 https://doi.org/10.24191/jeesr.v28i1.007

Particle-Swarm-Optimized Neural Networks for Baseline Energy Modelling in Institutional Buildings

Siti Solehah Md Ramli*, Mohammad Nizam Ibrahim, Anuar Mohamad @ Ahmad, Kamarulazhar Daud, Abdul Malek Saidina Omar

Abstract— **Accurate baseline energy models are essential for quantifying savings and guiding efficiency measures in institutional buildings under IPMVP-aligned M&V. This study evaluates multiple linear regression (MLR), a gradient-trained single-hidden-layer artificial neural network (ANN, tanh), and a particle-swarm-optimized ANN (PSO-ANN) for hourly electricity prediction in a Malaysian academic building using 1,584 hourly observations of energy use, outdoor temperature, and occupancy. Data are randomly partitioned 70/15/15 (train/validation/test). Each model is trained with 15 random seeds; for ANN and PSO-ANN, the hidden-layer size is swept from 5 to 30 neurons for each seed, retaining the configuration with the lowest validation-set MSE. A swarm size of 200 and a maximum of 1000 iterations are used to maintain search diversity and support convergence when optimizing the full weight–bias vector. On the test set, MLR, ANN, and PSO-ANN achieve RMSE of ≈23.5 kWh, 15.4 kWh, and 14.5 kWh, respectively (MSE ≈ 553.8 kWh², 238.6 kWh², and 210.0 kWh²). Relative to MLR, ANN reduces test MSE by 57 % and PSO-ANN by 62 %; the additional improvement over ANN is modest. Across seeds, PSO-ANN shows slightly lower central error; however, the improvement over ANN is not statistically significant ($p > 0.05$). Overall, PSO-ANN provides an accurate and reproducible baseline for institutional buildings, enabling benchmarking, retrofit planning, and IPMVP-oriented M&V within smart-campus operations.**

Index Terms— **Artificial Neural Network, Baseline Energy Modelling, Building Energy Forecasting, Energy Consumption, Multiple Linear Regression, Particle Swarm Optimization**.

I. INTRODUCTION

The rising energy demands of educational and institutional buildings underscore the need for precise baseline energy models. Such models enable stakeholders to quantify energy savings by comparing predicted and actual use post-intervention, in line with International Performance Measurement and Verification Protocol (IPMVP) guidance [1]. Accurate forecasts inform the deployment of energy conservation measures (ECMs) and support long-term sustainability planning. Multiple linear regression (MLR) remains widely used due to its simplicity and transparency. Strong predictive performance has been reported in institutional contexts such as campuses, school buildings, libraries, and laboratories with $R^2 \approx 0.80–0.95$ [2]–[5]. However, MLR cannot represent nonlinear interactions among weather, occupancy, and system dynamics, which degrades accuracy under fluctuating operating conditions [6]. Artificial neural networks (ANNs) address this nonlinearity and have delivered high accuracy for building-energy prediction across diverse inputs and horizons [7]–[9]. For educational buildings, studies have reported $R^2 \approx 0.95 –0.99$ with carefully selected inputs [10], [11], illustrating ANNs' strength in modelling both peak loads and highly variable operation. Yet gradient-based training can suffer from local minima and unstable convergence [12].

To mitigate these issues, hybrid metaheuristics such as particle swarm optimization (PSO -ANN) have been explored, where PSO tunes ANN weights and biases. Prior work shows PSO-ANN can match or outperform alternative hybrids such as GA-ANN, ICA-ANN, ABC-ANN while retaining a compact architecture suitable for near-real-time forecasting [13], [14], and achieve high predictive accuracy in building-related applications [15]. Although ensemble and deep-learning approaches such as random forests, deep neural networks, gradient boosting, support vector machines often yield strong accuracy on very large datasets, they typically trade off interpretability and computational efficiency [16], [17]. PSO-ANN offers a balanced alternative, combining nonlinearity with efficient search [18], [19] related applications outside the building domain also indicate favorable convergence behavior at scale [20]. PSO–LSTM variants show further gains in short-term forecasting [21]–[23] but generally require longer histories or high-frequency data.

Despite extensive data-driven work, there remains a lack of rigorous, controlled comparisons of MLR, gradient-trained ANN and PSO-ANN for hourly baseline modelling in institutional buildings, conducted under identical inputs and consistently random partitions, with seed-robust evaluation and

This manuscript is submitted on September 2, 2025, revised on January 14, 2026, accepted on January 15, 2026, and published on April 30, 2026.

Siti Solehah Md Ramli, Mohammad Nizam Ibrahim, Anuar Mohamad @ Ahmad, Kamarulazhar Daud, Abdul Malek Saidina Omar are from the Faculty of Electrical Engineering, Universiti Teknologi MARA Cawangan Pulau Pinang, Permatang Pauh Campus, 13500 Seberang Prai, Pulau Pinang, Malaysia

*Corresponding author
Email address: solehahramli@uitm.edu.my

non-parametric significance testing. This limits a fair assessment of the incremental value of search-based optimization over conventional ANN training for operational baselining.

This paper contributes:

(i) a rigorous baseline comparison of MLR, ANN, and PSO-ANN using 1,584 hourly observations from a Malaysian academic building with identical operational inputs (outdoor temperature, student occupancy and staff occupancy;

(ii) a reproducible evaluation protocol with 15 random seeds, reporting single-run test performance, and distributional robustness (median ± IQR) with Mann–Whitney and Cliff's δ;

(iii) an accuracy-versus-practicality assessment, showing that PSO-ANN delivers lower test error than MLR and gradient-trained ANN, with modest, statistically non-significant gains over ANN ($p > 0.05$), while retaining a compact, interpretable setup suitable for benchmarking and ECM planning.

II. METHODOLOGY

This study focuses on the case building, a five-story teaching block located at a public university in Malaysia, comprising administrative offices, lecturer rooms, and classrooms. As thermal comfort is provided almost exclusively by mechanical air-conditioning, this building serves as a representative case for baseline energy modelling (BEM) in institutional buildings. The analysis concentrates on the primary occupancy period (08:00–17:00) during working days to capture typical academic activities and heating, ventilation, and air conditioning (HVAC) operation patterns.

Hourly data were collected from 2nd October 2023 to 19th February 2024 during office hours, yielding 1,584 valid observations after removing weekends, holidays, and semester breaks. The temperature, occupancy of student, and staff were used as inputs, with energy consumption as the target output. Whole-building active power demand was measured using a Fluke 438 Series II three-phase analyzer, while outdoor dry-bulb temperatures were retrieved from the nearest weather underground station. Hourly student occupancy was reconstructed by cross-referencing room timetables with officially enrolled student headcounts, whereas staff presence was derived from the campus building's anonymized staff check-in records database. Aggregated, anonymized records were used with institutional approval. Each record thus includes four synchronized variables energy consumption, outdoor temperature, student occupancy, and staff occupancy providing a comprehensive, behavior-aware basis for model development.

The BEM framework implemented as shown in Fig. 1 comprises three parallel modelling streams. In the first stream, a multiple linear regression (MLR) model establishes the baseline. In the second, a single-hidden-layer feedforward artificial neural network (ANN) with randomly initialized weights and biases is trained via gradient-based optimization to capture nonlinear input–output relationships. In the third stream, the same ANN architecture is trained with Particle Swarm Optimization (PSO), wherein a swarm explores the weight–bias space under a validation-set objective, updating particle velocities and positions iteratively until convergence, yielding a particle-swarm-optimized artificial neural network (PSO-ANN). Finally, all three models are evaluated side-by-side on the hold-out data using the correlation coefficient (R), coefficient of determination (R^2), root mean square error (RMSE), Coefficient of Variation of the Root Mean Squared Error (CVRMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE).

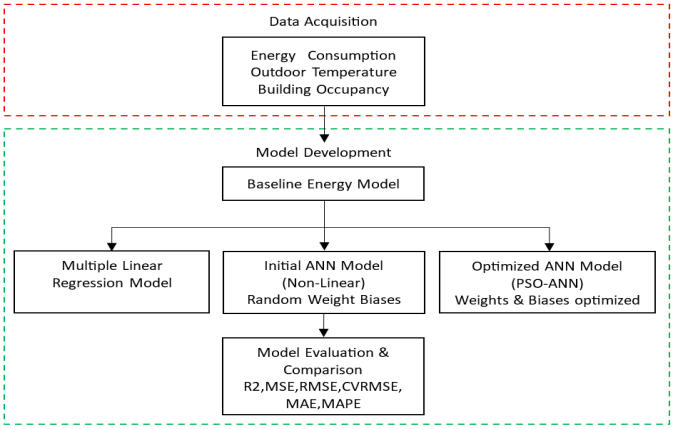

Fig. 1. Research Framework for Baseline Energy Modelling (BEM) Development at the Case Building.

In this study, data handling was carried out systematically to ensure the quality and suitability of data for modelling. Initially, raw hourly measurements including energy consumption, outdoor temperature, and occupancy of student and staff were synchronized and merged into one comprehensive dataset. This dataset was carefully cleaned by addressing missing values through deletion, interpolation, or statistical imputation, correcting anomalies based on defined thresholds and domain expertise, and standardizing all timestamps to a uniform datetime format. All inputs (temperature, occupancy of student and occupancy of staff) were min–max scaled to [−1,1] using statistics computed on the training subset only; the same transform was applied unchanged to validation and test subsets to avoid leakage. Each step was documented thoroughly to ensure reproducibility.

Subsequently, the cleaned dataset was randomly split into 70 % training, 15 % validation, and 15 % testing and prevent leakage. For the MLR baseline, a two-way split (70/30) was applied, corresponding to 1,108 training and 476 testing samples. A two-way split is used for MLR because no hyper-parameter tuning is required; all evaluation for every model is computed only on the held-out test set. For the neural models (ANN and PSO-ANN), a three-way split of 70/15/15 was adopted, yielding 1,108 training, 238 validation, and 238 testing samples. Each configuration (architecture and random initialization) was trained 15 independent times under different random seeds. For single-run comparisons, the run/architecture with the lowest validation-set MSE across the 15 runs was selected. To avoid selection bias and re-use of validation data, all evaluation metrics are computed on the held-out test set, providing an unbiased estimate of generalization.

Finally, each subset underwent further preprocessing, including handling invalid entries, and performing min–max normalization into the [-1, 1] range to facilitate efficient neural-network training and convergence is given as in (1) [24]:

$$x_{norm} = (x - x_{min}) \times \frac{\alpha_{max} - \alpha_{min}}{x_{max} - x_{min}} + \alpha_{min} \quad (1)$$

where x_{min} and x_{max} denote the minimum and maximum of the feature computed on the training set, while α_{min}= -1 and α_{max}=+1 define the target range. This normalization balances the influence of all predictors during gradient updates and accelerates convergence under the Levenberg–Marquardt algorithm.

Model evaluation verifies that baseline models match real-world performance using an independent test dataset. Model validation is crucial to verify that baseline energy models accurately represent the building's real-world performance. It evaluates how well model predictions match actual energy consumption data using an independent test dataset. Six statistical metrics were used such as:

Correlation Coefficient (R): Indicates the strength and direction of linear relationships between predicted and actual values (ideal value close to ±1) as in (2) [24]:

$$R = \frac{n(\sum_{i=1}^{n} x_i\, y_i) - (\sum_{i=1}^{n} x_i)(\sum_{i=1}^{n} y_i)}{\sqrt{[n\sum_{i=1}^{n} {x_i}^2 - (\sum_{i=1}^{n} x_i)^2][n\sum_{i=1}^{n} {y_i}^2 - (\sum_{i=1}^{n} y_i)^2]}} \quad (2)$$

Coefficient of Determination (R^2): Measures how well the independent variables explain variations in the energy consumption (ideal value close to 1) as in (3) [25]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_{actual,i} - y_{predict,i})^2}{\sum_{i=1}^{n}(y_{actual,i} - \bar{y}_{average,i})^2} \quad (3)$$

Mean Squared Error (MSE): Represents the average squared difference between predicted and actual values, showing overall accuracy as in (4) [25]:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_{actual,i} - y_{predict,i})^2 \quad (4)$$

Root Mean Squared Error (RMSE): The square root of MSE, expressing error in the same unit as energy consumption (smaller values indicate better accuracy) as in (5) [25]:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_{actual,i} - y_{predict,i})^2}{n}} \quad (5)$$

Coefficient of Variation of the Root Mean Squared Error (CVRMSE): RMSE normalized by the mean of the observed energy and expressed as a percentage; scale-independent, so smaller values indicate a better fit as in (6) [26] [27]:

$$CVRMSE = 100 \times \frac{\sqrt{\frac{\sum_{i=1}^{n}(y_{actual,i} - y_{predict,i})^2}{n-p}}}{\bar{y}_{actual}} \quad (6)$$

Mean Absolute Error (MAE): Average absolute differences between predicted and actual values, highlighting prediction accuracy without considering direction of errors as in (7) [25]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_{actual,i} - y_{predict,i}| \quad (7)$$

Mean Absolute Percentage Error (MAPE): Measures average prediction error as a percentage, allowing comparison regardless of scale (values close to 0 % indicate high accuracy) as in (8) [25]:

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_{actual,i} - y_{predict,i}}{y_{actual,i}}\right| \times 100 \quad (8)$$

where, $y_{actual,i}$ represent the actual energy consumption at time point i, $y_{predict,i}$ denotes the predicted energy consumption at point i and n is the number of data points in the dataset. While $\bar{y}_{actual} = \frac{1}{n}\sum_{i=1}^{n} y_{actual,i}$ is the mean of the observed values; and p is the number of estimated model parameters ($p = 0$ for an independent test set). Higher R and R^2 values indicate strong predictive relationships while lower values of MSE, RMSE, MAE, and MAPE reflect higher model accuracy [25]. All values reported are from the test subset. All modelling processes were carried out using MATLAB® R2023a.

A. Multiple Linear Regression (MLR) Specification

Regression analysis provides a structured approach for examining the statistical relationship between the dependent variable and the independent variables as influencing factors. It enables the identification and quantification of patterns in historical data, forming the basis for predictive modelling. In this study, Multiple Linear Regression (MLR) is employed to predict building energy consumption by simultaneously analyzing three predictors: outdoor temperature, student occupancy, and staff occupancy. The general MLR equation is expressed as in (9) [27]:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_{2,} + \cdots + \beta_r X_r + \epsilon \quad (9)$$

where y denote the dependent variable (output), $X_1, X_2, \ldots, X_r$ correspond to the independent variables (inputs), β_0 is the intercept, indicating the predicted value of y when all inputs are zero, $\beta_1, \beta_2, \ldots, \beta_r$ are the regression coefficients (slopes), indicating the magnitude and direction of the relationship between each independent variable and the

dependent variable and ϵ captures the remaining unexplained variability.

B. Feedforward Artificial Neural Network (ANN) Architecture and Training

A feedforward artificial neural network (ANN) is employed in this study using a single-hidden-layer architecture to balance model complexity and predictive accuracy. The input layer comprises three preprocessed features: outdoor temperature, student occupancy, and staff occupancy. These inputs are fully connected to a hidden layer with a hyperbolic tangent activation function, and the network output is produced by a linear output neuron to predict hourly energy consumption. The network components and signal flow are summarized in Fig. 2, with the forward-pass computations given in (10) until (12).

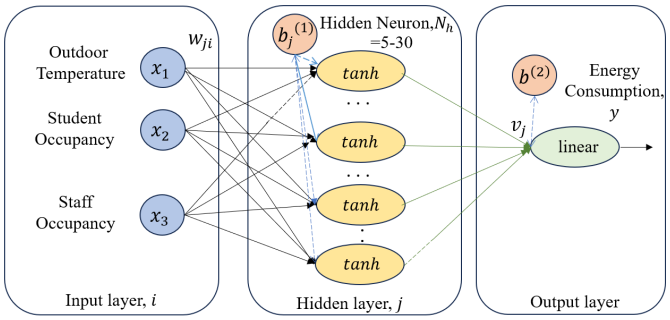

Fig. 2. Single-hidden-layer feedforward ANN used for energy consumption prediction.

For the hidden layer, the net input to hidden neuron, j is computed as in [28]:

$$z_j = \sum_{i=1}^{N_{in}} w_{ji}\, x_i + b_j^{(1)}, j = 1, \dots, N_h \tag{10}$$

and the hidden activation as in [28][29]:

$$h_j = \tanh(z_j) \tag{11}$$

The output layer aggregates the hidden activations using a linear activation function as in [28]:

$$y = \sum_{j=1}^{N_h} v_j\, h_j + b^{(2)} \tag{12}$$

where x_i denotes the i-th input feature (x_1, x_2, x_3); $N_{in} = 3$ is the number of inputs; w_{ij} is the weight connecting input i to hidden neuron j; $b_j^{(1)}$ is the bias of hidden neuron j; v_j is the weight connecting hidden neuron j to the output neuron; $b^{(2)}$ is the output-layer bias; $\tanh(\cdot)$ is the hidden-layer activation function; and $y$is the predicted energy consumption.

The hidden-layer size N_h is varied systematically from 5 to 30 neurons to identify an optimal network structure. The $\tanh(\cdot)$ activation is selected to support stable learning and effective representation of nonlinear relationships [30]. Prior to training, network weights and biases are randomly initialized within a controlled range to avoid early saturation and to keep activations within an effective operating region.

Network training is performed using the Levenberg–Marquardt (LM) algorithm [31]. During training, the ANN iteratively minimizes the mean squared error (MSE) by updating weights and biases based on the prediction error between $y$and the measured energy consumption. The dataset is partitioned into training, validation, and testing subsets: the training subset is used to estimate network parameters, the validation subset is used for model selection (including early stopping and choosing the optimal N_h), and the testing subset is used to evaluate final predictive performance on unseen data. The final ANN configuration is selected based on the lowest validation MSE and the highest coefficient of determination (R^2).

C. Particle Swarm -Optimized ANN (PSO-ANN) Procedure

Particle Swarm Optimization (PSO) is a metaheuristic optimization algorithm inspired by the collective behavior of bird flocks or fish schools. It operates using a swarm of particles, where each particle represents a potential solution. Particles dynamically update their positions based on individual experience (personal best) and collective knowledge of the swarm (global best). PSO was selected over more recent metaheuristics because this study is not to propose a new optimizer, but to establish a reproducible baseline-modeling benchmark using an optimizer that is (i) well-established and widely adopted, (ii) simple and easy to implement, and (iii) effective for continuous, high-dimensional search such as ANN weight–bias vectors [32],[33]. Meanwhile, newer algorithms such as Grey Wolf Optimizer (GWO), Whale Optimization Algorithm (WOA), and Salp Swarm Algorithm (SSA) can be competitive, they often introduce additional design choices and tuning complexity; moreover, PSO performance is known to be sensitive to core settings such as swarm size, which can materially affect outcomes and attribution of gains [33], [34]. The algorithm iteratively searches for an optimal solution by adjusting particles' velocities and positions using in (13) and (14) [35]:

Velocity update:

$$v_i(t+1) = \omega v_i(t) + c_1 r_1 \left(pbest_i - x_i(t)\right) + c_2 r_2 \left(gbest - x_i(t)\right) \tag{13}$$

Position update:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{14}$$

where $v_i(t+1)$ denotes the updated velocity of particle i at iteration $t+1$; $x_i(t+1)$ denotes the updated position of particle i at iteration $t+1$; ω is the inertia weight; c_1 and c_2 are the acceleration coefficients (cognitive and social influences); r_1 and r_2 are independent random numbers in r_1, $r_2 \sim U(0, 1)$; $pbest_i$ is the personal best position of particle i; and $gbest$ is the global best position among all particles.

All trainable parameters (weights and biases) are concatenated into a particle, and the validation-set MSE is used as fitness. Particle positions are updated by the inertial–cognitive–social rule until convergence; the inertia weight, ω is decreased from ω_{max} to ω_{min} to balance exploration (global search) and exploitation (local refinement). The global-best particle defines (W, b), which are then evaluated on the held-out test set, providing an unbiased and directly comparable estimate of generalization across methods. PSO is adopted for its rapid convergence in non-linear search spaces and minimal hyperparameter tuning. The iteration stops when improvements in the ANN's mean squared error (MSE) fall below a threshold or when a preset maximum number of iterations is reached.

In the PSO-ANN, all ANN weights and biases formed the particle position. A swarm size of 200 and a maximum of 1,000 iterations were used with a linear inertia schedule ($\omega_{max} = 0.9$ and $\omega_{min} = 0.4$) [36]. Multiple (c_1, c_2) combinations were evaluated within the evaluated range ($\approx$ 1.0–4.0) [37]. Since PSO performance is particularly sensitive to the inertia weight and the acceleration coefficients (c_1, c_2), and unbalanced settings such as $c_1 \gg c_2$ or $c_1 \ll c_2$ may induce excessive wandering or premature convergence, a discrete sweep of (c_1, c_2) was conducted [38]. The acceleration coefficients were evaluated using 20 (c_1, c_2) combinations: (1.2,1.8), (1.4,1.8), (1.0,4.0), (3.0,1.0), (2.0,3.0), (1.8,1.2), (1.1,1.3), (1.3,1.7), (1.0,1.0), (1.5,2.5), (1.4,1.4), (1.5,1.5), (1.6,1.8), (2.0,2.0), (1.5,2.0), (3.0,3.0), (1.9,1.9), (1.7,1.7), (1.3,1.3), and (1.1,1.1). The swarm size and iteration budget are problem-dependent design choices in PSO and were selected to preserve search diversity and support convergence in the high-dimensional ANN parameter space [33]. In particular, the swarm size (population size) is known to influence PSO behavior and performance, and thus was set accordingly [34]. Positions were clipped to $[-1,1]$ (weight bounds). The fitness function was the validation MSE, and convergence was tracked on the validation subset; the test set was used only for final reporting.

III. Results and Discussion

This section presents three analyses. Subsection A establishes the MLR baseline and summarizes its test-set fit and error characteristics. Subsection B evaluates a single-hidden-layer ANN using correlation/error metrics together with time-series overlays and absolute-error traces. Subsection C assesses the PSO-ANN with the same diagnostics to enable direct comparison.

A. Multiple Linear Regression (MLR)

Multiple Linear Regression (MLR) integrates three independent variables to predict building energy consumption (kWh). The regression shown in (15):

$$y = -6.8853 + 1.1221x_1 + 0.2135x_2 + 1.0482x_3 \quad (15)$$

where x_1 represents the outdoor temperature, x_2 represents the student occupancy, x_3 represents the staff occupancy and y represents the predicted energy consumption.

The coefficients in (15) were obtained by fitting a multiple linear regression model using ordinary least squares (OLS) with an intercept term. The coefficients were computed directly by minimizing the sum of squared residuals (no trial-and-error tuning). The predictors were used in their original units (no normalization/standardization), and rows with missing values in any of (x_1, x_2, x_3, y) were removed prior to model fitting. The regression was implemented in MATLAB using fitlm, and $R$and R^2 were obtained from the fitted model output.

Equation (15) indicates that energy consumption increases by approximately 1.12 kWh per 1 °C rise in outdoor temperature, 0.21 kWh per additional student, and 1.05 kWh per additional staff member. Notably, staff occupancy has the strongest impact, reflecting the substantial energy use linked to longer working hours, office equipment, and specific HVAC demands. The statistical performance of the MLR baseline model on the test set is summarized in Table I. As shown in Table X, the MLR baseline demonstrates strong predictive performance (R = 0.9674, R^2= 0.9410), indicating that 94.10 % of the variability in energy consumption is explained by the predictors. The corresponding error levels are MSE = 553.469 kWh², RMSE = 23.526 kWh, and CVRMSE = 24.91 %, indicating that the typical prediction error is about 23.5 kWh, or ~25 % of the mean measured load.

TABLE I. Statistical performance metrics of the MLR baseline model.

Metric	Value
R	0.9674
R^2	0.9410
MSE (kWh²)	553.469
RMSE (kWh)	23.526
CVRMSE (%)	24.91
MAE (kWh)	14.702
MAPE (%)	26.96

Fig. 3 illustrates how the MLR model predicted energy consumption aligns with the measured values across 476 test samples. Overall, the measured values and predicted values exhibit strong alignment, with both trends closely tracking each other across most of the samples. This visual agreement supports the regression model's predictive capability in capturing the temporal patterns of building energy consumption. Minor discrepancies are observed at certain peak periods, where predicted values slightly overestimate or underestimate actual consumption. These deviations are likely attributable to short-term operational fluctuations or unaccounted external factors such as weather anomalies and sporadic equipment usage.

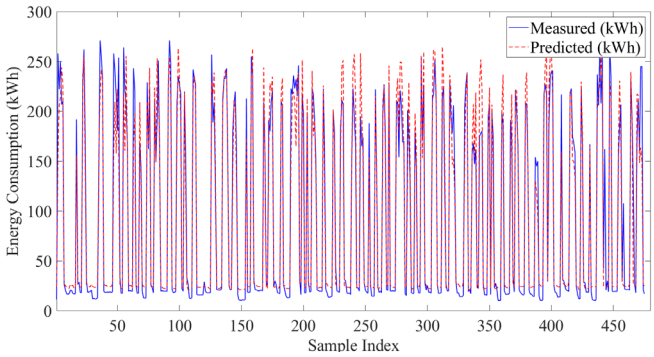

Fig. 3. Comparison of measured and predicted energy consumption of the MLR baseline energy model.

Fig. 4 plots the absolute error between measured and predicted energy consumption and expands on the earlier comparison. The error profile highlights fluctuations corresponding to prediction inaccuracies and shows notable spikes during periods of elevated usage. Overall, absolute errors remain low across most of the testing dataset with an MAE of approximately 14.70 kWh, reaffirming the model's reliability. However, the presence of intermittent error peaks suggests that during instances of high consumption variability, the model's prediction accuracy slightly diminishes. This performance indicates potential areas for improvement, such as employing adaptive learning models to dynamically adjust to real-time operational changes, thereby further enhancing forecasting precision.

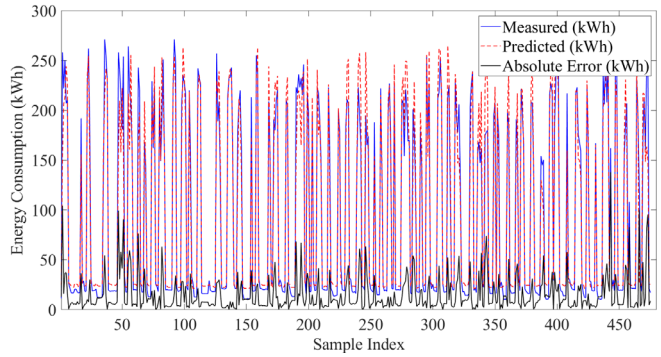

Fig. 4. Comparison of measured and predicted energy consumption with absolute error of the MLR baseline energy model

B. Artificial neural network (ANN)

Alongside the multiple linear regression models, feed-forward ANNs with hidden layers ranging from 5 to 30 neurons were developed. The performance evaluation of correlation metrics for each hidden-layer size is shown in Table II, while Fig. 5 plots the Rtrain, Rvalid, Rtest and Rall against neuron count, highlighting trends in generalization and overfitting, which indicate how effectively the ANN models account for variance in energy consumption. The optimal configuration uses 8 neurons in the hidden layer, achieving the highest $R = 0.9863$ and $R^2 = 0.9728$, which corresponds to explaining approximately 97.28 % of the observed variability. Models with 5 and 10 neurons also perform strongly ($R \approx 0.985$, $R^2 \approx 0.970$), but adding more neurons beyond 8 yields diminishing returns. At 20 neurons, for example, R falls to 0.9736 and R^2 to 0.9479, suggesting that greater complexity without adequate data justification leads to overfitting and reduced explanatory power. Consequently, the 8-neuron architecture strikes the best balance between accuracy and generalization, making it the most robust choice for predicting energy consumption.

Additionally, correlation analysis across the training, validation and testing datasets highlights how hidden-layer size affects model generalization also shown in Table II. From this table, the networks with very small hidden layers capture broad trends but fail to model subtler relationships. The 5-neuron model, for instance, records a training correlation of 0.9806 and a validation correlation of 0.9714, indicating that it accounts for most but not all of the observed variance. Its relatively lower validation value points to underfitting and insufficient representational capacity. When increased to 8 neurons, the network achieves the optimum trade-off between fitting the training data and generalizing to new cases.

Training correlation rises to 0.9811, validation correlation remains strong at 0.9699, testing correlation peaks at 0.9863 and overall correlation is 0.9802. These results correspond to explaining approximately 98 % of the variance across both seen and unseen samples without overemphasizing the training set. As hidden-layer size increases beyond 12 neurons, overfitting becomes apparent. By 16 neurons training correlation climbs further to 0.9824 while validation correlation falls to 0.9754. At 22 neurons validation correlation collapses to 0.4297 despite training remaining above 0.985. This divergence illustrates how excessive complexity allows memorization of training cases at the expense of true predictive power.

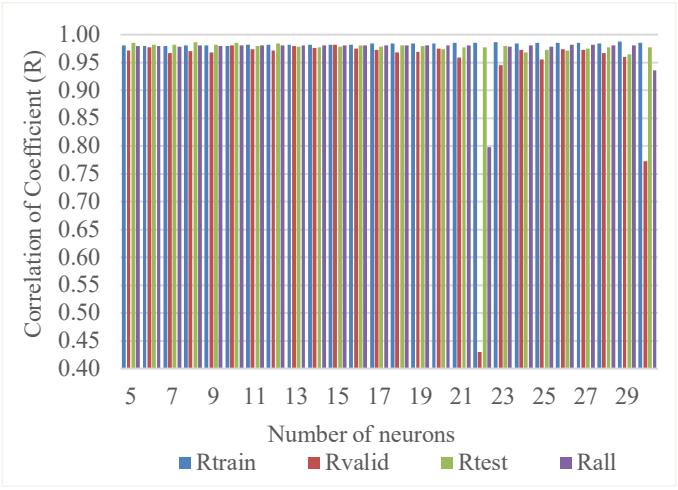

Fig. 5. Correlation coefficients vs. number of neurons in the hidden layer for the ANN baseline energy model.

TABLE II. PERFORMANCE EVALUATION OF CORRELATION METRICS FOR THE ANN BASELINE ENERGY MODEL.

Number of neurons	R	R^2	Rtrain	Rvalid	Rtest	Rall
5	0.9849	0.9700	0.9806	0.9714	0.9849	0.9799
6	0.9813	0.9629	0.9795	0.9773	0.9813	0.9794
7	0.9816	0.9636	0.9801	0.9674	0.9816	0.9783
8	0.9863	0.9728	0.9811	0.9700	0.9863	0.9802
9	0.9813	0.9629	0.9812	0.9678	0.9813	0.9793

10	0.9855	0.9712	0.9791	0.9804	0.9855	0.9802
11	0.9797	0.9598	0.9820	0.9737	0.9797	0.9804
12	0.9836	0.9675	0.9819	0.9719	0.9836	0.9807
13	0.9790	0.9583	0.9813	0.9799	0.9790	0.9807
14	0.9769	0.9544	0.9823	0.9757	0.9769	0.9804
15	0.9782	0.9569	0.9813	0.9814	0.9782	0.9809
16	0.9804	0.9612	0.9824	0.9754	0.9804	0.9811
17	0.9779	0.9563	0.9836	0.9726	0.9779	0.9810
18	0.9804	0.9611	0.9837	0.9679	0.9804	0.9806
19	0.9793	0.9591	0.9841	0.9691	0.9794	0.9811
20	0.9736	0.9479	0.9839	0.9744	0.9736	0.9810
21	0.9774	0.9553	0.9857	0.9588	0.9774	0.9805
22	0.9770	0.9545	0.9857	0.4297	0.9770	0.7980
23	0.9792	0.9588	0.9863	0.9451	0.9792	0.9789
24	0.9684	0.9378	0.9847	0.9730	0.9684	0.9805
25	0.9722	0.9452	0.9851	0.9551	0.9722	0.9780
26	0.9715	0.9437	0.9855	0.9737	0.9715	0.9815
27	0.9749	0.9504	0.9856	0.9730	0.9749	0.9818
28	0.9772	0.9550	0.9847	0.9668	0.9772	0.9810
29	0.9646	0.9305	0.9872	0.9597	0.9646	0.9802
30	0.9767	0.9540	0.9850	0.7726	0.9767	0.9356

In addition, error-based measures provide a vital complement to the correlation analysis by highlighting how predictive precision evolves with hidden-layer size. Table III presents the performance evaluation of ANN baseline energy model with varying hidden layer neurons. The 8-neuron network attains the lowest MSE of 238.59 kWh^2 and RMSE of 15.45 kWh, while also achieving a MAE of 10.09 kWh and a MAPE of 15.92 %. These results demonstrate that this configuration not only minimizes average squared deviations but also maintains uniformly low absolute and relative errors across the full dataset. In terms of calibration, the 8-neuron model also attains a CVRMSE of 17.45 %, comparable to the lowest values observed across configurations (≈16.5 % –17.7 %) and well below commonly used hourly M&V thresholds (≈ 30 %). By contrast, the 5-neuron model achieves the smallest MAE of 9.72 kWh, illustrating its capacity to produce tightly centered predictions, yet its MSE and MAPE increase to 247.83 kWh^2 and 17.87 % respectively. This inconsistency indicates that although a very small network may excel at limiting magnitude errors for many observations, it lacks the representational power to deliver stable accuracy throughout. Networks with 16 hidden neurons yield the lowest percentage error at 15.27 % but incur greater squared-error penalties, with MSE of 316.99 kWh^2 and RMSE of 17.80 kWh, reflecting wider variability in prediction errors. Mid-range designs, with 10 to 12 neurons, strike a balance by producing moderate values across all metrics without achieving peak performance in any single category.

At the upper end of complexity, especially with 20 neurons, overfitting becomes pronounced as MSE and RMSE reach 433.02 kWh^2 and 20.81 kWh respectively, revealing that excessive model capacity leads to memorization of the training set at the expense of genuine generalization. Overall, these findings confirm that an 8-neuron hidden layer offers the optimal compromise between complexity and robustness, making it the most effective choice for reliable and accurate energy-consumption forecasting. Across the tested architectures, CVRMSE spans approximately 16.5 % –24.7 %, with the 8-neuron design remaining among the most efficient in balancing accuracy and robustness.

TABLE III. PERFORMANCE EVALUATION OF ANN BASELINE ENERGY MODEL.

Number of neurons	MSE (kWh^2)	RMSE (kWh)	CVRMSE (%)	MAE (kWh)	MAPE (%)
5	247.834	15.743	16.393	9.719	17.869
6	303.455	17.420	17.697	10.552	19.005
7	320.512	17.903	18.509	11.111	16.507
8	238.593	15.446	17.453	10.093	15.918
9	329.832	18.161	20.603	11.306	16.535
10	243.022	15.589	16.516	9.784	18.250
11	341.795	18.488	17.601	11.522	18.964
12	285.383	16.893	17.836	10.267	17.816
13	377.611	19.432	21.149	11.957	17.772
14	376.338	19.399	19.627	11.193	16.792
15	359.295	18.955	20.534	11.139	18.017
16	316.987	17.804	17.286	10.832	15.266
17	405.778	20.144	21.853	13.262	19.480
18	345.264	18.581	19.544	11.485	18.953
19	371.065	19.263	21.296	11.821	15.691
20	433.020	20.809	23.107	12.477	17.408
21	397.132	19.928	20.574	12.578	17.244
22	388.749	19.717	21.019	12.390	17.854
23	359.892	18.971	19.269	11.669	19.428
24	541.623	23.273	24.743	13.208	18.778
25	473.166	21.752	21.577	11.821	16.789
26	486.761	22.062	23.220	12.806	19.025
27	461.169	21.474	22.609	13.405	18.201
28	372.118	19.290	18.807	11.627	18.516
29	547.122	23.390	22.739	13.099	20.626
30	408.176	20.203	20.817	12.440	17.850

Following the demonstrated optimal performance of the 8-neuron model in correlation and error measures, its hourly predictions are compared against actual energy consumption. Figure 6 plots the network's predictions (red line) alongside measured baseline use (blue line) for all 1,584 observations. The prediction profile closely mirrors observed fluctuations, capturing both pronounced peaks and subtle troughs in energy use without noticeable lag or systematic bias. High-demand events align accurately in amplitude and timing, while low-usage intervals are equally well represented. The residuals appear randomly distributed around zero, confirming that the model reproduces the underlying temporal dynamics rather than merely matching aggregate error statistics.

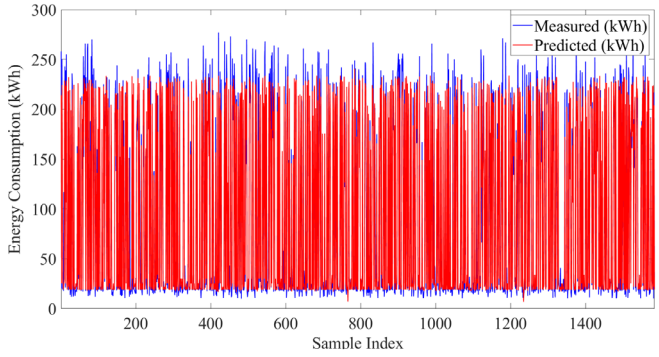

Fig. 6. Energy consumption prediction of the ANN baseline energy model with 8 neurons.

Extending the time-series analysis to unseen data, Fig. 7 presents the model's performance on the 238-sample test set, which was kept entirely separate from training and validation. The red dotted prediction curve closely follows the blue

measured energy consumption curve, accurately capturing both sharp demand spikes and extended low-usage periods. Besides, the prediction errors remain small, randomly distributed, and free of systematic bias. Such performance on completely unseen data confirms the 8-neuron configuration's robustness and its suitability for real-world energy-use forecasting.

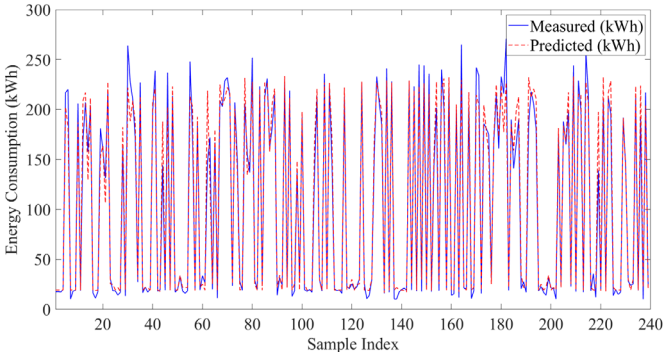

Fig. 7. Energy consumption prediction of the ANN baseline energy model with 8 neurons (testing data).

Fig. 8 further enriches the comparison by plotting absolute error (black line) alongside measured energy consumption (blue line) and model predictions (red line) for the same 238 test samples. Error magnitudes remain uniformly low which is typically below 10 kWh and seldom exceeding 20 kWh even during the most extreme demand spikes. This consistently small, unbiased error distribution confirms that the 8-neuron network delivers precise and reliable forecasts under all operating conditions.

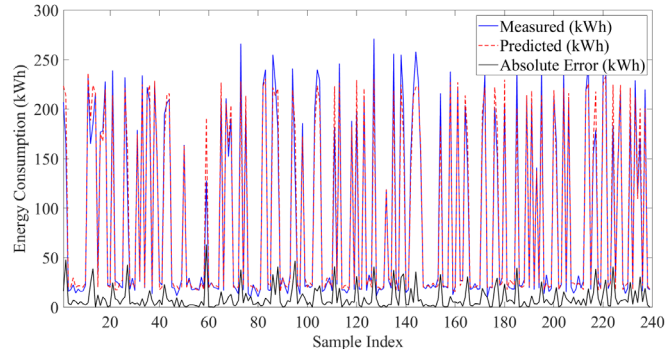

Fig. 8. Energy consumption prediction vs. absolute error of the ANN baseline energy model with 8 neurons (testing data).

C. Particle swarm -optimized ANN (PSO-ANN)

This study further enhanced ANN performance by employing Particle Swarm Optimization (PSO) to fine-tune network weights and biases, systematically testing hidden-layer sizes from 5 to 30 neurons. Performance was evaluated using correlation metrics and prediction errors to determine the optimal configuration. The correlation analysis of the PSO-ANN baseline energy model is summarized in Table III, while Fig. 9 plots the correlation coefficients (Rtrain, Rvalid, Rtest, Rall) across different hidden-layer sizes. By systematically varying the number of hidden neurons, this analysis seeks to identify the configuration that delivers the greatest predictive accuracy without introducing unnecessary model complexity. Across all 26 network configurations tested, the PSO-ANN exhibits consistently high performance, with R values exceeding 0.985 and corresponding R^2 values above 0.970. Even the most compact network consisting of only five hidden neurons achieves R = 0.9861 and R^2 = 0.9719, demonstrating that the PSO algorithm effectively tunes connection weights and biases to capture the dominant patterns in the building energy data. Such uniformly strong outcomes confirm both the robustness of the PSO training process and the relevance of the selected input features for modeling energy consumption.

As the number of hidden neurons increases, the model's complexity grows and its capacity to approximate nonlinear relationships improves, peaking at 30 neurons where the highest R^2 of 0.9747 (with R = 0.9873) is observed. A similarly strong result is obtained with 23 neurons (R = 0.9872, R^2 = 0.9746). Beyond roughly 23 hidden units, however, further increases yield only marginal improvements in goodness of fit, indicating diminishing returns and suggesting that excessively large networks add little predictive benefit while increasing the risk of over-parameterization.

Despite the overall trend of rising accuracy with network size, slight dips in performance at 7, 10, and 16 neurons highlight the stochastic nature of PSO and the sensitivity to initial conditions. This variability underlines the importance of conducting multiple training runs or employing ensemble techniques when selecting a final architecture. Balancing accuracy against parsimony, a hidden-layer size in the range of 20–25 neurons is therefore recommended, as it achieves near-peak performance explaining over 97 % of the variance with moderate complexity, thereby minimizing training time and the potential for overfitting.

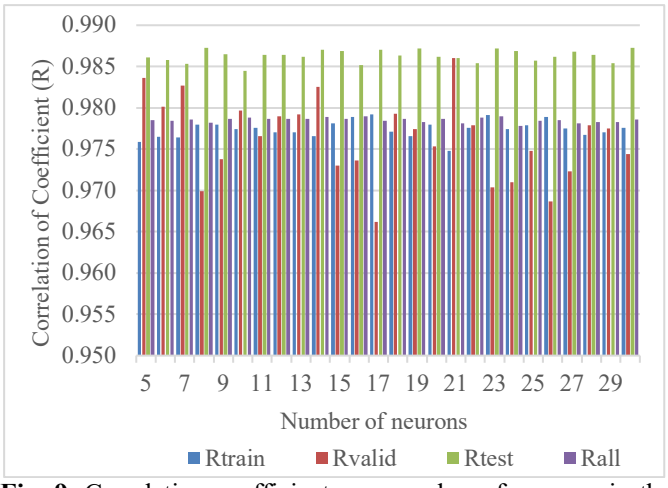

Fig. 9. Correlation coefficients vs. number of neurons in the hidden layer for the PSO-ANN baseline energy model.

Additionally, Table IV also presents the PSO-ANN's predictive strength separately on training (Rtrain), validation (Rvalid), and test (Rtest) subsets, as well as the aggregated correlation (Rall) across all 1,584 hourly observations. By partitioning the data into a 70 %,15 %,15 % split, these metrics reveal both how well the network fits the data it sees during weight optimization and how robustly it generalizes to unseen

cases. Consistently high values for Rtrain around 0.978 and Rvalid and Rtest hovering near 0.98. This indicate that the PSO process effectively tunes the ANN to capture the underlying energy-use patterns without severe overfitting.

On the training subset, Rtrain remains remarkably stable across hidden-layer sizes, fluctuating only between 0.9759 and 0.9789. This narrow band demonstrates that even the smallest networks learn the dominant trends in the data with minimal variance, while larger networks do not dramatically improve fit on data already seen. The slight peak in training correlation at 16 neurons (Rtrain = 0.9789) suggests an optimal capacity for complex pattern capture, but gains beyond this point are virtually negligible. Validation performance likewise exhibits minimal spread, ranging from Rvalid = 0.9662 at 17 neurons up to 0.9836 at 5 neurons. Although the highest validation correlation occurs with only five hidden units, this result likely reflects stochastic fluctuations in PSO rather than a true advantage of extremely small networks. In general, validation and test correlations track one another closely, confirming that the model's behavior on the hold-out sets is consistent and that the PSO-ANN does not suffer from over-tuning on the training data.

Across all observations, the overall correlation Rall remains above 0.978 for every architecture, peaking at 0.9790 with 23 hidden neurons. This near-uniform excellence underscores the PSO-ANN's reliability regardless of specific split or network size, the model explains over 95.5 % of the variance in energy consumption. Given the diminishing returns in Rall beyond approximately 20 neurons and the occasional dip in validation at mid-range sizes, a hidden-layer size between 20 and 25 neurons is recommended to balance slightly higher aggregate accuracy with manageable model complexity and training cost.

TABLE IV. PERFORMANCE EVALUATION OF CORRELATION METRICS FOR THE PSO-ANN BASELINE ENERGY MODEL.

Number of neurons	R	R^2	Rtrain	Rvalid	Rtest	Rall
5	0.9861	0.9719	0.9759	0.9836	0.9861	0.9785
6	0.9858	0.9716	0.9765	0.9801	0.9858	0.9784
7	0.9853	0.9707	0.9764	0.9827	0.9853	0.9786
8	0.9873	0.9741	0.9780	0.9699	0.9873	0.9782
9	0.9865	0.9723	0.9780	0.9738	0.9865	0.9787
10	0.9845	0.9692	0.9774	0.9797	0.9845	0.9788
11	0.9864	0.9727	0.9776	0.9766	0.9864	0.9787
12	0.9864	0.9726	0.9770	0.979	0.9864	0.9787
13	0.9862	0.9724	0.9770	0.9792	0.9862	0.9787
14	0.9870	0.9723	0.9766	0.9825	0.987	0.9789
15	0.9869	0.9731	0.9781	0.973	0.9869	0.9787
16	0.9852	0.9704	0.9789	0.9736	0.9852	0.979
17	0.9870	0.9739	0.9792	0.9662	0.9870	0.9784
18	0.9863	0.9724	0.9771	0.9793	0.9863	0.9787
19	0.9872	0.9745	0.9766	0.9774	0.9872	0.9783
20	0.9862	0.9711	0.978	0.9753	0.9862	0.9787
21	0.9860	0.9723	0.9748	0.986	0.9860	0.9781
22	0.9854	0.9710	0.9776	0.9779	0.9854	0.9788
23	0.9872	0.9746	0.9791	0.9704	0.9872	0.979
24	0.9869	0.9739	0.9774	0.971	0.9869	0.9778
25	0.9857	0.9717	0.9779	0.9748	0.9857	0.9784
26	0.9862	0.9727	0.9789	0.9687	0.9862	0.9785
27	0.9868	0.9737	0.9775	0.9723	0.9868	0.9781
28	0.9864	0.9729	0.9767	0.9779	0.9864	0.9783
29	0.9854	0.9710	0.9770	0.9775	0.9854	0.9783
30	0.9873	0.9747	0.9776	0.9744	0.9873	0.9786

Next, the PSO-ANN's accuracy is assessed using four complementary error metrics, as presented in Table V. Considering all four metrics in parallel thus offers a nuanced view of how prediction errors respond to changes in hidden-layer size. Across the 26 network configurations, MSE ranges from approximately 210 kWh^2 to 255 kWh^2, with the lowest value occurring at 23 hidden neurons and a secondary minimum at 17 neurons. RMSE follows the same pattern, varying between about 14.49 kWh and 15.98 kWh. These trends indicate that mid-range architectures deliver the most precise squared-error predictions, whereas very small or very large networks produce slightly greater dispersion in their squared deviations. From a calibration standpoint, CVRMSE ranges from 14.16 % to 18.00 % on the held-out test set across the 26 configurations; notably, the 23-neuron model records 14.72 %, well below commonly used hourly M&V thresholds.

Turning to absolute errors, MAE remains below 10.5 kWh for every configuration and reaches its minimum of 9.23 kWh at 23 neurons. In relative terms, MAPE values lie between 15.8 % and 22.1 %, with the most favorable percentage errors observed at 6 and 30 neurons. The tight clustering of both absolute and relative errors highlights the PSO-ANN's ability to sustain stable predictive performance across a broad spectrum of hidden-layer sizes.

Taken together, these findings demonstrate that configuring the PSO-ANN with approximately 20 to 25 hidden neurons strikes the optimal balance between error minimization and architectural simplicity. In particular, the 23-neuron model simultaneously achieves the lowest MSE, RMSE and MAE, making it the most effective choice for accurate energy-use forecasting while keeping computational demands moderate.

After confirming that a 23-neuron hidden layer delivers the best error metrics, Fig. 10 overlays PSO-ANN's hourly predictions (red) with the measured energy-use data (blue). The time-series comparison provides a visual complement to the error metrics values, showing how closely the network mimics real consumption patterns. Throughout the full range of operation, the red and blue lines remain tightly paired, capturing low-load periods (around 20–40 kWh) and daily peaks (around 250–280 kWh) with minimal deviation. Slight underestimation of the very highest spikes appears, but these differences are small relative to the total variation. Such strong correspondence confirms that the PSO-tuned 23-neuron model generalizes effectively and produces reliable hour-by-hour forecasts.

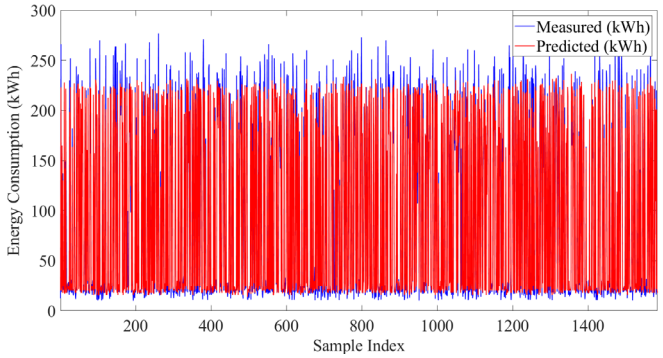

Fig. 10. Overall energy consumption prediction of the PSO-ANN baseline energy model with 23 neurons.

TABLE V. PERFORMANCE EVALUATION OF PSO-ANN BASELINE ENERGY MODEL.

Number of neurons	MSE (kWh^2)	RMSE (kWh)	CVRMSE (%)	MAE (kWh)	MAPE (%)
5	232.566	15.250	16.291	10.422	22.129
6	242.492	15.572	14.499	10.212	15.790
7	234.262	15.305	15.992	9.396	17.294
8	234.605	15.316	16.670	9.949	16.262
9	241.859	15.551	16.425	9.987	17.842
10	246.397	15.697	15.848	9.833	19.688
11	231.607	15.218	16.239	9.579	16.904
12	234.873	15.325	14.316	10.227	18.598
13	234.695	15.319	14.985	9.956	20.009
14	221.899	14.896	15.385	9.533	19.371
15	239.477	15.475	15.627	10.135	16.071
16	250.342	15.822	16.792	9.818	16.841
17	214.445	14.643	14.154	9.385	17.042
18	230.600	15.185	14.947	9.839	19.056
19	219.885	14.828	16.768	9.108	17.827
20	255.372	15.980	16.959	10.513	17.169
21	243.983	15.620	15.478	9.930	17.823
22	242.244	15.564	15.934	9.994	15.813
23	210.026	14.492	14.720	9.225	16.632
24	239.765	15.484	17.196	9.764	15.960
25	238.834	15.454	16.325	9.723	17.890
26	237.431	15.408	16.739	9.942	17.345
27	221.995	14.899	14.186	9.646	18.343
28	220.263	14.841	18.001	9.429	18.624
29	236.472	15.377	14.858	9.590	18.782
30	236.764	15.387	14.765	9.654	15.789

Following the full-dataset comparison, Fig. 11 focuses specifically on the 1,108 samples used for training, overlaying PSO-ANN's predictions on the measured values. This plot provides a more detailed view of how well the network has fit the data it was optimized on, complementing the quantitative Rtrain ≈ 0.978 and low training-set error metrics reported earlier. Visually, the red and blue lines track one another closely throughout the training window, capturing the recurring daily cycles from low-load troughs near 20 kWh to peak demands around 270 kWh. Occasional mismatches occur at the highest peaks, where the model slightly underpredicts extreme spikes, but these deviations are minimal relative to the overall amplitude. The tight alignment confirms that PSO effectively tuned the 23-neuron network to learn the underlying energy-use patterns in the training set without significant overfitting.

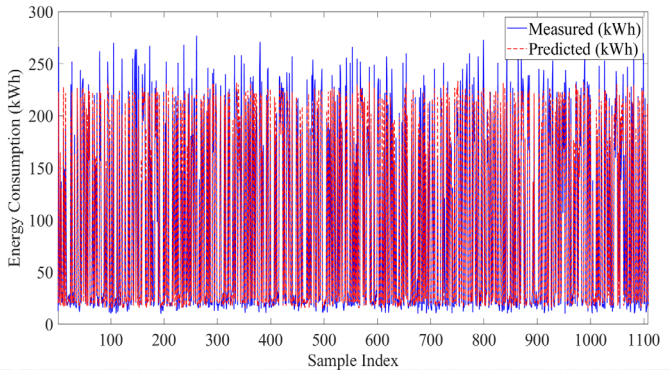

Fig. 11. Comparison of measured and predicted hourly energy (PSO-ANN, 23-neuron; training set).

Following the close fit on the training set, Fig. 12 presents the PSO-ANN's predictions alongside the measured energy-use data for the 238-sample validation set. The model reliably tracks both the low-load intervals around 0–20 kWh and the daytime peaks near 200–250 kWh, with only minor timing offsets early in the series and slight amplitude errors around sample indices 60 and 180. These small discrepancies remain negligible compared to the overall variation, demonstrating that the 23-neuron network generalizes effectively beyond its training data.

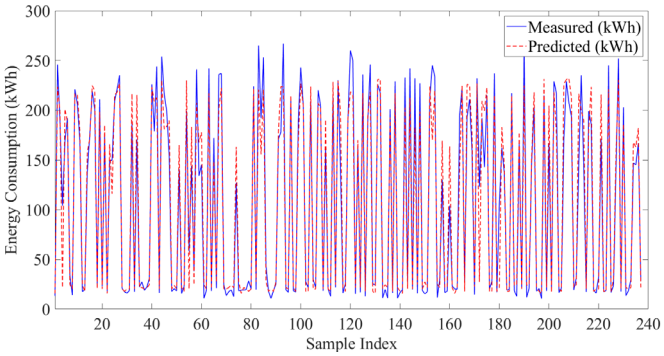

Fig. 12. Comparison of measured and predicted hourly energy (PSO-ANN, 23-neuron; validation set).

Fig. 13 presents the PSO-ANN's predictions (red dashed line) alongside the measured energy-use data (solid blue line) for the 238-sample testing set. This graph completes the series by demonstrating how well the 23-neuron network generalizes to entirely unseen data, reinforcing the robustness indicated by Rtest ≈ 0.986 and the low error metrics reported earlier. Throughout the testing period, the predicted and measured curves remain tightly paired, accurately reflecting both off-peak intervals near 0–20 kWh and daytime peaks around 200–260 kWh. A few slight mismatches occur, such as a modest underprediction of the spike at sample index 30 and minor overshoots near index 120, but these differences are negligible relative to the full range of variation. The close correspondence between red and blue traces confirms that the PSO-tuned 23-neuron model maintains its predictive fidelity on out-of-sample data.

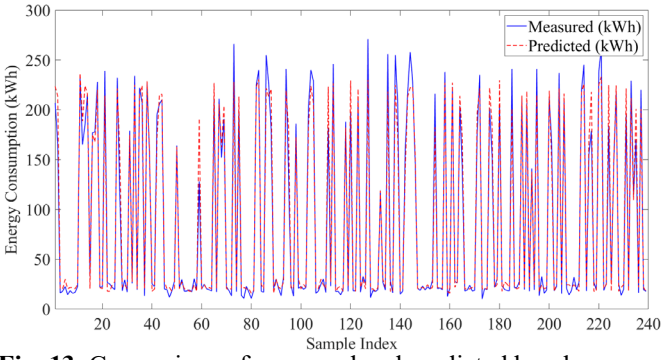

Fig. 13. Comparison of measured and predicted hourly energy (PSO-ANN, 23-neuron; testing set).

In addition to the measured-versus-predicted comparison, Fig. 14 overlays the absolute error (black line, right axis) onto the same time series of energy consumption. Plotting error magnitude alongside the measured and predicted curves clearly highlights the instances where the PSO-ANN's forecasts deviate most, with a few spikes which around sample index 40 approaching 65 kWh. Across the remainder of the testing window, absolute errors generally remain between 5 kWh and 15 kWh, reinforcing the low RMSE and MAE reported earlier. This combined view of prediction trace and point-wise error confirms that the 23-neuron PSO-ANN not only follows the overall energy-use pattern but does so with limited localized deviation, underscoring its reliability for hour-by-hour forecasting.

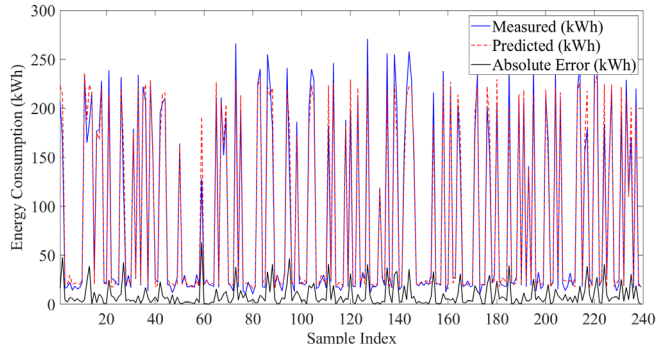

Fig. 14. Comparison of measured and predicted hourly energy with absolute error (PSO-ANN, 23-neuron; testing set).

IV. Statistical Robustness

PSO-based and neural models are stochastic, so single-run results can be influenced by random initialization, swarm dynamics, or data resampling. A robustness analysis across multiple independent runs offers a clearer view of repeatability and stability. Median and interquartile range (IQR) are reported because they are resistant to outliers and do not assume normality. For statistical comparison without a pairing of runs and without distributional assumptions, the Mann–Whitney rank-sum test is used; practical importance is conveyed by the Cliff's delta (δ) effect size, which quantifies the magnitude and direction of differences between two error distributions. By contrast, MLR is deterministic for a fixed train/test split; repeated refits are identical and therefore excluded from this robustness analysis.

Table VI summarizes test-set RMSE across 15 independent runs using quartiles. Q1 and Q3 denote the 25th and 75th percentiles, the Median is the 50th percentile, and IQR = Q3 − Q1 measures central dispersion (a smaller IQR indicates more consistent performance). For the ANN with 8-neuron, Q1 = 18.507 kWh, Median = 20.570 kWh, and Q3 = 21.334 kWh (IQR = 2.827 kWh). For PSO-ANN (23 neurons), Q1 = 18.251 kWh, Median = 19.125 kWh, and Q3 = 20.579 kWh (IQR = 2.327 kWh). Relative to ANN, the PSO-ANN distribution shifts downward: Q1 decreases by 0.2560 kWh and Q3 by 0.7553 kWh; the IQR contracts by 0.4993 kWh (≈ 17.7 %), indicating reduced run-to-run variability. The median improvement is 1.444 kWh (≈ 7.0 % relative to the ANN median). A Mann–Whitney U test yields p = 0.372, which is not significant at α = 0.05; Cliff's delta δ = 0.1911 indicates a small effect. Overall, PSO-ANN delivers slightly lower and more tightly dispersed errors, but the gain is modest at the current sample size.

TABLE VI. Statistical Robustness Across 15 Seeds (Test-Set RMSE).

Model	Q1	Median	Q3	IQR
ANN	18.507	20.570	21.334	2.827
PSO-ANN	18.251	19.125	20.579	2.327

V. Comparative Performance Analysis

The three modelling approaches reveal a clear and progressive performance gradient, as summarized in Table VII. The multiple linear regression (MLR) baseline captures the dominant linear trends, achieving $R^2 = 0.9410$ with MSE = 553.469 kWh². (RMSE = 23.526 kWh; CVRMSE = 24.912 %). Replacing the linear model with an 8-neuron feed-forward ANN improves predictive accuracy, increasing R^2 to 0.9728 (3.38 % vs. MLR) and reducing MSE to 238.593 kWh² (56.89 % reduction), reflecting the ANN's ability to capture nonlinear relationships between weather and occupancy. Further refinement using PSO to optimize ANN weights and biases yields a 23-neuron PSO-ANN with $R^2 = 0.9746$ (3.57 % vs. MLR) and MSE = 210.027 kWh² (62.05 % reduction), alongside RMSE and CVRMSE reductions of 38.40 % and 40.91 % relative to MLR. Relative to the ANN, PSO-ANN provides an additional 11.97 % reduction in MSE (6.18 % and 15.66 % reductions in RMSE and CVRMSE, respectively), indicating that global search-based weight optimization provides a measurable and incremental improvement.

TABLE VII. Performance Comparison Results for MLR, ANN, PSO-ANN.

Metric	MLR	ANN	PSO-ANN	ANN vs MLR (%)	PSO-ANN vs MLR (%)	PSO-ANN vs ANN (%)
Neurons	–	8	23	–	–	–
Rtrain	–	0.9810	0.9791	–	–	–
Rvalid	–	0.9699	0.9704	–	–	–
Rtest	–	0.9863	0.9872	–	–	–
Rall	0.9674	0.9802	0.9790	1.32	1.2	−0.12

R^2	0.9410	0.9728	0.9746	3.38	3.57	0.19
MSE (kWh^2)	553.469	238.593	210.027	56.89	62.05	11.97
RMSE (kWh)	23.526	15.446	14.492	34.34	38.4	6.18
CVRMSE (%)	24.912	17.453	14.720	29.94	40.91	15.66

VI. CONCLUSION

This study compared Multiple Linear Regression (MLR), an Artificial Neural Network (ANN), and a Particle-Swarm-Optimized ANN (PSO-ANN) for hourly baseline energy forecasting in a five-story institutional building using three predictors: outdoor temperature, student occupancy, and staff occupancy. Using the data partitions defined in methodology, MLR provided a transparent baseline (R^2 = 0.9410; MSE = 553 kWh^2; CVRMSE = 24.91 %). Replacing the linear framework with an 8-neuron feedforward ANN improved fidelity (R^2 = 0.9728; MSE = 239 kWh^2; CVRMSE = 17.45 %), highlighting the value of modeling nonlinearity. Further refinement via Particle Swarm Optimization produced a 23-neuron PSO-ANN with the highest coefficient and lowest errors (R^2 = 0.9746; MSE = 210 kWh^2; CVRMSE = 14.72 %). In robustness tests across 15 runs, PSO-ANN showed a lower median RMSE and a narrower IQR than ANN; however, the gain was modest, with a small effect size, and not statistically significant at $\alpha = 0.05$ (Mann–Whitney U test: $p = 0.3725$; Cliff's $\delta = 0.1911$). This indicates that PSO mainly improves stability and run-to-run consistency rather than delivering a reliably larger accuracy gain over ANN. Overall, the results highlight a progression in accuracy, error consistency, and generalizability from MLR to ANN to PSO-ANN. The trends suggest that PSO-ANN is a practical option for baseline energy modelling in institutional settings, particularly when consistency is valued. Operationally, the observed stability can support more precise energy benchmarking, earlier anomaly detection, and data-driven planning. Future research may generalize this framework to diverse building types and dynamic operating environments, while also integrating other metaheuristic algorithms to further enhance predictive performance without compromising reproducibility.

ACKNOWLEDGMENT

The authors gratefully acknowledge the Universiti Teknologi MARA, and the Faculty of Electrical Engineering, Universiti Teknologi MARA Cawangan Pulau Pinang, for their support and assistance in the preparation and publication of this paper.

REFERENCES

[1] National Renewable Energy Laboratory (NREL), International performance measurement, and verification protocol: Concepts and options for determining energy and water savings, Volume I (Revised), Rep. NREL/TP-710-31505, Golden, CO, USA, 2002, pp. 1–93. doi: https://doi.org/10.2172/15000242.

[2] R. Mustapa, N. Dahlan, A. Yassin, and A. Nordin, "Lecture and non-lecture week baseline energy model development and energy prediction in Malaysia educational building," J. Phys.: Conf. Ser., vol. 1432, 2020, Art. no. 012001. doi: https://doi.org/10.1088/1742-6596/1432/1/012001.

[3] H. B. Juričić and H. Krstić, "Comparing MLR and ANN models for school building electrical energy prediction in Osijek-Baranja County in Croatia," Energy Reports, vol. 12, 2024, pp. 3595–3606. doi: https://doi.org/10.1016/j.egyr.2024.09.039.

[4] W. Y. Hong and B. N. N. M. Nura'Liyah, "Lighting energy consumption estimation models for a library building with different lighting scenarios," in Proc. 2021 IEEE Asia-Pacific Conf. on Computer Science and Data Engineering (CSDE), 2021, pp. 1–5. doi: https://doi.org/10.1109/CSDE53843.2021.9718403.

[5] R. F. Mustapa, A. H. M. Nordin, M. A. Hairuddin, W. S. W. Ibrahim, S. A. M. Saleh, and N. Y. Dahlan, "Educational building's energy consumption independent variables analysis using linear regression model: A comparative study," in Proc. 2023 IEEE 3rd Int. Conf. in Power Engineering Applications (ICPEA), 2023, pp. 202–207. doi: https://doi.org/10.1109/ICPEA56918.2023.10093222.

[6] M. Khalil, A. S. McGough, Z. Pourmirza, M. Pazhoohesh, and S. Walker, "Machine learning, deep learning and statistical analysis for forecasting building energy consumption—A systematic review," Eng. Appl. Artif. Intell., vol. 115, 2022, Art. no. 105287. doi: https://doi.org/10.1016/j.engappai.2022.105287.

[7] S. J. Mohammed, H. Abdel-khalek, and S. M. Hafez, "Predicting performance measurement of residential buildings using machine intelligence techniques (MLR, ANN and SVM)," Iran. J. Sci. Technol., Trans. Civ. Eng., vol. 46, 2022, pp. 3429–3451. doi: https://doi.org/10.1007/s40996-021-00742-4.

[8] Z. Dong, J. Liu, B. Liu, K. Li, and X. Li, "Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification," Energy Build., vol. 241, 2021, Art. no. 110929. doi: https://doi.org/10.1016/j.enbuild.2021.110929.

[9] A. Ali, R. Jayaraman, A. Mayyas, B. Alaifan, and E. Azar, "Machine learning as a surrogate to building performance simulation: Predicting energy consumption under different operational settings," Energy Build., vol. 286, 2023, Art. no. 112940. doi: https://doi.org/10.1016/j.enbuild.2023.112940.

[10] A. Alshibani, "Prediction of the energy consumption of school buildings," Appl. Sci., vol. 10, 2020, Art. no. 5885. doi: https://doi.org/10.3390/app10175885.

[11] A. A. Jaber, A. A. Saleh, and H. F. M. Ali, "Prediction of hourly cooling energy consumption of educational buildings using artificial neural network," Int. J. Adv. Sci. Eng. Inf. Technol., vol. 9, 2019, pp. 159–166. doi: https://doi.org/10.18517/IJASEIT.9.1.7351.

[12] N. Nehra, P. Sangwan, and D. Kumar, "Artificial neural networks—A comprehensive review," in Handbook of Machine Learning for Computational Optimization, ch. 11. Boca Raton, FL, USA: CRC Press, 2021, pp. 203–227. doi: https://doi.org/10.1201/9781003138020-11.

[13] L. T. Le, H. Nguyen, J. Dou, and J. Zhou, "A comparative study of PSO-ANN, GA-ANN, ICA-ANN, and ABC-ANN in estimating the heating load of buildings' energy efficiency for smart city planning," Appl. Sci., vol. 9, 2019, Art. no. 2630. doi: https://doi.org/10.3390/app9132630.

[14] H. Chenglei, L. Kangji, L. Guohai, and P. Lei, "Forecasting building energy consumption based on hybrid PSO-ANN prediction model," in Proc. 2015 34th Chinese Control Conf. (CCC), 2015, pp. 8243–8247. doi: https://doi.org/10.1109/ChiCC.2015.7260948.

[15] S. E. Aslay, "Building energy prediction model with AI-based PSO-ANN approach integrating architectural and HVAC processes," Build. Environ., 2025, Art. no. 113305. doi: https://doi.org/10.1016/j.buildenv.2025.113305.

[16] R. Olu-Ajayi, H. Alaka, I. Sulaimon, F. Sunmola, and S. Ajayi, "Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques," J. Build. Eng., vol. 45, 2022, Art. no. 103406. doi: https://doi.org/10.1016/j.jobe.2021.103406.

[17] Q. Fu, K. Li, J. Chen, J. Wang, Y. Lu, and Y. Wang, "Building energy consumption prediction using a deep-forest-based DQN method," Buildings, vol. 12, 2022, Art. no. 131. doi: https://doi.org/10.3390/buildings12020131.

[18] M. Shariati, M. S. Mafipour, P. Mehrabi, A. Bahadori, Y. Zandi, M. N. A. Salih, H. Nguyen, J. Dou, X. Song, and S. Poi-Ngian, "Application of a hybrid artificial neural network–particle swarm optimization (ANN-PSO) model in behavior prediction of channel shear connectors embedded in normal and high-strength concrete," *Applied Sciences*, vol. 9, 2019, Art. no. 5534. doi: https://doi.org/10.3390/app9245534.

[19] J. Alsarraf, H. Moayedi, A. S. A. Rashid, M. A. Muazu, and A. Shahsavar, "Application of PSO–ANN modelling for predicting the exergetic performance of a building integrated photovoltaic/thermal

system," Eng. Comput., vol. 36, 2020, pp. 633–646. doi: https://doi.org/10.1007/s00366-019-00721-4.
[20] H. Moayedi, M. Mehrabi, M. Mosallanezhad, A. S. A. Rashid, and B. Pradhan, "Modification of landslide susceptibility mapping using optimized PSO-ANN technique," Eng. Comput., vol. 35, 2019, pp. 967–984. doi: https://doi.org/10.1007/s00366-018-0644-0.
[21] V. Gundu and S. P. Simon, "PSO–LSTM for short-term forecast of heterogeneous time series electricity price signals," J. Ambient Intell. Humaniz. Comput., vol. 12, 2021, pp. 2375–2385. doi: https://doi.org/10.1007/s12652-020-02353-9.
[22] S. Mahjoub, S. Labdai, L. Chrifi-Alaoui, L. Marhic, and L. Delahoche, "Short-term occupancy forecasting for a smart home using optimized weight updates based on GA and PSO algorithms for an LSTM network," Energies, vol. 16, 2023, Art. no. 1641. doi: https://doi.org/10.3390/en16041641.
[23] J. Wang, Z. Li, and C. Pan, "Energy-efficient trajectory planning with curve splicing based on PSO-LSTM prediction," Control Eng. Pract., vol. 150, 2024, Art. no. 106009. doi: https://doi.org/10.1016/j.conengprac.2024.106009.
[24] P. J. Muhammad Ali, "Investigating the Impact of Min-Max Data Normalization on the Regression Performance of K-Nearest Neighbor with Different Similarity Measurements," ARO—The Scientific Journal of Koya University, vol. 10, no. 1, pp. 85–91, Jun. 2022, doi: 10.14500/aro.10955
[25] A. Jierula, S. Wang, T.-M. Oh, and P. Wang, "Study on Accuracy Metrics for Evaluating the Predictions of Damage Locations in Deep Piles Using Artificial Neural Networks with Acoustic Emission Data," Applied Sciences, vol. 11, no. 5, Art. no. 2314, Mar. 2021, doi: 10.3390/app11052314.
[26] ASHRAE, *ASHRAE Guideline 14-2023: Measurement of Energy, Demand and Water Savings*. Atlanta, GA, USA: ASHRAE, 2023.
[27] Bonneville Power Administration (BPA), Regression for M&V: Reference Guide, 2023. [Online]. Available: https://www.bpa.gov/-/media/Aep/energy-efficiency/measurement-verification/3-bpa-mv-regression-reference-guide.pdf
[28] [28] C. M. Bishop and H. Bishop, *Deep Learning: Foundations and Concepts*. Springer, 2024. doi: 10.1007/978-3-031-45468-4.
[29] S. Bouraya and A. Belangour, "A comparative analysis of activation functions in neural networks: unveiling categories," *Bulletin of Electrical Engineering and Informatics*, vol. 13, no. 5, pp. 3301–3308, 2024.
[30] M. Gustineli, "A survey on recently proposed activation functions for deep learning," preprint, 2022. [Online]. Available: https://doi.org/10.31224/2245
[31] MathWorks, "trainlm — Levenberg–Marquardt backpropagation," MATLAB Deep Learning Toolbox Documentation. Accessed: Jan. 2026. [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/trainlm.html
[30] D. Freitas, L. G. Lopes, and F. Morgado-Dias, "Particle Swarm Optimisation: A Historical Review Up to the Current Developments," *Entropy*, vol. 22, no. 3, Art. no. 362, Mar. 2020, doi: 10.3390/e22030362.
[31] E. H. Houssein, A. G. Gad, K. Hussain, and P. N. Suganthan, "Major Advances in Particle Swarm Optimization: Theory, Analysis, and Application," *Swarm and Evolutionary Computation*, vol. 63, Art. no. 100868, 2021, doi: 10.1016/j.swevo.2021.100868.
[32] A. P. Piotrowski, J. J. Napiorkowski, and A. E. Piotrowska, "Population Size in Particle Swarm Optimization," *Swarm and Evolutionary Computation*, vol. 58, Art. no. 100718, Nov. 2020, doi: 10.1016/j.swevo.2020.100718.
[33] E. Twumasi, E. A. Frimpong, N. K. Prah II, and D. B. Gyasi, "A novel improvement of particle swarm optimization using an improved velocity update function based on local best murmuration particle," *Journal of Electrical Systems and Information Technology*, vol. 11, art. no. 42, Oct. 2024, doi: 10.1186/s43067-024-00168-8.
[34] D. E. Ratnawati, M. Marjono, W. Widodo, and S. Anam, "PSO-ELM with Time-varying Inertia Weight for Classification of SMILES Codes," *International Journal of Intelligent Engineering and Systems*, vol. 13, no. 6, pp. 522–532, 2020, doi: 10.22266/ijies2020.1231.46
[35] Y. M. Hendrawan, F. H. Anshori, A. Pratama, H. B. Harja, Pandoe, and A. P. Danadibrata, "Implementation of Particle Swarm Optimization (PSO) Method in Minimum Quantity Lubrication (MQL) Optimization to Obtain Optimal Machining in CNC Milling Machine," in Proc. 5th Int. Conf. on Applied Science and Technology on Engineering Science (iCAST-ES 2022), pp. 821–827, 2023, doi: 10.5220/0011891500003575
[36] M. Jain, V. Saihjpal, N. Singh, and S. B. Singh, "An Overview of Variants and Advancements of PSO Algorithm," Applied Sciences, vol. 12, no. 17, Art. no. 8392, 2022, doi: 10.3390/app12178392.