant issue in the future. Some scientists attribute the rise in energy consumption to people's inclination to maintain high living standards. The increase in energy consumption is believed to stem from people's tendency to achieve higher living standards (Liu et al., 2016). Consequently, the need to propose an optimal prediction tool for building energy consumption has become a central focus of numerous studies, leading to the development of various assessment systems and numerical methods.

According to the 2023 data from the International Energy Agency (IEA), residential buildings consume 29 % of the final energy worldwide. This consumption is distributed as follows: electricity (21 %), heating and cooling (7 %), and heavy fuels (1 %). In developing countries, 70 % of the energy used in residential buildings is for heating and cooling, whereas in developed countries, this proportion is 40 %. Of the energy used in residential buildings, 72 % comes from fossil fuels (coal, oil, natural gas), while 28 % comes from renewable energy sources (solar, geothermal, biomass). Carbon emissions from residential buildings account for 21 % of the total global greenhouse gas emissions, with 70 % of these emissions resulting from the fossil fuels used for heating and cooling. Implementing insulation in buildings, using energy-efficient

E-mail address: i.tugal@alparslan.edu.tr.

https://doi.org/10.1016/j.egyr.2025.03.048
Received 29 December 2024; Received in revised form 16 March 2025; Accepted 26 March 2025
Available online 8 April 2025

appliances, and investing in renewable energy sources such as solar panels can significantly enhance energy efficiency. According to the IEA, improving energy efficiency in residential buildings could reduce global energy consumption by 10 % by 2030. One of the United Nations Sustainable Development Goals is to double the rate of improvement in energy efficiency in buildings by 2030. The European Union has set a target for all buildings to be net zero energy by 2050 (IEA, 2023).

Given the record high and low temperatures worldwide, the importance of being efficient in heating and cooling our buildings is increasingly evident (Tugal and Sevgin, 2023). Efforts should be made to enhance building comfort, reduce energy bills, and minimize the carbon footprint. Increasing your building's energy efficiency not only improves comfort and saves money but also helps the environment. Achieving this in harmony is the ultimate goal. The key to this balance lies in the precise calculation of cooling (CL) and heating (HL) loads. CL and HL are considered two critical criteria in the construction industry (Zhao et al., 2023). Various parameters, such as building characteristics and climatic conditions, are taken into account to estimate cooling and heating capacities (Calotă et al., 2024). Estimating HL and CL from simple features like wall area, height, etc., helps determine energy performance (Aloshan and Aldali, 2024). Accurate HL and CL predictions enable decision-makers to make investment decisions that benefit efficiency, savings, and the environment. The cornerstone of a comfortable indoor environment is the accurate calculation of heating and cooling loads. It determines the amount of heating or cooling needed to keep an area comfortable. Incorrect calculations can lead to discomfort due to uncomfortably high or low temperature fluctuations. There are numerous real-life examples demonstrating how incorrect load calculations can disrupt indoor comfort (Nicol and Humphreys, 2002). A well-calculated system, on the other hand, consumes less energy, leading to significant savings on electricity bills. Case studies show how precise load calculations significantly reduce energy consumption and translate into noticeable financial benefits (Fan et al., 2024).

HVAC refers to the devices and technologies used for heating and cooling an enclosed space, as well as controlling humidity and air quality. An HVAC system helps create a healthier indoor environment by providing enhanced indoor air quality. Properly sizing HVAC systems to meet the specific heating and cooling needs of a building is crucial. HVAC systems constitute a significant portion of energy consumption in both commercial and residential buildings. These systems are among the highest energy consumers in such buildings. By ensuring the efficient operation of HVAC systems, energy consumption and energy bills can be significantly reduced. The alignment between a building's actual energy needs and the heating and cooling capacity provided by the HVAC system must be verified through measurements. Insufficient capacity results in low energy efficiency and high energy consumption (Solano et al., 2021). Thus, precise load estimation is crucial for the optimal operation and control of HVAC systems. In the context of HVAC, load calculation is the process of determining the precise heating and cooling requirements for a specific area. It involves a series of calculations that consider factors such as insulation, occupancy, and external conditions. Accurately predicting a building's energy consumption is challenging due to the complex and nonlinear relationship between these influencing parameters and building energy performance (Xu et al., 2022). The advantages of load calculation in HVAC design are undeniable. It results in improved system performance, energy savings, and proper equipment sizing.

Optimizing energy efficiency in HVAC systems is crucial for businesses and homeowners, representing an innovative and responsible approach. Embracing energy efficiency not only provides cost savings but also contributes to a sustainable future. Implementing energy-saving strategies can reduce energy consumption, lower operational costs, improve indoor comfort, and promote environmental sustainability. To optimize energy efficiency in HVAC systems, it is important to understand how these systems consume energy and to be aware of relevant energy efficiency ratings and standards. Analyzing a building's energy performance is essential for reducing energy use. Methods that are either too large or too small can lead to energy waste and reduced efficiency. Accurate load calculations and considering factors such as insulation, building orientation, and occupancy help determine the appropriate system size (Vakiloroaya et al., 2014).

In their study, Li et al. focused on optimizing HVAC systems in passive buildings to improve energy efficiency while maintaining indoor comfort. The researchers modeled and simulated the HVAC system of a passive building in Jinan, China, using TRNSYS, GenOpt, and Java. They achieved a 17.31 % reduction in energy consumption by applying heat storage and return air duct strategies. In addition, optimization techniques using improved PSO and Hooker-Jeeves algorithm further reduced energy usage by 19 % compared to the baseline model. The proposed control and optimization framework demonstrates significant energy saving potential, contributing to advances in building energy saving (Li et al., 2024).

It is evident that the energy performance of buildings is directly related to the optimal modeling of HVAC systems. There are two types of approaches for BPE measurements: traditional deterministic methods and data-driven methods. Traditional methods are challenging and costly. Moreover, the results obtained are often only valid for the specific building analyzed, leading to significant scalability issues. Therefore, data-driven techniques are widely used to improve the efficiency of HVAC systems because they are cost-effective and provide satisfactory results (Grillone et al., 2020). In this study, a data-driven approach was preferred. Through data, machine learning methods can identify relationships between the state variables (inputs and outputs) of the analyzed system without having explicit or detailed knowledge about the system's physical behavior.

Bazazzadeh et al. investigate the impact of climate change on building energy consumption and evaluate the effectiveness of machine learning (ML) in predicting HVAC optimization outcomes. Using future weather data for 6 United Statescities generated from the HadCM3 climate model under the A2 GHG scenario, energy performance simulations were conducted on a Department of Energy office building prototype. By adjusting HVAC setbacks, the study analyzed energy savings and thermal comfort improvements. The Light Gradient Boosting Model Regressor (LGBMR) achieved 88 % accuracy in predicting these improvements, demonstrating ML's potential as a valuable tool for assessing HVAC optimization strategies under future climate conditions (Bazazzadeh et al., 2025).

The Multi-Layer Perceptron (MLP) is widely employed to forecast energy consumption in buildings. Training data-driven models generally involves utilizing historical data concerning external environmental factors and energy usage. The aim of the training process is to identify the pertinent parameters of these models and to investigate the intricate nonlinear relationships between various input and output datasets (Solano et al., 2021; Xu et al., 2022). Training deep learning models like MLP often involves optimization techniques such as backpropagation algorithms and gradient descent. However, metaheuristic optimization algorithms can also be used either in place of or in conjunction with these traditional optimization techniques for training MLPs. Metaheuristic algorithms are particularly useful for large datasets or complex MLP models (Zerouali et al., 2023). These algorithms aim to find global minimum or maximum values of the model, potentially leading to improved performance and faster training times (Eker et al., 2021).

Stochastic algorithms are effective in overcoming the challenge of early prediction of thermal loads. In light of this, Moayedi and Mosavi focused on the evaluation of a new hybrid model for predicting cooling loads (CL) in residential buildings. This model combines artificial neural networks (ANN) with stochastic fractal search (SFS) algorithm. To compare its performance, grasshopper optimization algorithm (GOA) and firefly algorithm (FA) are also considered and compared with SFS. The analysis examines the nonlinear effects of eight independent factors on CL by optimizing the model structure with each algorithm. The results reveal that all three metaheuristic algorithms are effective in optimizing ANN with correlations exceeding 90 %. In particular, the use of GOA, FA and SFS reduced the prediction error by approximately 23 %, 18 % and 36 %, respectively. Furthermore, the performance measurements consistently showed that SFS outperformed the benchmark algorithms (Moayedi and Mosavi, 2021).

Metaheuristic algorithms are commonly used to solve search and optimization problems, possessing the ability to explore a wide search space to find global minimum or maximum values. Therefore, during the training of MLPs, metaheuristic algorithms can be employed to optimize the model parameters (such as weights and biases) (Moayedi et al., 2020; Rojas et al., 2022). Additionally, these algorithms are often used to optimize the hyperparameters or network structure of MLPs, such as the number of hidden layers, learning rate, or activation functions (Chandra and Sharma, 2016; Smithson et al., 2016). In this study, various metaheuristic optimization algorithms were used to optimize these hyperparameters of MLPs for predicting heating and cooling loads in HVAC systems. The use of metaheuristic algorithms varies depending on the specific problem domain and dataset, requiring careful configuration and tuning. Moreover, these algorithms typically involve higher training times and computational costs compared to traditional optimization techniques. Therefore, careful selection and implementation of metaheuristic algorithms are crucial (Eker et al., 2023).

This study aims to optimize the hyperparameters of a neural network using a dataset used for analyzing heating and cooling loads in buildings. An optimized regression model was developed to predict both heating and cooling loads. Various metaheuristic optimization methods were compared for enhancing the MLP regression model. The number of hidden layers, learning rate, and selection of activation functions were optimized using different optimization techniques. The GWO is improved by proposing an entropy-based approach for the search and exploitation balance. The effectiveness of each method was evaluated based on mean squared error (MSE) values to determine which method yielded better results. The impact of each method on the accuracy of the model was demonstrated.

In this era where environmental consciousness is paramount, this study can also be viewed as an effort to reduce our carbon footprint, given that we have not yet transitioned to renewable energy sources. In the second section, we discussed our innovations and contributions. The third section examined similar studies conducted in literature. In the fourth section, we conducted data and statistical analysis. The fifth section explained the methodology and methods used, including MLP, optimization techniques, and our study. The sixth section analyzed and discussed the results. The final section is the conclusion.

2. Main novelties and contributions

Accurate heating and cooling load predictions are crucial for optimizing HVAC system performance and enhancing energy efficiency in buildings. However, optimizing the hyperparameters of Multi-Layer Perceptron (MLP) models remains a significant challenge due to the complexity of parameter tuning and the risk of suboptimal solutions. This study systematically evaluates ten different optimization algorithms for fine-tuning MLP hyperparameters to improve predictive accuracy in HVAC systems. The tested methods include Particle Swarm Optimization (PSO), Genetic Algorithm (GA), Grey Wolf Optimization (GWO), Harris Hawks Optimization (HHO), Whale Optimization Algorithm (WOA), Cross Entropy (CE), Hunting Search (HUS), Biogeography-Based Optimization (BBO), Harmony Search (HSA), and Cuckoo Search. By comparing these algorithms, we provide a comprehensive evaluation of their strengths and weaknesses in enhancing MLP model performance for heating and cooling load predictions.

A major contribution of this study is the development of Entropy-Based Grey Wolf Optimization (ENTGWO), an improved version of GWO designed to enhance search efficiency and convergence reliability. Unlike traditional GWO, ENTGWO dynamically adjusts its nonlinear convergence factor ($\rightarrow$a) based on iteration count and population entropy, improving the balance between exploration and exploitation. This entropy-driven strategy reduces the risk of premature convergence, allowing the algorithm to explore a wider range of solutions and avoid local optima. The proposed ENTGWO framework is not limited to HVAC applications. Its adaptive nature makes it a promising approach for various optimization problems in energy efficiency, computational intelligence, and beyond.

To ensure a rigorous performance evaluation, this study conducts an extensive comparative analysis of all eleven optimization methods applied to MLP hyperparameter tuning. The results demonstrate that different optimization techniques significantly impact prediction accuracy and convergence speed. Notably, ENTGWO outperforms all other methods by achieving lower error rates and faster convergence, proving its effectiveness in enhancing machine learning-based HVAC load prediction.

Beyond immediate applications, this study contributes to the broader field of AI-driven energy optimization by introducing an adaptive, entropy-driven optimization strategy. These findings offer practical recommendations for engineers, researchers, and energy policymakers aiming to develop smarter, more efficient building systems. Furthermore, the ENTGWO framework establishes a strong foundation for future studies exploring metaheuristic-based optimization techniques for a wide range of prediction and classification tasks. The methodology presented here can be adapted to other complex engineering problems, making it a versatile tool for data-driven decision-making in energy efficiency and sustainable development.

3. Literature

Yao et al. (2023) propose an entropy-enhanced Grey Wolf Optimizer (IEGWO) to address issues like low diversity, imbalanced exploration and exploitation, and premature convergence in the standard GWO. The improvements include entropy-based initialization, a dynamic position update strategy to maintain diversity, and a nonlinear convergence method to optimize search balance. Evaluated on CEC2014 and CEC2017 benchmarks, and applied to engineering and real-world problems, the IEGWO demonstrates enhanced accuracy and robustness over comparable algorithms for global optimization tasks.

Kim and Cho (2019) developed a new neural network model called CNN-LSTM to predict residential energy consumption. This model combines two different types of neural networks: Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM). CNN is used to analyze various factors affecting energy consumption, while LSTM models complex patterns and trends in time-series data. They tested the model with a large dataset collected from a real residential setting. The results demonstrated that the CNN-LSTM model achieves higher accuracy, especially for electricity consumption, compared to traditional prediction methods.

Moradzadeh et al. (2020) compared two machine learning methods for predicting heating and cooling loads in buildings: Support Vector Regression (SVR) and Multi-Layer Perceptron (MLP). They found that SVR performed better in predicting cooling loads, while MLP provided better results for heating load prediction. Both methods demonstrated higher accuracy compared to traditional prediction methods.

Roy et al. (2020) proposed a Deep Neural Network (DNN) model for predicting heating and cooling loads in residential buildings. They compared the performance of DNN with other models and found that DNN successfully predicted heating and cooling loads. This suggests that DNN models are potentially accurate options for predicting energy consumption in buildings.

Olu-Ajayi et al. (2022) developed a model using machine learning techniques for residential buildings to predict annual energy consumption. They proposed deep learning, ensemble learning, and other machine learning models for prediction, while investigating the impact of data size on model performance. The significant benefit of their study is providing architects with a model that can predict the annual average energy consumption of a building early in the design process.

Dasi et al. (2024) strengthened support vector regression and extreme gradient boosting methods using metaheuristic algorithms. They compared the accuracy of six meta-heuristic algorithms through different statistical indices. The findings indicate that the combination of Satin Bowerbird Optimizer and extreme gradient boosting was the most effective approach. This hybrid algorithm demonstrated superior performance with minimum error values and consistent convergence across various datasets.

Kumar et al. (2018) propose an innovative method to enhance energy load assessment in buildings. Initially, they explore the connection between building design and structural characteristics with heating and cooling loads. Subsequently, they employ new methods called Extreme Learning Machine (ELM) and Online Sequential ELM (OSELM) to predict these loads. A significant aspect of their study is the real-time prediction capability of OSELM, which allows for online forecasting while data continuously flows. They develop a total of 24 different models using various feature sets and activation functions, comprising 12 ELM and 12 OSELM models. These models' accuracy, computational performance, and efficiency are compared with existing models.

Y. Zhou et al. (2022) introduce a novel workflow for predicting heating and cooling loads of buildings at both individual and neighborhood levels during early design stages. Their workflow involves creating meta-models using seven input variables that significantly influence building loads. Load profiles of different building types such as offices, commercial spaces, and hotels are simulated collectively using EnergyPlus software. They employ a sequence-to-sequence (Seq2Seq) model based on one-dimensional convolutional neural networks (1D-CNN), a deep learning method, to rapidly predict hourly building loads throughout the year. This meta-modeling workflow is applied as a case study on a neighborhood in Shanghai, China. The prediction results exhibit strong correlations with actual loads, achieving R^2 values of 0.9978 and 0.9975 for heating and cooling loads, respectively. The workflow expands the application domain of physics-based methods in building design and enhances the temporal resolution of traditional data-driven methods. It stands out in engineering by providing the required precision and ease of use, with low prediction errors and fast computational speed.

Wang et al. (2023) propose an artificial intelligence-based control strategy for predicting the energy demand of building HVAC systems one day in advance. The method utilizes XGBoost models to forecast energy usage and indoor temperature. Innovative approaches are employed to enhance the strategy's performance, such as customizing the initial population of genetic algorithms, using a progressively decreasing control structure, and separating control action periods with sampling. The strategy is validated on the TRNSYS-Python joint simulation test platform. Results demonstrate that the developed XGBoost models achieve high performance in predicting energy usage (CV-RMSE: 4.52 %) and indoor temperature (CV-RMSE: 0.40 %) for the next 10 minutes using only one week of historical data. Moreover, customizing the initial population reduces computational costs by 78 %.

Jang et al. (2022) compared the impact of operational data on the prediction performance of three different models based on the Long Short-Term Memory (LSTM) algorithm for a non-commercial building. Each model was fed with different types of data inputs: only building ambient data, building ambient data + external environmental data, and building ambient data + external environmental data + operational data. The study found that incorporating additional variables related to building operational habits improved the performance of the prediction models. When information about changes in building operational habits was included in the energy consumption model, it increased the similarity between the operational pattern and the prediction compared to other models. Some studies in the literature for hyperparameter and parameter optimization are shown in Table 1.

4. Data and statistical analysis

In this study, the dataset used for heating and cooling loads of buildings was obtained from the UCI (Lichman, n.d.) data repository, a data mining resource. The dataset consists of 12 different residential buildings with the same volume but varying in parameters such as glazing area, glazing area distribution, and orientation. The dataset includes 768 samples and 10 features. It comprises eight attributes and two responses. The aim is to use the eight features to predict each of the two responses. The attributes include Relative Compactness, Surface Area, Wall Area, Roof Area, Overall Height, Orientation, Glazing Area, and Glazing Area Distribution (Tsanas and Xifara, 2012). The outputs are heating load and cooling load values. The study aims to predict the building's heating and cooling loads based on real values. You can refer to these studies (Bouktif et al., 2018; Tsanas and Xifara, 2012) for more detailed information about the data set.

The summary statistics Table 2 provides insights into the statistical distributions of the features and target variables used in HVAC systems. This data offers a comprehensive understanding of building characteristics and loads that affect the performance of HVAC systems and provides critical information for optimization. It is particularly important for understanding the impact of building features on heating and cooling loads. While some features show variability within a narrow range, others exhibit a wide range of variability.

Fig. 1 illustrates different building structures and their Relative Compactness (RC) values. RC is used to represent different building types and is calculated by comparing the volume-to-surface area ratio of

Table 1
Related studies using ML models with expert-knowledge-based improvements(Lu et al., 2023).

Ref.	Year	Prediction	Model	Optimization	Improvement Type
(Papadopoulos et al., 2018)	2018	HL and CL	DT, RF, GBM, Extra Tree	Grid search	Hyperparameter optimization
(Seyedzadeh et al., 2019)	2019	HL and CL	ANN, SVM, Gaussian process, RF, GBM	Grid search	Hyperparameter optimization
(Seyedzadeh et al., 2020)	2020	HL and CL	RF	GA	Hyperparameter optimization
(Moayedi et al., 2020)	2020	CL	ANN	EHO, ACO, HHO	Parameter optimization
(Z. Guo et al., 2020)	2020	HL and CL	ANN	WOA, SHO, SSA, WDO	Parameter optimization
(G. Zhou et al., 2020)	2020	HL and CL	ANN	ABC, PSO	Parameter optimization
(Huang and Li, 2021)	2021	HL and CL	Wavelet neural network	ABC	Hyperparameter optimization
(Xu et al., 2022)	2022	HL and CL	ANN	GA, PSO, ACO, BBO	Parameter optimization
(Liang et al., 2022)	2022	HL and CL	ANN	HGS; Wrapper-based feature selection	Parameter optimization, Feature engineering
(J. Guo et al., 2023)	2023	HL and CL	LightGBM	Random Search, Grid Search, Tree-Structured Parzen Estimator, Covariance Matrix Adaptation Evolution Strategy	Hyperparameter optimization, Feature engineering
(Eker et al., 2024)	2024	HL and CL	ANN	SSA	Parameter optimization

Table 2
Statistical analysis of dataset.

Features	No. of Possible Values	Descriptive Index								
		Mean	Standard Error	Median	Mode	Standard Deviation	Sample Variance	Skewness	Min	Max
Relative Compactness (X1)	12	0.76	0.00	0.75	0.98	0.11	0.01	0.50	0.62	0.98
Surface Area (X2)	12	671.71	3.18	673.75	514.50	88.09	7759.16	−0.13	514.50	808.50
Wall Area (X3)	7	318.50	1.57	318.50	294.00	43.63	1903.27	0.53	245.00	416.50
Roof Area (X4)	4	176.60	1.63	183.75	220.50	45.17	2039.96	−0.16	110.25	220.50
Overall Height (X5)	2	5.25	0.06	5.25	7.00	1.75	3.07	0.00	3.50	7.00
Orientation (X6)	4	3.50	0.04	3.50	2.00	1.12	1.25	0.00	2.00	5.00
Glazing Area (X7)	4	0.23	0.00	0.25	0.10	0.13	0.02	−0.06	0.00	0.40
Glazing Area distribution (X8)	6	2.81	0.06	3.00	1.00	1.55	2.41	−0.09	0.00	5.00
Heating load (Y1)	586	22.31	0.36	18.95	15.16	10.09	101.81	0.36	6.01	43.10
Cooling load (Y2)	636	24.59	0.34	22.08	21.33	9.51	90.50	0.40	10.90	48.03

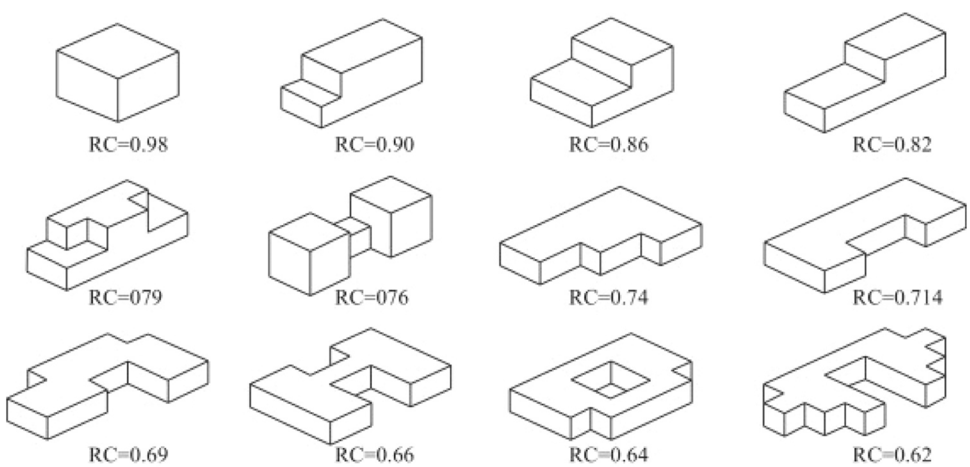


Fig. 1. Building shapes.

a building with that of another building of the same volume but a more compact shape. Relative Compactness is a measure of the surface area relative to the volume of a building and is an important parameter for energy efficiency. A building with a high RC value indicates that it has a maximum volume with a minimum surface area, which generally helps to minimize heat loss and gain. Relative Compactness significantly affects a building's energy performance. Fig. 1 provides valuable insights into understanding how different building designs impact the energy efficiency of HVAC systems (Afzal et al., 2023).

5. Methods and methodology

5.1. Multi-Layer Perceptron

MLP (Multi-Layer Perceptron) is a type of artificial neural network and is widely used as a deep learning model. It has a broad range of applications, including classification, regression, pattern recognition, natural language processing, image processing, and many more (Gardner and Dorling, 1998). Recognized as one of the fundamental building blocks of deep learning, MLP is a robust solution for addressing

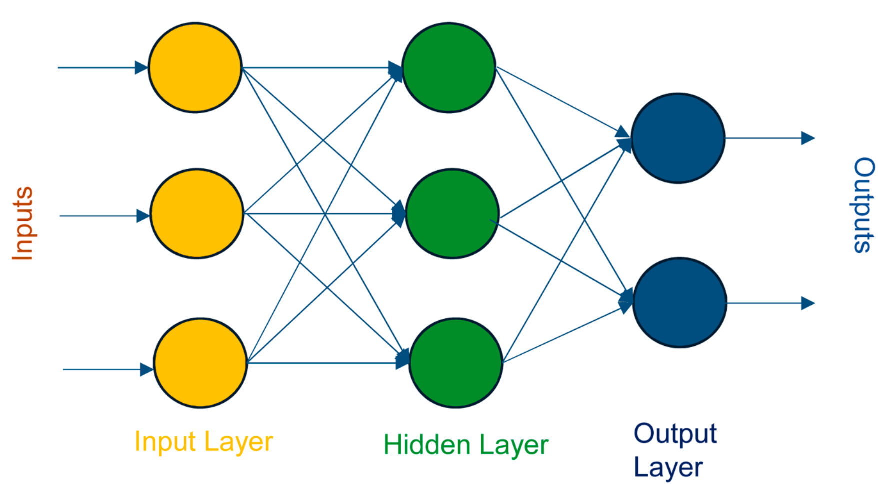


Fig. 2. Schematic of MLP structure.

complex issues. An MLP is composed of an input layer, one or more hidden layers, and an output layer, each containing multiple artificial neurons, also known as perceptrons as illustrated in Fig. 2.

The MLP begins with the input layer. This layer consists of the data or features fed into the model. The input layer comprises nodes representing each feature in the problem. Essentially, it represents the raw data the model receives. The inputs are denoted as $x_1, \quad x_2, \quad x_3, \quad \ldots\ldots, \quad x_n$. For instance, in an image classification problem, each pixel might correspond to an input node.

The MLP can have one or more hidden layers. Each hidden layer consists of interconnected neurons. These hidden layers receive input signals, apply weights, activate them, and pass the outputs to the next layer. Hidden layers are used to learn complex relationships within the data. There are weights associated with connections between each layer, which connect the input and output layers or other hidden layers. The output of a neuron in a hidden layer is calculated according to Eq. (1).

$$z_j = f\left(\sum\nolimits_{i=1}^{n} w_{ij} x_i + b_j\right) \tag{1}$$

Here z_j is the output of the j-th neuron. w_{ij} is the weight from the i-th input to the j-th neuron. x_i is the i-th input. b_j is the bias term for rhe j-th neuron. f is the activation function. Each neuron processes its inputs through an activation function. This function determines the output of the neuron. Commonly used activation functions include sigmoid, ReLU (Rectified Linear Activation), tanh, and softmax.

The final layer of an MLP is the output layer, which contains the model's predictions. For a classification problem, each class might correspond to an output node. In regression problems, there could be a single output node. The output of a neuron in the output layer is computed similarly to the neurons in the hidden layers, as shown in Eq. (2):

$$y_k = f\left(\sum\nolimits_{j=1}^{m} w_{jk} z_j + b_k\right) \tag{2}$$

Eqs. (1) and (2) define the feedforward process of an MLP (Eker et al., 2021; Ogundokun et al., 2022). During training, the weights and bias terms are optimized using the backpropagation algorithm. This optimization process involves updating the weights and bias terms to minimize the error. The training of an MLP is performed using the backpropagation algorithm to minimize the difference between the model's outputs and the input data. In this process, the weights in each layer of the network are updated using an optimization algorithm. For each neuron in the output layer, the error is calculated. By backpropagating the errors, the error signals for each weight and bias term are computed. The weight and bias values are updated based on the error signals. The steps are repeated for a specified number of iterations or until the error falls below a certain threshold. This is how neural networks are trained. The basic backpropagation algorithm for updating weights can be summarized by the following Eq. (3), where n represents the learning rate, and δ_j represents the error term for the j-th neuron (Shadkani et al., 2021; X. Xie et al., 2023).

$$w_{ij} \leftarrow w_{ij} + \eta \delta_j x_i \tag{3}$$

MLPs offer several advantages. Unlike single-layer neural networks, MLPs possess a universal approximation property. This means that with a sufficient number of hidden layers and neurons, an MLP can approximate any continuous function to a desired degree of accuracy. MLPs can learn complex structures in datasets and are highly successful in detecting intricate patterns. This capability enhances their applicability in fields such as image processing, natural language processing, speech recognition, and other complex data analysis applications. Additionally, MLPs can be configured with varying numbers of hidden layers and neurons, making the model flexible and adaptable to various types of problems.

There are several key factors that impact the applicability and performance of MLPs. Depending on the problem to be solved and the characteristics of the available dataset, these factors must be taken into account. Very large and complex MLP models tend to overfit the training data. This means the model fits the training data very well but loses its generalization ability and performs poorly on new data. As the complexity of the network and the number of parameters used during training increase, the risk of overfitting also rises. This can negatively affect the model's overall performance and reduce accuracy when applied to new data. Training MLPs on large datasets and complex models requires significant computational power. Training the model can take a long time and requires efficient use of computational resources. The quality and quantity of the training data directly impact the performance of the MLP. High-quality, well-labeled, and sufficient data help the model learn better and generalize well to new data. The architecture of the MLP, including the number of layers and the number of neurons per layer, influences the model's ability to learn and generalize. Balancing the model complexity to avoid underfitting and overfitting is essential. The choice of activation functions, optimization algorithms, and regularization techniques (such as dropout or weight decay) also play a vital role in the training and performance of MLPs.

There are many hyperparameters involved in configuring MLPs, and they significantly affect the model's accuracy and performance. Proper adjustment and optimization of these hyperparameters are crucial for enhancing the model's success. Hyperparameters are parameters set before the learning process begins and are not updated during training. These configurations encompass parameters like the learning rate, activation functions, number of hidden layers, batch size, number of neurons per layer, and additional settings. Optimizing the architecture and hyperparameters of MLP models can also be viewed as an optimization problem (L. Yang and Shami, 2020). Hyperparameter tuning is a crucial part of the machine learning workflow, and finding the right balance can significantly enhance model performance. It often involves iterative experimentation and careful analysis of results to identify the best set of hyperparameters.

Hyperparameter tuning involves several techniques to optimize the performance of machine learning models. Grid search systematically explores a predefined set of hyperparameter combinations, evaluating each one using cross-validation to find the optimal set that maximizes performance metrics. Random search, on the other hand, randomly samples hyperparameters from predefined distributions, making it more efficient in high-dimensional spaces compared to grid search. Bayesian optimization builds a probabilistic model of the objective function to intelligently select hyperparameters for evaluation, leveraging past evaluations to focus on promising areas. These techniques, along with gradient-based optimization and advanced methods like Hyperband, collectively aim to balance exploration of different hyperparameter configurations with exploitation of promising ones, ultimately improving model accuracy and efficiency (Ogundokun et al., 2022; Sezer et al., 2020).

5.2. Optimization methods

The PSO (Kennedy and Eberhart, 1995), BBO (Simon, 2008), CE (Rubinstein, 1997), Cuckoo Search (X.-S. Yang and Suash Deb, 2009), GA (Holland, 1975; Katoch et al., 2021), WOA (Mirjalili and Lewis, 2016), HSA (Zong Woo Geem et al., 2001), HHO (Heidari et al., 2019), HUS (Oftadeh and Mahjoob, 2009), iGWO (Mirjalili et al., 2014; Nadimi-Shahraki et al., 2021), and the proposed ENTGWO are applied to solve this problem. Each algorithm in our study was run 30 times, with 100 iterations per run. The average MSE error value of the obtained values was calculated. The HL and CL values obtained by optimizing MLP hyperparameters with optimization methods were compared.

5.2.1. Grey Wolf Optimization

Grey Wolf Optimization (GWO) is inspired by the hierarchical structure and social behaviors of grey wolves, which are social animals known for their hunting strategies. The algorithm aims to find the minimum or maximum of an optimization function by modeling the hunting strategies and social behaviors of grey wolves. In GWO, a group of wolves represents a population where each wolf represents a solution (i.e., a position). There exists a hierarchy among these wolves, with alpha (α), beta (β), delta (δ), and omega (ω) representing the four types of grey wolves used to simulate leadership hierarchy mathematically. Alpha is considered the most suitable solution to model the social hierarchy of wolves. Beta and delta represent the second and third best solutions, respectively, while omega represents the rest of the candidate solutions. The algorithm applies three main steps inspired by wolf hunting behaviors: searching for prey, surrounding the prey, and attacking the prey.

$$D_{\alpha,j}^{t+1} = \left| C_1 \times X_{\alpha,j}^{t} - X_{i,j}^{t} \right| \tag{4}$$

$$D_{\beta,j}^{t+1} = \left| C_2 \times X_{\beta,j}^{t} - X_{i,j}^{t} \right| \tag{5}$$

$$D_{\delta,j}^{t+1} = \left| C_3 \times X_{\delta,j}^{t} - X_{i,j}^{t} \right| \tag{6}$$

$$X_{ad1,j}^{t+1} = X_{\alpha,j}^{t} - A_1 \times D_{\alpha,j}^{t+1} \tag{7}$$

$$X_{ad2,j}^{t+1} = X_{\beta,j}^{t} - A_2 \times D_{\beta,j}^{t+1} \tag{8}$$

$$X_{ad3,j}^{t+1} = X_{\delta,j}^{t} - A_3 \times D_{\delta,j}^{t+1} \tag{9}$$

$$X_{i,j}^{t+1} = \left(X_{ad1,j}^{t+1} + X_{ad2,j}^{t+1} + X_{ad3,j}^{t+1}\right) \Big/ 3 \tag{10}$$

$$C = 2 \times rand \tag{11}$$

$$A = (2 \times rand - 1) \times a \tag{12}$$

$$a = 2 \times \left(1 - \frac{1}{Max_{iter}}\right) \tag{13}$$

The equations (4)–(13) use $X_{i,j}^{t}$ to represent the position of the i-th wolf in the j-th dimension at iteration t. The leading wolves, α, β, and δ, are denoted by their positions as X_α, X_β, and X_δ, respectively. The distances to these dominant wolves are represented by D_α, D_β, and D_δ, while the position adjustments relative to α, β, and δ are expressed as X_{ad1}, X_{ad2}, and X_{ad3}. Additionally, A and C are search coefficients used in the position update process, with A_1, A_2, A_3 and C_1, C_2, C_3 being specific instances of these parameters. The term Max_{iter} indicates the total number of iterations allowed. *rand* is a random number in the range [0,1]. a indicates that the exploration rate decreases linearly from 2 to 0 as iterations progress and can be called the linear component (Mirjalili et al., 2014; Nadimi-Shahraki et al., 2021; H. Xie et al., 2020). a linear component is often referred to as the exploration-exploitation trade-off parameter or similar names. When the prey stops moving, the gray wolf completes the hunting process by attacking. For the approaching prey, the value of a is gradually reduced and therefore the fluctuation range of A is also reduced. In other words, as the value of a decreases during the iteration, the corresponding value of A also varies in the interval $[-a, a]$. When $|A| < 1$, the wolves attack the prey (Jin and Fan, 2022).

Wolves move within a search space, with alpha wolves determining the positions of potential prey (i.e., good solutions) and guiding the pack towards them. Beta and delta wolves follow alpha wolves and adjust their positions accordingly. Wolves follow a specific movement strategy aimed at moving towards the good solutions identified by the alpha wolves and updating their positions within the pack. This strategy is employed to discover better solutions and optimize the pack. The movement and position updates of the wolves are repeated until a specific condition is met, continuing until the best solution is found or a predetermined criterion is satisfied.

GWO is effective in addressing complex and multi-dimensional optimization problems. By modeling the social behavior and hunting strategies of wolves, the algorithm provides a natural diversity and leadership mechanism, enabling its application in a wide range of problem domains (Mirjalili et al., 2014). In this study, the improved GWO method proposed in (Nadimi-Shahraki et al., 2021) was used instead of the classical GWO. The improved GWO algorithm was developed to address problems such as lack of population diversity, imbalance between exploration and exploitation, and early convergence, and its effectiveness in optimization tasks was further increased. Our proposed entropy-based GWO algorithm took this to a better level.

5.2.2. Entropy based Grey Wolf Optimization

GWO offers several advantages over classical methods such as PSO and GA, thanks to its multi-leader guided search strategy and dynamically adjustable search scopes. However, it suffers from limitations such as local stagnation, slow convergence, and insufficient fine-tuning around the best leaders (H. Xie et al., 2020).

The linear component (a) parameter plays a crucial role in GWO by controlling the distances of leader wolves (Alpha, Beta, Delta) from the prey (i.e., the optimal solution). It dynamically adjusts the wolves' movements toward or away from the prey, controlling the search intensity and range. This parameter regulates the balance between global exploration (searching new areas) and exploitation (refining existing solutions), a critical aspect of the search performance in metaheuristic algorithms. Typically, linear component decreases linearly from 2 to 0 as iterations progress, gradually narrowing the search space for more focused exploitation (Mirjalili et al., 2014). However, this linear decay pattern in the original GWO algorithm can cause an abrupt reduction in the search area during exploration and insufficient attention to promising regions near the leaders during exploitation.

In this study, the linear component parameter is dynamically adjusted using an entropy-based approach. Entropy is a concept that measures the uncertainty or disorder of a system. Higher entropy means more uncertainty and more variety, while lower entropy means more order and predictability (Tuğal and Karcı, 2019). When entropy, which measures the diversity of the population, is high, the parameter value is increased to enable the algorithm to explore a broader search space. High entropy indicates a more dispersed population, suggesting that leaders should maintain a wider search radius to capitalize on the diversity. This method aims to prevent the algorithm from stagnating in local optima and accelerates the convergence rate.

$$H = \sum_{i=1}^{n} p_i \quad \log_2(p_i) \tag{14}$$

In this study, the following steps are followed to calculate the entropy (H) of a population.

1. **Creating a Histogram:** Population positions are taken. The population (solution set) is divided into cells and the number of solutions in each cell is determined. The histogram is the number of solutions in each cell. In this study, the number of cells was selected as 10 when producing solutions for heating load and cooling load. This value can be played with according to the solution space. The number of cells and the number of populations are important in calculating entropy values. They affect entropy values.
2. **Calculating Probabilities:** The number of solutions in each cell is divided by the total number of solutions. p is the probability values of each cell. Cells with a probability of 0 are removed. It is necessary because the logarithm does not work with 0.
3. **Calculating Entropy:** The entropy of the population is calculated using the Shannon Entropy formula in Eq. (14). A more evenly distributed population (i.e., the probabilities in the cells are more equal) has a higher entropy. Entropy (H) returns a scalar value that measures how diverse the population is.

In this study, the linear component parameter of the GWO is dynamically adjusted based on both the number of iterations and population entropy, deviating from the standard GWO approach. When the entropy, representing population diversity, exceeds a predefined threshold (set as $H > 1.0$ for Y2 and $H > 1.584$ for Y1), the parameter increases ($a = \min(a + 0.5, 2.0)$ with an upper limit of 2.0), allowing broader exploration. Otherwise, it follows the default linear decay. This dual adjustment strategy enhances convergence and mitigates local optima. Fig. 3 shows how a linear component behaves with the change in entropy in our data set. The threshold value here must be chosen correctly for a good solution. Initially, this value was chosen randomly between 0 and 2. The change in entropy was monitored and converted to an average value. Later, better results were obtained with values closer to the average in the experiments. In summary, the proposed GWO variant introduces dynamic search space adjustments with a nonlinear exploration scheme, which expands the search area during exploration and restricts it during exploitation. This improvement enhances the balance between search diversification and intensification, significantly increasing the likelihood of reaching the global optimum.

5.3. Hyperparameter tuning in MLP

Tuning various hyperparameters is vital to improve the performance of an MLP model and achieve optimum results. Optimization techniques play a pivotal role in navigating this expansive search space, enabling the discovery of hyperparameter configurations that allow the model to operate at peak effectiveness. The process of hyperparameter optimization revolves around four key components: an estimator equipped with an objective function, a defined search space encompassing possible parameter values, a chosen optimization method, and an evaluation function. These components cooperate synergistically to determine the optimal set of hyperparameters that optimize the performance of the model. In this study, the hyperparameters that are desired to be optimized (best fit to the dataset) are the number of hidden layers, activation function and learning rate. These parameters significantly affect the performance of the model and require the use of optimization techniques to find their best values. By employing an optimization algorithm that minimizes a specific objective function tailored to a particular problem, we can optimize the hyperparameters of an MLP Regressor model.

Learning Rate: Learning rate is a numerical value that determines how much the neural network weights change in the context of optimization. It defines how fast a network updates its parameters while minimizing the loss function. Therefore, this parameter is important for the optimizer and the loss function (Jena et al., 2021). It controls the magnitude of weight updates in each iteration of training, typically a small positive number, ensuring more stable and controlled weight updates.

Number of Hidden Layers: Specified as a tuple or list, each element represents the number of neurons in a hidden layer. This hyperparameter configures the structure of hidden layers in a MLP, making the internal representation of the model more complex and flexible. The hyperparameter used for hidden layer sizes is used to determine the number of hidden layers and neurons in each layer of an artificial neural network, influencing the model's complexity and capacity. However, improper selection can lead to issues such as overfitting or underfitting.

Activation Function: An activation function is a crucial component in artificial neural networks, governing how input signals are transformed into output signals by individual neurons. It transforms the input value into a specific range, deciding whether the neuron should be activated. Activation functions significantly impact the learning capacity and performance of neural networks. Choices include sigmoid, tanh, ReLU, and softmax activation functions.

Optimizing these hyperparameters effectively enhances the MLP model's capability to learn and generalize patterns from data, improving its overall performance in various applications. In this study Mean Squared Error (MSE) was defined as the objective function to measure the accuracy of various structures of the networks. MSE signifies the average squared difference between predicted values and actual values, commonly used as a metric to evaluate the performance of prediction models. MSE represents the mean of the squared variances between predicted and actual values. It is employed to measure how well a model makes predictions; a low MSE indicates that the model makes predictions with high accuracy, while a high MSE indicates that the model's predictions are inaccurate. MSE is expressed as follows:

$$MSE = \frac{1}{N}\sum_{t=1}^{N}(y_i - \widehat{y}_i)^2 \tag{15}$$

According to the Eq. (15), when measuring the accuracy of a model, N is the number of data points, y_i is actual values, and $\widehat{y}_i$ is predicted values (Tuğal, 2024).

Fig. 4 illustrates the flow diagram of the method used. Using the specified objective function, attempts were made to determine optimal hyperparameters within the specified range using optimization techniques. The study aimed to assess which of the selected optimization methods could more effectively discover suitable hyperparameters up to the specified iteration.

6. Results and discussion

The primary objective of this study is to optimize the performance of the MLP model used to predict heating load and cooling load values in HVAC systems. To achieve this goal, the hyperparameters of the MLP model were optimized using 10 different optimization methods. The study aimed to identify which optimization algorithm could best determine the hyperparameters. The optimization methods used in the study include PSO, GA, GWO, ENTGWO, HHO, WOA, CE, HUS, BBO,

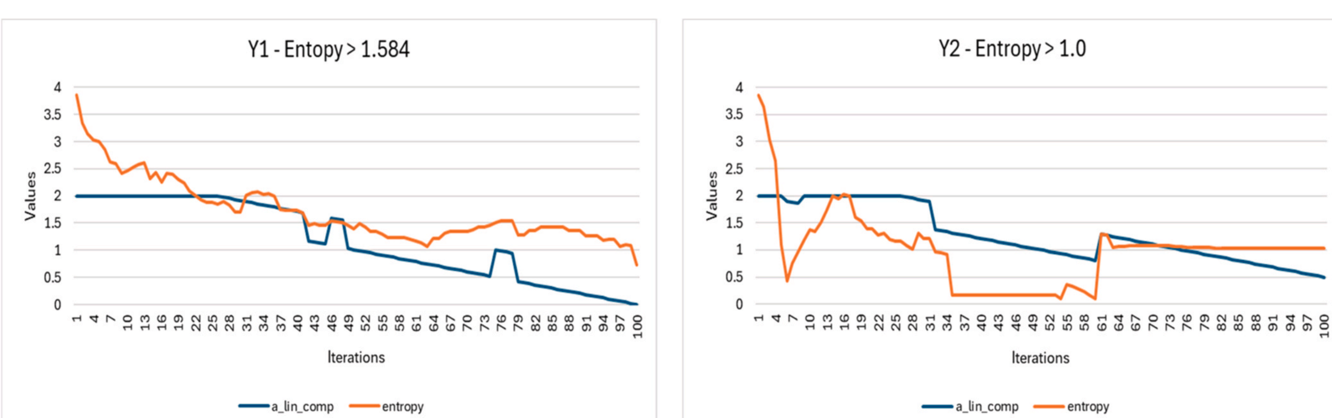


Fig. 3. Entropy based a linear component selection.

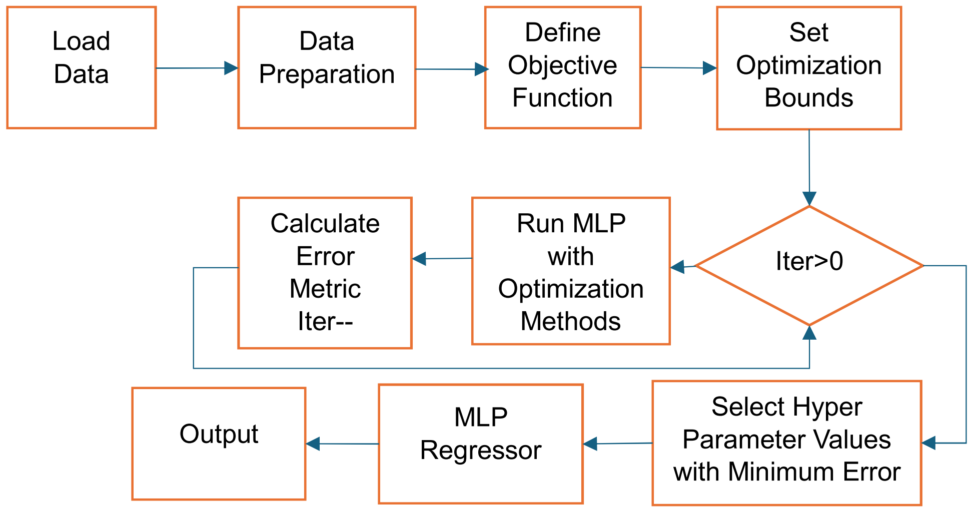


Fig. 4. MLP optimization flow diagram.

HSA, and CUCKOO. The effectiveness of each method was evaluated based on its potential to improve the model's prediction accuracy. 30 % of the dataset, which consists of 768 samples, was reserved for testing. Training and testing data were randomly selected. The data were subjected to normalization.

Table 3 shows the performance of various algorithms and optimization methods used in HVAC system measurements with MSE in the literature. In Table 3, the last row provides the results of our study.

The obtained results have shown that the choice of optimization method significantly impacts the model's performance in predicting HVAC system HL and CL values. Among the eleven optimization algorithms tested, ENTGWO, GWO and WOA demonstrated the highest accuracy, yielding the best performance in terms of MSE. These methods were particularly effective in fine-tuning the hyperparameters of the MLP model, resulting in improved predictive accuracy. In contrast, the PSO algorithm was noted for its rapid convergence, quickly finding optimal or near-optimal solutions. This indicates that while PSO may not achieve the lowest error rates, its speed makes it a viable option for applications where time is a critical factor. The findings of this study align with those reported in the existing literature, further validating the efficacy of ENTGWO, GWO and WOA in optimization tasks. These results suggest that ENTGWO, GWO and WOA are robust methods for enhancing the performance of MLP models in complex, multi-dimensional optimization problems such as HVAC system load predictions. Additionally, the study underscores the importance of selecting an appropriate optimization algorithm to achieve the best possible outcomes in model performance.

Fig. 5 presents the Y1 and Y2 values obtained after 50 iterations. From the graph, it is evident how different methods vary in effectiveness and which methods exhibit better performance with lower error rates for Y1 and Y2. Specifically, the ENTGWO, GWO, WOA, and PSO algorithms achieved significantly lower error values, indicating their superior performance. Conversely, the HSA method performed poorly, particularly for the Y2 value. This comparison underscores the importance of selecting an appropriate optimization method to enhance the predictive accuracy of the model.

Fig. 6 shows the iteration count on the horizontal axis and the MSE values on the vertical axis. When examining the performance of the optimization algorithms for HL (Y1), it is evident that the MSE values decrease with the number of iterations. This indicates that the optimization algorithms improve the solution quality over time. Fig. 6 illustrates how the MSE values change with the number of iterations for each algorithm. The PSO algorithm demonstrates a faster initial performance compared to other algorithms, quickly reducing the MSE values. On the other hand, the BBO algorithm shows higher MSE values and relatively poorer performance. Similar performances are observed among the GA, CUCKOO, HHO, and WOA algorithms, which generally achieve comparable MSE values. The HSA algorithm exhibits minimal reduction in MSE values with iterations, indicating less effectiveness. Although the CE and HUS algorithms perform well, their changes in MSE values from the beginning are not substantial. The ENTGWO algorithm achieves the lowest MSE values, showing a continuous reduction in MSE values as iterations progress. In our HVAC model, ENTGWO has yielded more successful results for HL, consistently demonstrating improved performance and lower MSE values compared to other algorithms. This analysis highlights the effectiveness of the ENTGWO algorithm in enhancing model accuracy over iterative optimization processes.

Table 4 compares the performance of the algorithms, making it easier to identify which algorithm performs best on average and which has the most consistent performance. Mean performance indicates the average performance of the algorithm over all iterations. Standard deviation indicates the variability in the performance of the algorithm over all iterations. The algorithm with the lowest standard deviation is more consistent. The summary statistics with Table 4 provides a quantitative comparison.

When considering the CL (Y2) problem, the performance of the HSA algorithm remains lower compared to others, showing minimal improvement over iterations. The PSO and WOA algorithms appear as the next best performers after ENTGWO and GWO, exhibiting robust performance trends. Although GA initially performs well in early iterations, it converges to similar MSE values as CE in later iterations. The CUCKOO algorithm stabilizes after initial iterations. BBO and HUS algorithms start with a relatively slower pace, demonstrating stable performance but not achieving low MSE values. The MSE values of the GWO algorithm become more stable and consistently low as iterations progress, indicating both rapid achievement of good results and maintenance of these results over time. Compared to other well-performing algorithms like PSO, WOA, GA, and CE, GWO, ENTGWO achieves lower MSE values both in initial and overall performance. This underscores ENTGWO as the most effective optimization algorithm for the CL problem, demonstrating its ability to consistently achieve and maintain superior results.

Fig. 7 illustrates the difference in error between the first and last iterations. A small change indicates low performance of the optimization algorithm in searching for hyperparameter values. In HL prediction for 100 iterations, the HUS, HSA, and CE algorithms yield the smallest error difference, indicating that they are less effective in optimizing hyperparameters compared to others. GWO (0.0831) and ENTGWO (0.1165) show the highest changes, indicating stronger optimization. The same behavior can also be observed for optimization algorithms in 50

Table 3
The error values of HL and CL obtained with MSE using different methods.

Ref.	Year	Model	Evaluation Metrics	Results
(Navarro-Gonzalez and Villacampa, 2019)	2019	Octahedric Regression (OR)	MSE	HL: 2.289, CL: 2.731
(Sajjad et al., 2020)	2020	Gated Recurrent Unit(GRU)	MSE	HL: 0.7215, CL: 0.9791
(Moradzadeh et al., 2020)	2020	MLP SVR	MSE	MLP – HL: 0.2335, CL: 6.896 SVR – HL: 0.7838, CL: 3.024
(Huang and Li, 2021)	2021	ACO	MSE	ACO – HL: 3.478, CL: 1.9291
(Chaganti et al., 2022)	2022	LSTM CNN LSTM-CNN	MSE	LSTM - HL: 1.73, CL: 1.92 CNN - HL: 2.17, CL: 2.03 LSTM - CNN HL: 2.33, CL: 2.18
(Kavitha et al., 2022)	2022	HHO	MSE	HHO- HL: 0.90572, CL: 0.81743
(Pachauri and Ahn, 2022)	2022	WCA PSO TLBO SFLA	MSE	WCA – HL: 0.2442, CL: 0.5441 PSO – HL: 0.2205, CL: 0.4794 TLBO – HL: 0.2075, CL: 0.4794 SFLA – HL: 0.2048, CL: 0.4772
(Salami et al., 2023)	2023	LR DT RF	MSE	LR – HL: 9.2129, CL: 9.8814 DT – HL: 0.3852, CL: 4.0697 RF – HL: 0.2395, CL: 2.9853
Our study	2025	PSO BBO CUCKOO HHO GA WOA CE HUS HSA GWO ENTGWO	MSE	PSO- HL: 0.2141, CL: 0.7805 BBO- HL: 0.2406, CL: 0.9914 CUCKOO- HL: 0.2270, CL: 0.9336 HHO- HL: 0.2303, CL: 0.9101 GA- HL: 0.2116, CL: 0.8236 WOA- HL: 0.2086, CL: 0.7537 CE- HL: 0.2105, CL: 0.8244 HUS- HL: 0.2194, CL: 1.0088 HSA- HL: 0.2607, CL: 1.3600 GWO- HL: 0.2007, CL: 0.7346 ENTGWO- HL: 0.1994, CL: 0.7215

iterations. ENTGWO performed well for HL. In CL, HHO (4.046–4.0834) shows an abnormally high difference, which suggests instability or an issue with its optimization process for 50 and 100 iterations. HHO had difficulty stabilizing during iterations. HHO seems to be more successful for CL since it starts the search with a high MSE error value. However, it can be seen in Fig. 6 that it does not have a lower error metric at the end of the iteration. For CL, GWO and ENTGWO again showed strong optimization. Overall, GWO and ENTGWO appear to be the most effective metaheuristic algorithms for optimizing ANN hyperparameters in this study.

Fig. 8 contains the MSE error rates of the optimization algorithms after 100 iterations. It shows the distribution of the methods, their central tendency (median), quartiles (IQR) and possible outliers. It allows us to compare the statistical distribution of the performance of each method. A high median value indicates that the methods generally give better results, while a wide IQR or a large number of outliers may indicate inconsistency or imbalance of the method. When Fig. 8 is examined, the algorithm that makes the most errors is HSA. The algorithm with the most outliers is CUCKOO for HL and CL. HUS and CE have no outliers for HL. HSA has no outliers for CL. The error rate is stable. The algorithm with the lowest error and good performance is ENTGWO. It can be said that the proposed method is generally successful for HL and CL.

The advantages of this approach can be listed as follows. The integration of ENTGWO with MLP improves prediction performance compared to other traditional optimization methods, reducing prediction error. Metaheuristic algorithms improve MLP training by optimizing hyperparameters without requiring exhaustive manual tuning. The proposed method is well-suited for complex, multi-dimensional data where traditional regression models may struggle. The entropy-based search mechanism of ENTGWO enhances global exploration, avoiding local optima that may hinder learning. The ENTGWO model outperformed other models, demonstrating its reliability and efficiency for this type of predictive modeling.

The disadvantages of this approach can be listed as follows. Metaheuristic algorithms require additional computational resources compared to traditional methods. The performance of the optimization models depends on the proper tuning of optimization parameters, such as population size and iteration limits. While ENTGWO is less prone to local optima, poor parameter selection may affect its ability to find the optimal solution efficiently.

Based on these evaluations, we can conclude that the ENTGWO algorithm demonstrated the best performance and achieved the lowest MSE values for both HL and CL problems. It is observed that algorithms achieved lower MSE values for HL. This underscores ENTGWO's effectiveness in such optimization problems. However, it's important to note that this study is limited to specific datasets and hyperparameter ranges, and different datasets and parameters may yield different results.

7. Conclusion

This study underscores the critical role of accurate load prediction in HVAC systems for optimizing energy efficiency and sustainability in buildings. The proposed entropy-based Grey Wolf Optimization (ENTGWO) algorithm effectively optimized the hyperparameters of a Multi-Layer Perceptron (MLP) model, achieving significant

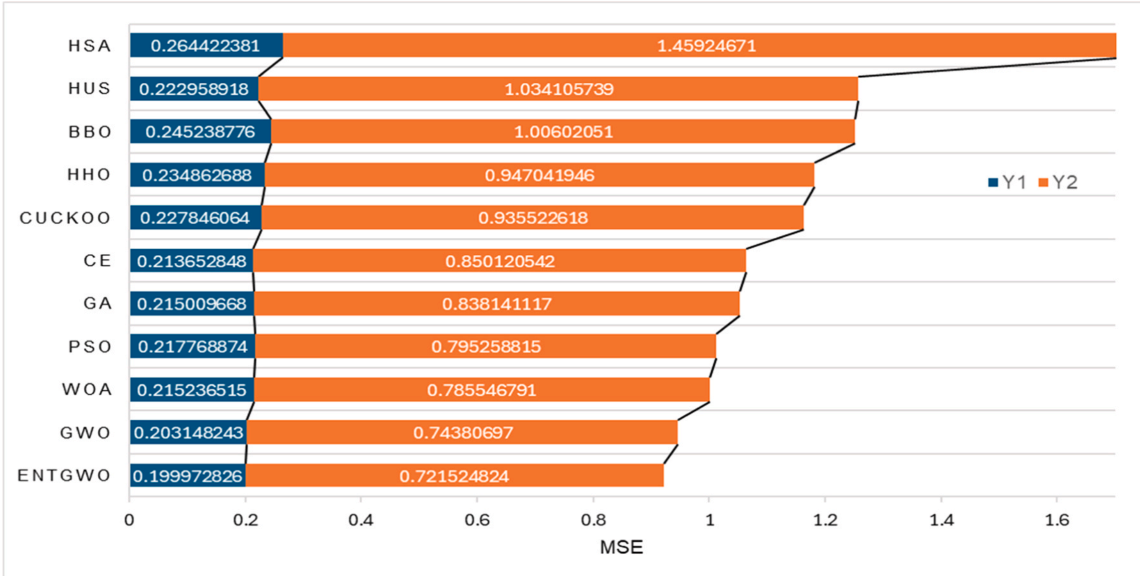
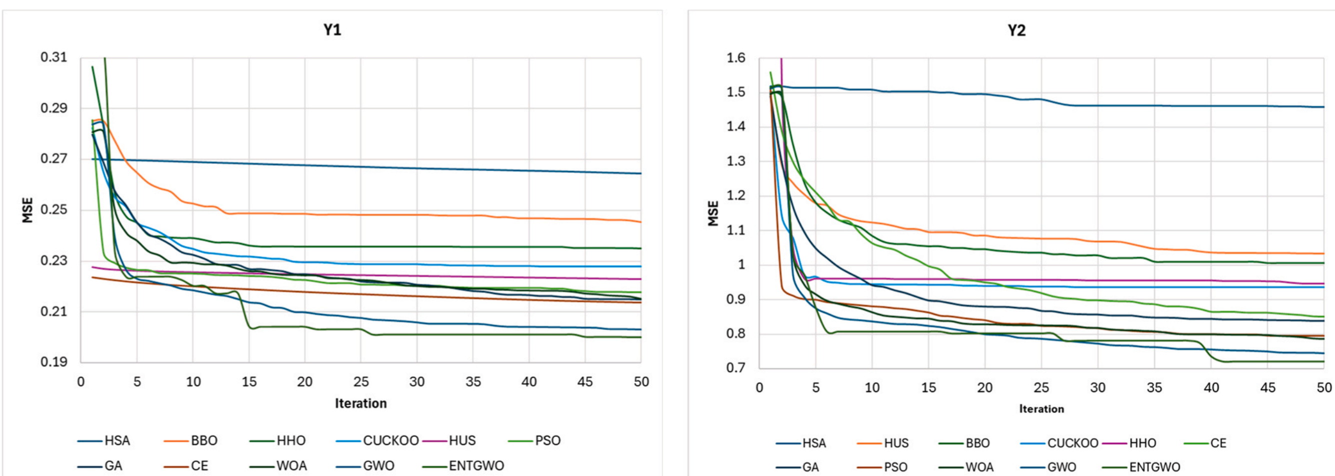


Fig. 5. Average minimum MSE value across 50 iterations for each optimization method.



(a) Heating Load (Y1) (b) Cooling Load (Y2)

Fig. 6. Convergence of optimization algorithms based on MSE values over 50 iterations.

Table 4
Mean performance and standard deviation of each algorithm.

Algorithm	Y1 (Heating Load)		Y2 (Cooling Load)	
	Mean Performance	Standard Deviation	Mean Performance	Standard Deviation
PSO	0.219280	0.007830	0.818228	0.079242
BBO	0.246468	0.008293	1.035371	0.084117
CUCKOO	0.230153	0.007944	0.946122	0.061686
HHO	0.235321	0.009538	0.985205	**0.410005**
GA	0.219824	0.011764	0.870127	0.096188
WOA	0.218423	0.012024	0.813566	0.107628
CE	0.214612	0.003374	0.907120	0.127906
GWO	0.207319	0.012967	0.779528	0.113694
HUS	0.222992	0.001931	1.058067	0.067660
HSA	0.264795	0.002738	1.442880	0.047586
ENTGWO	**0.205303**	**0.017534**	**0.770103**	0.118993

improvements in both heating and cooling load predictions. By dynamically adjusting the linear component based on entropy and iteration count, ENTGWO demonstrated superior performance in preventing premature convergence and enhancing the search capability of the algorithm. The comparative analysis with other metaheuristic approaches, including GWO and WOA, revealed that ENTGWO consistently achieved lower prediction errors and faster convergence, making it a promising approach for HVAC system optimization.

The findings of this study hold substantial theoretical and practical implications. From a theoretical perspective, the integration of entropy-based adjustments in swarm intelligence algorithms presents a novel approach to balancing exploration and exploitation in optimization processes. The results suggest that dynamically adapting control parameters based on entropy can enhance optimization performance across various machine learning applications beyond HVAC load prediction. Practically, the successful application of ENTGWO to HVAC systems offers a valuable tool for engineers and building designers to improve energy management strategies. The reduction in prediction error directly translates to more accurate system sizing, reduced operational costs, and lower carbon footprints, which are crucial for achieving energy efficiency targets in the built environment.

Despite these promising results, certain limitations must be acknowledged. The study focused on optimizing an MLP-based

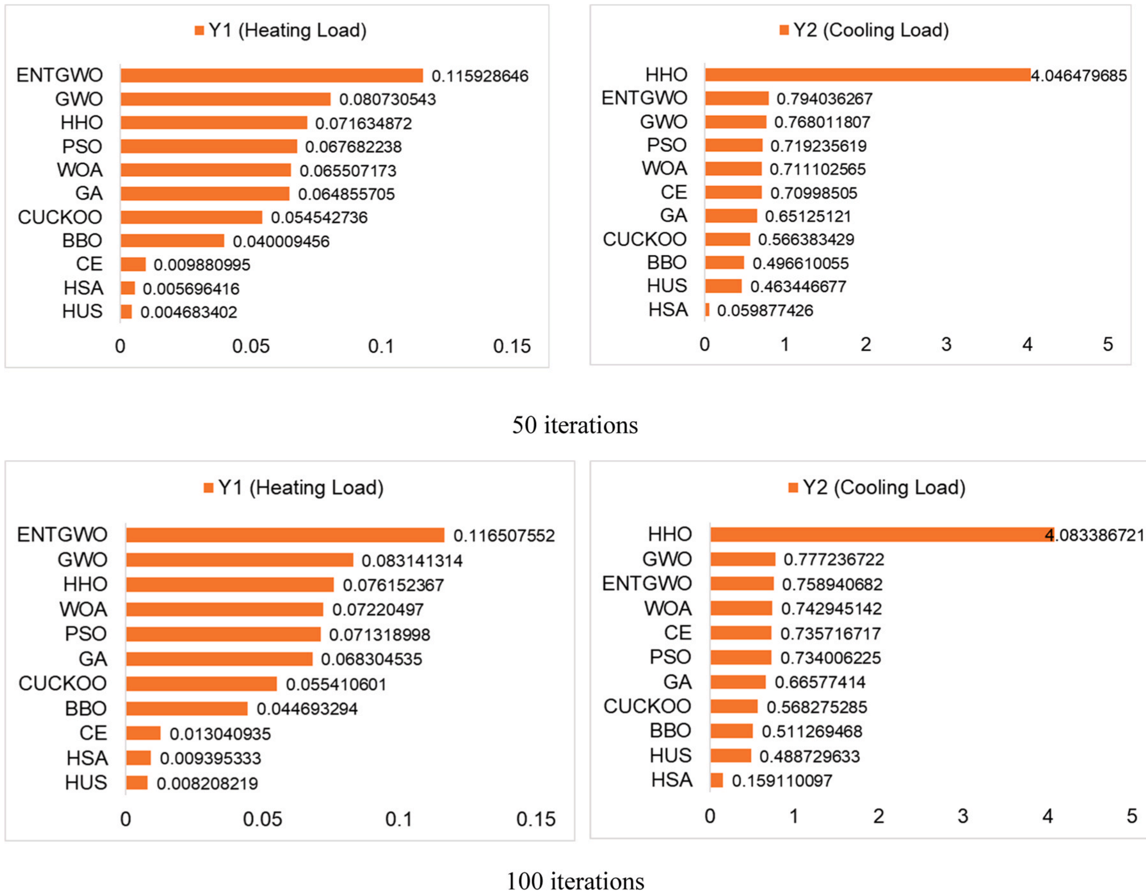


Fig. 7. Success of algorithms in search space from the beginning with 50 and 100 iterations.

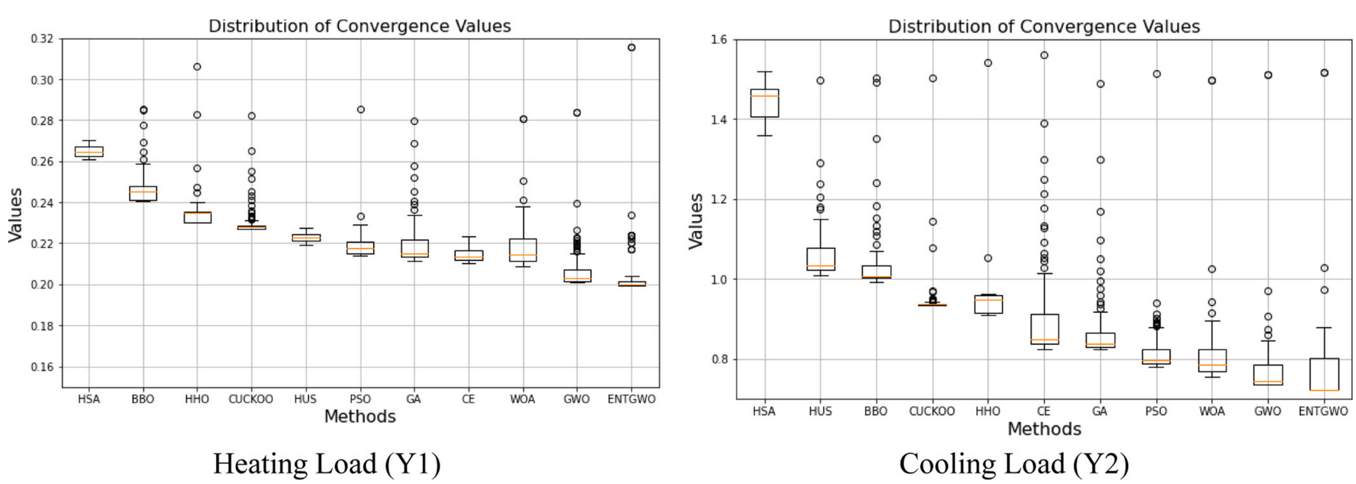


Fig. 8. Success of algorithms in search space from the beginning.

prediction model, and while ENTGWO proved effective in this context, further validation is necessary across different deep learning architectures and real-world datasets. Additionally, the proposed model was tested under static conditions, and its adaptability to dynamic environmental variations remains an open research question. Future research should explore the application of ENTGWO in real-time HVAC control systems, integrating IoT-based sensor data for continuous optimization. Furthermore, expanding the framework to include hybrid optimization techniques, such as combining ENTGWO with deep reinforcement learning, could further enhance predictive accuracy and system adaptability.

In conclusion, this study provides strong evidence that entropy-driven metaheuristic optimization can significantly improve the predictive accuracy of ANN-based models in energy-related applications. The proposed ENTGWO model not only refines HVAC load forecasting but also sets the foundation for future advancements in AI-driven energy efficiency solutions. By extending this approach to broader domains, such as smart grid management and renewable energy forecasting, researchers and practitioners can leverage its capabilities to drive sustainable technological innovations in the energy sector.

CRediT authorship contribution statement

İhsan Tuğal: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Funding

No funding was received for conducting this study.

Declaration of Competing Interest

The authors have no competing interests to declare that are relevant to the content of this article.

Data availability

Data will be made available on request.

References

Afzal, S., Ziapour, B.M., Shokri, A., Shakibi, H., Sobhani, B., 2023. Building energy consumption prediction using multilayer perceptron neural network-assisted models; comparison of different optimization algorithms. Energy 282, 128446. https://doi.org/10.1016/j.energy.2023.128446.

Aloshan, M., Aldali, K., 2024. Empirical study of facade retrofits for optimizing energy efficiency and cooling in school buildings in Saudi Arabia. Energy Rep. 12, 4105–4128. https://doi.org/10.1016/j.egyr.2024.09.076.

Bazazzadeh, H., Hoseinzadeh, S., Mohammadi, M.M., Garcia, D.A., 2025. AI- aided surrogate model for prediction of HVAC optimization strategies in future conditions in the face of climate change. Energy Rep. 13, 1834–1845. https://doi.org/10.1016/j.egyr.2025.01.033.

Bouktif, S., Fiaz, A., Ouni, A., Serhani, M., 2018. Optimal Deep Learning LSTM model for electric load forecasting using feature selection and genetic algorithm: comparison with machine learning approaches. Energies 11 (7), 1636. https://doi.org/10.3390/en11071636.

Calotă, R., Bode, F., Souliotis, M., Croitoru, C., Fokaides, P.A., 2024. Bridging the gap: discrepancies in energy efficiency and smart readiness of buildings. Energy Rep. 12, 5886–5898. https://doi.org/10.1016/j.egyr.2024.11.060.

Chaganti, R., Rustam, F., Daghriri, T., Díez, I., de la, 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.

Chandra, B., Sharma, R.K., 2016. Deep learning with adaptive learning rate using laplacian score. Expert Syst. Appl. 63, 1–7. https://doi.org/10.1016/j.eswa.2016.05.022.

Dasi, H., Ying, Z., Ashab, M.F. Bin, 2024. Proposing hybrid prediction approaches with the integration of machine learning models and metaheuristic algorithms to forecast the cooling and heating load of buildings. Energy 291, 130297. https://doi.org/10.1016/j.energy.2024.130297.

Dermentzis, G., Ochs, F., Franzoi, N., 2021. Four years monitoring of heat pump, solar thermal and PV system in two net-zero energy multi-family buildings. J. Build. Eng. 43, 103199. https://doi.org/10.1016/j.jobe.2021.103199.

Eker, E., Atar, Ş., Şevgin, F., Tuğal, İ., 2024. Optimization of non-linear problems using salp swarm algorithm and solving the energy efficiency problem of buildings with salp swarm algorithm-based multi-layer perceptron algorithm. Electrica 24 (2), 436–449. https://doi.org/10.5152/electrica.2024.23193.

Eker, E., Kayri, M., Ekinci, S., Izci, D., 2021. A new fusion of ASO with SA algorithm and its applications to MLP training and DC motor speed control. Arab. J. Sci. Eng. 46 (4), 3889–3911. https://doi.org/10.1007/s13369-020-05228-5.

Eker, E., Kayrı, M., Ekinci, S., Kacmaz, M.A., 2023. Performance evaluation of PDO algorithm through benchmark functions and MLP training. Electrica 23 (3), 597–606. https://doi.org/10.5152/electr.2023.22179.

Fan, Y., Li, N., He, Y., Zhang, Y., 2024. User preference-driven optimization method for building retrofit planning using a neural network discrepancy model. Energy Rep. 12, 4241–4251. https://doi.org/10.1016/j.egyr.2024.09.061.

Gardner, M., Dorling, S., 1998. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. Atmos. Environ. 32 (14–15), 2627–2636. https://doi.org/10.1016/S1352-2310(97)00447-0.

Grillone, B., Danov, S., Sumper, A., Cipriano, J., Mor, G., 2020. A review of deterministic and data-driven methods to quantify energy efficiency savings and to predict retrofitting scenarios in buildings. Renew. Sustain. Energy Rev. 131, 110027. https://doi.org/10.1016/j.rser.2020.110027.

Guo, Z., Moayedi, H., Foong, L.K., Bahiraei, M., 2020. Optimal modification of heating, ventilation, and air conditioning system performances in residential buildings using the integration of metaheuristic optimization and neural computing. Energy Build. 214, 109866. https://doi.org/10.1016/j.enbuild.2020.109866.

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. Build. Environ. 236, 110252. https://doi.org/10.1016/j.buildenv.2023.110252.

Heidari, A.A., Mirjalili, S., Faris, H., Aljarah, I., Mafarja, M., Chen, H., 2019. Harris hawks optimization: algorithm and applications. Future Gener. Comput. Syst. 97, 849–872. https://doi.org/10.1016/j.future.2019.02.028.

Holland, J., 1975. Adaptation in Natural and Artificial Systems. University of Michigan Press.

Huang, Y., Li, C., 2021. Accurate heating, ventilation and air conditioning system load prediction for residential buildings using improved ant colony optimization and wavelet neural network. J. Build. Eng. 35, 101972. https://doi.org/10.1016/j.jobe.2020.101972.

IEA, 2023. Track. Build. 2023. ⟨https://www.iea.org/energy-system/buildings⟩.

Jang, J., Han, J., Leigh, S.-B., 2022. Prediction of heating energy consumption with operation pattern variables for non-residential buildings using LSTM networks. Energy Build. 255, 111647. https://doi.org/10.1016/j.enbuild.2021.111647.

Jena, B., Saxena, S., Nayak, G.K., Saba, L., Sharma, N., Suri, J.S., 2021. Artificial intelligence-based hybrid deep learning models for image classification: the first narrative review. Comput. Biol. Med. 137, 104803. https://doi.org/10.1016/j.compbiomed.2021.104803.

Jin, J., Fan, J., 2022. Renyi entropy threshold selection on circular histograms based on GWO algorithm. 4th Int. Conf. Nat. Lang. Process. (ICNLP) 2022, 125–132. https://doi.org/10.1109/ICNLP55136.2022.00028.

Katoch, S., Chauhan, S.S., Kumar, V., 2021. A review on genetic algorithm: past, present, and future. Multimed. Tools Appl. 80 (5), 8091–8126. https://doi.org/10.1007/s11042-020-10139-6.

Kavitha, R.J., Thiagarajan, C., Priya, P.I., Anand, A.V., Al-Ammar, E.A., Santhamoorthy, M., Chandramohan, P., 2022. Improved Harris Hawks optimization with hybrid deep learning based heating and cooling load prediction on residential buildings. Chemosphere 309, 136525. https://doi.org/10.1016/j.chemosphere.2022.136525.

Kennedy, J., Eberhart, R., 1995. Particle swarm optimization. Proc. ICNN'95 - Int. Conf. Neural Netw. 4, 1942–1948. https://doi.org/10.1109/ICNN.1995.488968.

Khanlari, A., Sözen, A., Sahin, B., Di Nicola, G., Afshari, F., 2020. Experimental investigation on using building shower drain water as a heat source for heat pump systems. Energy Sources Part A: Recov. Util. Environ. Eff. 1–13. https://doi.org/10.1080/15567036.2020.1796845.

Kim, T.-Y., Cho, S.-B., 2019. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 182, 72–81. https://doi.org/10.1016/j.energy.2019.05.230.

Kumar, S., Pal, S.K., Singh, R.P., 2018. A novel method based on extreme learning machine to predict heating and cooling load through design and structural attributes. Energy Build. 176, 275–286. https://doi.org/10.1016/j.enbuild.2018.06.056.

Li, J., Wang, M., Fang, P., Zhao, Q., Zheng, X., Gao, H., 2024. Modeling and optimization of a passive building HVAC system based on improved PSO and H-J Algorithm. Energy Rep. 11, 2415–2438. https://doi.org/10.1016/j.egyr.2024.02.011.

Liang, R., Le-Hung, T., Nguyen-Thoi, T., 2022. Energy consumption prediction of air-conditioning systems in eco-buildings using hunger games search optimization-based artificial neural network model. J. Build. Eng. 59, 105087. https://doi.org/10.1016/j.jobe.2022.105087.

Lichman, M.. (n.d.). UCI Machine Learning Repository [http://archive.ics.uci.edu/ML]. Irvine, CA: University of California, School of Information and Computer Science. Retrieved 4 January 2018, from http://archive.ics.uci.edu/ML.

Liu, L., Chen, T., Yin, Y., 2016. Energy consumption and quality of life: energy efficiency index. Energy Procedia 88, 224–229. https://doi.org/10.1016/j.egypro.2016.06.152.

Lu, C., Li, S., Reddy Penaka, S., Olofsson, T., 2023. Automated machine learning-based framework of heating and cooling load prediction for quick residential building design. Energy 274, 127334. https://doi.org/10.1016/j.energy.2023.127334.

Ming, R., Yu, W., Zhao, X., Liu, Y., Li, B., Essah, E., Yao, R., 2020. Assessing energy saving potentials of office buildings based on adaptive thermal comfort using a tracking-based method. Energy Build. 208, 109611. https://doi.org/10.1016/j.enbuild.2019.109611.

Mirjalili, S., Lewis, A., 2016. The whale optimization algorithm. Adv. Eng. Softw. 95, 51–67. https://doi.org/10.1016/j.advengsoft.2016.01.008.

Mirjalili, S., Mirjalili, S.M., Lewis, A., 2014. Grey wolf optimizer. Adv. Eng. Softw. 69, 46–61. https://doi.org/10.1016/j.advengsoft.2013.12.007.

Moayedi, H., Mosavi, A., 2021. Suggesting a stochastic fractal search paradigm in combination with artificial neural network for early prediction of cooling load in residential buildings. Energies 14 (6), 1649. https://doi.org/10.3390/en14061649.

Moayedi, H., Mu'azu, M.A., Foong, L.K., 2020. Novel swarm-based approach for predicting the cooling load of residential buildings based on social behavior of elephant herds. Energy Build. 206, 109579. https://doi.org/10.1016/j.enbuild.2019.109579.

Moradzadeh, A., Mansour-Saatloo, A., Mohammadi-Ivatloo, B., Anvari-Moghaddam, A., 2020. Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings. Appl. Sci. 10 (11), 3829. https://doi.org/10.3390/app10113829.

Nadimi-Shahraki, M.H., Taghian, S., Mirjalili, S., 2021. An improved grey wolf optimizer for solving engineering problems. Expert Syst. Appl. 166, 113917. https://doi.org/10.1016/j.eswa.2020.113917.

Navarro-Gonzalez, Villacampa, 2019. An octahedric regression model of energy efficiency on residential buildings. Appl. Sci. 9 (22), 4978. https://doi.org/10.3390/app9224978.

Nicol, J.F., Humphreys, M.A., 2002. Adaptive thermal comfort and sustainable thermal standards for buildings. Energy Build. 34 (6), 563–572. https://doi.org/10.1016/S0378-7788(02)00006-3.

Oftadeh, R., Mahjoob, M.J., 2009. A new meta-heuristic optimization algorithm: hunting search. Fifth Int. Conf. Soft Comput., Comput. Words Percept. Syst. Anal. Decis. Control 2009, 1–5. https://doi.org/10.1109/ICSCCW.2009.5379451.

Ogundokun, R.O., Misra, S., Douglas, M., Damaševičius, R., Maskeliūnas, R., 2022. Medical Internet-of-Things based breast cancer diagnosis using hyperparameter-optimized neural networks. Future Internet 14 (5), 153. https://doi.org/10.3390/fi14050153.

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Sunmola, F., Ajayi, S., 2022. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. J. Build. Eng. 45, 103406. https://doi.org/10.1016/j.jobe.2021.103406.

Pachauri, N., Ahn, C.W., 2022. Regression tree ensemble learning-based prediction of the heating and cooling loads of residential buildings. Build. Simul. 15 (11), 2003–2017. https://doi.org/10.1007/s12273-022-0908-x.

Papadopoulos, S., Azar, E., Woon, W.-L., Kontokosta, C.E., 2018. Evaluation of tree-based ensemble learning algorithms for building energy performance estimation. J. Build. Perform. Simul. 11 (3), 322–332. https://doi.org/10.1080/19401493.2017.1354919.

Peng, Y., Shen, H., Tang, X., Zhang, S., Zhao, J., Liu, Y., Nie, Y., 2023. Energy consumption optimization for heating, ventilation and air conditioning systems based on deep reinforcement learning. IEEE Access 11, 88265–88277. https://doi.org/10.1109/ACCESS.2023.3305683.

Rojas, M.G., Olivera, A.C., Vidal, P.J., 2022. Optimising multilayer perceptron weights and biases through a cellular genetic algorithm for medical data classification. Array 14, 100173. https://doi.org/10.1016/j.array.2022.100173.

Roy, S.S., Samui, P., Nagtode, I., Jain, H., Shivaramakrishnan, V., Mohammadi-ivatloo, B., 2020. Forecasting heating and cooling loads of buildings: a comparative performance analysis. J. Ambient Intell. Humaniz. Comput. 11 (3), 1253–1264. https://doi.org/10.1007/s12652-019-01317-y.

Rubinstein, R.Y., 1997. Optimization of computer simulation models with rare events. Eur. J. Oper. Res. 99 (1), 89–112. https://doi.org/10.1016/S0377-2217(96)00385-2.

Sajjad, M., Khan, S.U., Khan, N., Haq, I.U., Ullah, A., Lee, M.Y., Baik, S.W., 2020. Towards efficient building designing: heating and cooling load prediction via multi-output model. Sensors 20 (22), 6419. https://doi.org/10.3390/s20226419.

Salami, B.A., Abba, S.I., Adewumi, A.A., Dodo, U.A., Otukogbe, G.K., Oyedele, L.O., 2023. Building energy loads prediction using bayesian-based metaheuristic optimized-explainable tree-based model. Case Stud. Constr. Mater. 19, e02676. https://doi.org/10.1016/j.cscm.2023.e02676.

Seyedzadeh, S., Pour Rahimian, F., Rastogi, P., Glesk, I., 2019. Tuning machine learning models for prediction of building energy loads. Sustain. Cities Soc. 47, 101484. https://doi.org/10.1016/j.scs.2019.101484.

Seyedzadeh, S., Rahimian, F.P., Oliver, S., Glesk, I., Kumar, B., 2020. Data driven model improved by multi-objective optimisation for prediction of building energy loads. Autom. Constr. 116, 103188. https://doi.org/10.1016/j.autcon.2020.103188.

Sezer, O.B., Gudelek, M.U., Ozbayoglu, A.M., 2020. Financial time series forecasting with deep learning: a systematic literature review: 2005–2019. Appl. Soft Comput. 90, 106181. https://doi.org/10.1016/j.asoc.2020.106181.

Shadkani, S., Abbaspour, A., Samadianfard, S., Hashemi, S., Mosavi, A., Band, S.S., 2021. Comparative study of multilayer perceptron-stochastic gradient descent and gradient boosted trees for predicting daily suspended sediment load: the case study of the Mississippi River, U.S. Int. J. Sediment. Res. 36 (4), 512–523. https://doi.org/10.1016/j.ijsrc.2020.10.001.

Simon, D., 2008. Biogeography-based optimization. IEEE Trans. Evolut. Comput. 12 (6), 702–713. https://doi.org/10.1109/TEVC.2008.919004.

Smithson, S.C., Yang, G., Gross, W.J., Meyer, B.H., 2016. Neural networks designing neural networks. Proc. 35th Int. Conf. Comput. -Aided Des. 1–8. https://doi.org/10.1145/2966986.2967058.

Solano, J.C., Caamaño-Martín, E., Olivieri, L., Almeida-Galárraga, D., 2021. HVAC systems and thermal comfort in buildings climate control: an experimental case study. Energy Rep. 7, 269–277. https://doi.org/10.1016/j.egyr.2021.06.045.

Tsanas, A., Xifara, A., 2012. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy Build. 49, 560–567. https://doi.org/10.1016/j.enbuild.2012.03.003.

Tugal, I., Sevgin, F., 2023. Analysis and forecasting of temperature using time series forecasting methods a case study of Mus. Therm. Sci. 27 (4 Part B), 3081–3088. https://doi.org/10.2298/TSCI2304081T.

Tuğal, İ., 2024. Comparative analysis of LSTM architectures for wind speed forecasting: a case study in Muş, Turkey. Turk. J. Nat. Sci. 13 (4), 107–119. https://doi.org/10.46810/tdfd.1525648.

Tuğal, İ., Karcı, A., 2019. Comparisons of Karcı and Shannon entropies and their effects on centrality of social networks. Phys. A: Stat. Mech. Appl. 523, 352–363. https://doi.org/10.1016/j.physa.2019.02.026.

Vakiloroaya, V., Samali, B., Fakhar, A., Pishghadam, K., 2014. A review of different strategies for HVAC energy saving. Energy Convers. Manag. 77, 738–754. https://doi.org/10.1016/j.enconman.2013.10.023.

Wang, H., Chen, Y., Kang, J., Ding, Z., Zhu, H., 2023. An XGBoost-Based predictive control strategy for HVAC systems in providing day-ahead demand response. Build. Environ. 238, 110350. https://doi.org/10.1016/j.buildenv.2023.110350.

Woo Geem, Zong, Kim, Joong Hoon, Loganathan, G.V., 2001. A new heuristic optimization algorithm: harmony search. Simulation 76 (2), 60–68. https://doi.org/10.1177/003754970107600201.

Xie, X., Pu, Y.-F., Wang, J., 2023. A fractional gradient descent algorithm robust to the initial weights of multilayer perceptron. Neural Netw. 158, 154–170. https://doi.org/10.1016/j.neunet.2022.11.018.

Xie, H., Zhang, L., Lim, C.P., 2020. Evolving CNN-LSTM models for time series prediction using enhanced grey wolf optimizer. IEEE Access 8, 161519–161541. https://doi.org/10.1109/ACCESS.2020.3021527.

Xu, Y., Li, F., Asgari, A., 2022. Prediction and optimization of heating and cooling loads in a residential building based on multi-layer perceptron neural network and different optimization algorithms. Energy 240, 122692. https://doi.org/10.1016/j.energy.2021.122692.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: theory and practice. Neurocomputing 415, 295–316. https://doi.org/10.1016/j.neucom.2020.07.061.

Yang, X.-S., Suash Deb, 2009. Cuckoo search via levy flights. World Congr. Nat. Biol. Inspired Comput. (NaBIC) 2009, 210–214. https://doi.org/10.1109/NABIC.2009.5393690.

Yao, K., Sun, J., Chen, C., Cao, Y., Xu, M., Zhou, X., Tang, N., Tian, Y., 2023. An information entropy-based grey wolf optimizer. Soft Comput. 27 (8), 4669–4684. https://doi.org/10.1007/s00500-022-07593-9.

Zerouali, B., Santos, C.A.G., de Farias, C.A.S., Muniz, R.S., Difi, S., Abda, Z., Chettih, M., Heddam, S., Anwar, S.A., Elbeltagi, A., 2023. Artificial intelligent systems optimized by metaheuristic algorithms and teleconnection indices for rainfall modeling: the case of a humid region in the mediterranean basin. Heliyon 9 (4), e15355. https://doi.org/10.1016/j.heliyon.2023.e15355.

Zhao, J., Yuan, X., Duan, Y., Li, H., Liu, D., 2023. An artificial intelligence (AI)-driven method for forecasting cooling and heating loads in office buildings by integrating building thermal load characteristics. J. Build. Eng. 79, 107855. https://doi.org/10.1016/j.jobe.2023.107855.

Zhou, Y., Liang, Y., Pan, Y., Yuan, X., Xie, Y., Jia, W., 2022. A deep-learning-based meta-modeling workflow for thermal load forecasting in buildings: method and a case study. Buildings 12 (2), 177. https://doi.org/10.3390/buildings12020177.

Zhou, G., Moayedi, H., Bahiraei, M., Lyu, Z., 2020. Employing artificial bee colony and particle swarm techniques for optimizing a neural network in prediction of heating and cooling loads of residential buildings. J. Clean. Prod. 254, 120082. https://doi.org/10.1016/j.jclepro.2020.120082.