ENERGY CONSUMPTION PREDICTION IN SMART FACTORIES USING DATA-DRIVEN DEEP LEARNING MODELS

Thi Phuong Quyen Nguyen*, Thi My Ha Nguyen, Thi Cuc Nguyen, Nguyen Phuong Thao Nguyen

The University of Danang - University of Science and Technology, Vietnam

*Corresponding author: ntpquyen@dut.udn.vn

(Received: February 06, 2026; Revised: March 17, 2026; Accepted: March 19, 2026)

DOI: 10.31130/ud-jst.2026.24(3).087E

Abstract - Accurate energy consumption prediction plays a critical role in improving efficiency and sustainability in smart factories. This study develops a data-driven framework based on deep learning (DL) models for short-term energy consumption forecasting using multivariate time-series sensor data. Three DL models are employed: Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM), and Deep Recurrent Neural Networks (DRNNs). To enhance model performance, Genetic Algorithm (GA) is integrated to optimize key hyperparameters related to temporal dependency, network configuration, and training strategy. Experimental results on a real-world smart manufacturing dataset demonstrate that GA-based DL models consistently outperform their non-optimized counterparts. GA-optimized models achieve an average reduction of approximately 8–15% in Mean Absolute Percentage Error (MAPE) on the test set, with the GA-CNN model demonstrating the best overall performance. These results confirm the effectiveness of combining evolutionary optimization with DL for robust and accurate energy consumption prediction in smart factory environments.

Key words - Deep learning; energy consumption prediction; data-driven; smart manufacturing

1. Introduction

The rapid evolution of Industry 4.0 and the widespread adoption of sensor networks and industrial Internet of Things (IIoT) technologies have transformed traditional manufacturing environments into smart factories, where large volumes of operational data are continuously generated [1]. Efficient energy consumption management has become a critical priority due to its direct impact on operational costs, sustainability, and environmental performance in industrial systems. Accurately forecasting energy consumption not only enables proactive energy planning but also facilitates adaptive control strategies that can significantly reduce waste and enhance overall productivity [2].

Traditional energy prediction approaches, often based on statistical or traditional machine learning models, struggle to capture the complex temporal and nonlinear relationships inherent in multivariate industrial energy data. Several studies in the literature highlight the limitations of conventional methods and the potential of advanced machine learning algorithms. For instance, support vector regression and random forest models are commonly used in energy forecasting, but have limitations when modeling long-term dependencies and the non-linearity of consumption data [3].

In contrast, deep learning (DL) techniques such as Deep Recurrent Neural Networks (DRNNs) [4], Long Short-Term Memory (LSTM) [5], and Convolutional Neural Networks (CNNs) [6] have demonstrated superior performance in forecasting applications by automatically learning hierarchical feature representations and capturing temporal dynamics in time-series data. Prior research has shown the effectiveness of hybrid approaches combining CNN and LSTM structures for load forecasting in buildings and grid systems, indicating enhanced predictive capability compared to standalone models [3].

Despite these advances, most existing work focuses on building or grid-level forecasting problems [7]. Research specifically addressing industrial energy consumption forecasting in smart factory settings remains limited. Smart manufacturing data is often high-dimensional and noisy due to complex process interactions, which requires robust models for accurate forecasting. Although DL models have shown strong potential in capturing complex patterns, their performance is highly sensitive to the choice of temporal window size and model hyperparameters [8]. Inappropriate parameter settings can significantly degrade prediction accuracy. Therefore, adaptive optimization strategies such as Genetic Algorithms (GAs) are needed to tailor DL models to the specific characteristics of industrial data. This integration helps improve prediction robustness and supports effective energy management in smart factory systems.

The primary contributions of this work are outlined below:

1. Propose a data-driven framework that integrates GA optimization with DL models for energy consumption prediction in smart factories.

2. Conduct a comprehensive evaluation on real multivariate manufacturing process data, comparing GA-optimized CNN, LSTM, and DRNN models under a unified experimental setting using Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE) metrics.

3. Demonstrate that GA-enhanced CNN models can outperform recurrent architectures in industrial energy forecasting, highlighting the importance of local temporal pattern extraction for manufacturing process-level energy management.

2. Related works

Predicting energy consumption in industrial and manufacturing environments has attracted growing research interest due to its importance for cost control, operational efficiency, and sustainable production. Early studies in this area mainly relied on statistical and conventional machine learning methods, such as linear regression, autoregressive integrated moving average (ARIMA), support vector regression (SVR), and random forest, to model industrial energy consumption patterns [9]. While these approaches offer interpretability, they often struggle to capture the nonlinear relationships and temporal dependencies present in manufacturing sensor data. This limitation has motivated the adoption of more advanced, data-driven learning methods. Several comparative studies on real-world energy time series have reported that DL techniques generally outperform traditional models in forecasting accuracy, highlighting their potential for industrial energy prediction tasks [10].

In recent years, DL models have been widely adopted for industrial energy consumption forecasting due to their ability to learn temporal representations directly from multivariate time-series data. Convolutional and recurrent architectures have been widely explored to capture both local and sequential patterns in energy data [11]. Oliveira et al. [12] compared CNN, LSTM, CNN–LSTM, and temporal convolutional network (TCN) models for short-term industrial energy consumption forecasting, highlighting the robustness of deep architectures under varying manufacturing conditions. Other studies focusing on specific industrial sectors, such as steel manufacturing, have benchmarked machine learning and DL approaches, including ANN, k-NN, random forest, and LSTM, with recurrent models often achieving superior performance [13]. More recently, advanced frameworks such as attention-based recurrent networks have been proposed to improve forecasting accuracy in using heterogeneous industrial sensor data [8].

Despite their advantages, DL models remain highly sensitive to hyperparameter configurations, including temporal window size, network depth, and the number of hidden units. Suboptimal parameter settings can lead to unstable training behavior and degraded prediction performance [14]. To mitigate these issues, metaheuristic optimization techniques, particularly Genetic Algorithms, have been increasingly employed to optimize DL models. GA has proven effective in exploring large and complex search spaces without relying on gradient information [15]. Previous studies have shown that GA-based optimization can improve both convergence stability and forecasting accuracy in time-series prediction tasks [16, 17]. However, the use of GA-optimized DL frameworks for energy consumption prediction at the manufacturing process level remains relatively limited.

3. Methodology

This study proposes a data-driven framework for forecasting energy consumption in smart factories, consisting of four main stages: (i) collecting and preprocessing production data, (ii) developing and training DL models (CNN, DRNN, LSTM), (iii), developing a GA-based DL models to improve prediction accuracy, and (iv) evaluating model performance using consistent criteria. The proposed framework is designed to ensure fair comparisons across different models while accurately reflecting the characteristics of real-world industrial data.

3.1. Data Collection and Pre-processing

The dataset used in this study was collected from a smart manufacturing process and is publicly available on the Kaggle platform [18]. This is a time-series dataset collected from a sensor-based smart manufacturing system, intended to capture the operating behavior of equipment and production processes through a variety of operational parameters. The data include measured variables such as machine temperature, speed, vibration levels, production quality score, optimal condition, and energy consumption recorded at multiple sensor points across the factory. The dataset is organized in a tabular format, with each record representing a continuous observation time point and containing multiple sensor features related to the production process.

The data are structured as a multivariate time series and sampled at fixed time intervals, making them suitable for short-term and medium-term energy forecasting tasks in a smart factory environment. Since the sensor features are measured on different scales (e.g., temperature in °C and energy consumption in kW), the data are normalized using Min–Max scaling to bring all variables onto a common scale. This helps prevent certain variables from dominating others during model training.

Min–Max scaling is applied to map all input features to the same value range, namely $[0, 1]$. For each feature x, the normalized value x' is computed using the following formula:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where, $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values of that feature in the training dataset, respectively. To prevent data leakage, the parameters $x_{\min}$ and $x_{\max}$ are estimated solely from the training set and then applied consistently to the validation and test sets. This approach ensures that information from future data does not influence the model training process.

Min–Max scaling is commonly used for DL models. It usually uses activation functions such as sigmoid or ReLU to maintain a stable input range and improve convergence speed during training. In addition, this method enhances the stability of distance-based machine learning models, such as SVR.

Because the data are time series, the dataset is split based on temporal order to prevent information leakage from future observations. Specifically, the first 70% of the data are used as the training set, while the remaining 30% from later periods are reserved for testing. This splitting strategy ensures that the models rely solely on past information to forecast future energy consumption, thereby accurately reflecting their generalization capability in real-world smart factory environments.

3.2. Deep Learning Architectures

3.2.1. 1D Convolutional Neural Network – 1D CNN

A one-dimensional convolutional neural network (1D-CNN) model was developed to predict energy consumption in a smart factory, using multivariate time-series data collected from a smart manufacturing system. The dataset used in the study includes features reflecting the operating state of the production process, such as sensor measurements, process variables, and energy consumption-related indicators at each point in time. These features are organized as time series and used as input to the CNN model. Figure 1 illustrates the architecture of the CNN model applied in this study.

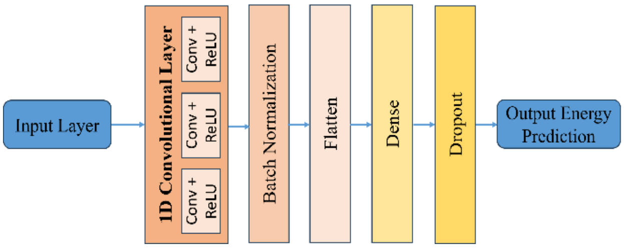

Figure 1. *One-dimensional CNN architecture for predicting energy consumption in smart factories*

The structure of the CNN model includes the following main layers:

• Input layer: the input layer receives multivariate one-dimensional time series data, where each sample consists of successive observations over time of features in the smart manufacturing dataset. Specifically, the input features include variables representing production line operating conditions (e.g., sensor measurements, equipment parameters, and state variables). These variables are grouped within a fixed time window to predict energy consumption at the next time point.

• One-dimensional convolutional layer (Conv1D): a one-dimensional convolutional layer with 32 filters and a kernel size of 3 is employed to extract short-term temporal features from the input sequence. Same padding and ReLU activation are applied to preserve sequence length and enhance nonlinearity.

• Batch Normalization Layer: applied after the Conv1D layer to normalize feature distributions to improve training stability and convergence.

• Flatten Layer: transforms the convolutional output into a one-dimensional vector for input to the fully connected layers.

• Dense Layer: a fully connected layer with 64 neurons and ReLU activation learns nonlinear relationships between extracted features and energy consumption.

• Dropout Layer: a dropout rate of 0.3 is used to reduce overfitting during training.

• Output Layer: consists of a single neuron with linear activation to produce the predicted energy consumption value.

Unlike traditional CNN architectures designed for two-dimensional image data, the CNN model in this study has been adapted to process multivariate time-series data for smart factory energy consumption forecasting.

3.2.2. Long Short-Term Memory – LSTM

The LSTM model is widely popular in the DL community to address the limitations of traditional feedback neural networks (RNNs) in long-term information retention and the gradient decay problem during training. In this study, LSTM was used to learn the temporal relationship between past operating states and current energy consumption. Figure 2 illustrates the LSTM architecture, which includes the following layers:

• Input layer: receives the input time series in window size format and the number of input variables.

• LSTM Layer 1: consists of 128 hidden units, uses the activation function "*tanh*", and returns sequences = "*True*" to retain the output sequence for the next layer. This layer learns basic sequential features from the time series.

• LSTM Layer 2: consists of 64 hidden units, continuing to learn more advanced features. Setting the parameter "return_sequences" to "*True*" allows the model to continue processing the sequence through multiple layers.

• LSTM Layer 3: consists of 32 hidden units, the final layer with the return sequences = "*False*" to return the final hidden state to be fed into the Dense layer. This layer synthesizes the information from the entire sequence to make a prediction.

• Dropout Layer: after each LSTM layer, a Dropout layer at a rate of 0.2 is used to reduce overfitting by randomly discarding 20% of the connections during training.

• Dense Layer (Full Connection): consists of one output neuron with a continuous value, used to predict future energy consumption.

Conventional LSTM architectures often use 1 to 2 LSTM layers with a large number of hidden units to exploit time-series features [52]. However, in energy consumption forecasting for smart manufacturing environments using multivariate time-series data, exploiting deep features across multiple time levels is very important. The proposed model in this study employs three LSTM layers with progressively fewer numbers of hidden units (128 → 64 → 32). A dropout rate of 0.2 is applied after each layer to help the model learn general features in the upper layers while refining features in the lower layers.

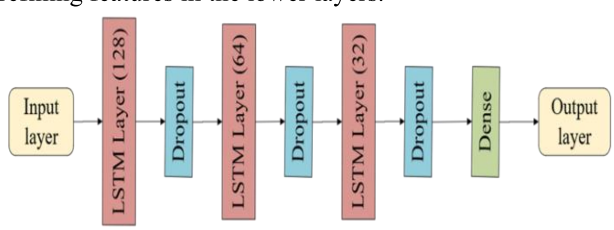

Figure 2. *LSTM model architecture for predicting energy consumption*

3.2.3. Deep Recurrent Neural Network (DRNN)

DRNNs extend traditional recurrent neural networks by stacking multiple feedback layers to enhance the representation and modeling of complex temporal relationships in time-series data. Unlike LSTMs, which use gating mechanisms to control the information flow, DRNNs rely on deeper feedback layers to learn hierarchical representations of time series. Figure 3 presents the DRNN model architecture for predicting energy consumption.

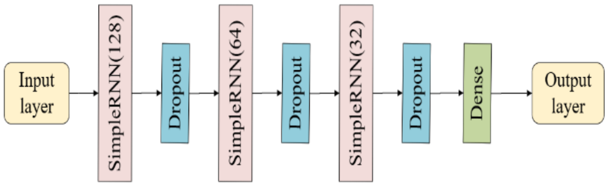

Figure 3. *DRNN model architecture for predicting energy consumption*

The model structure is similar to an LSTM, consisting of 3 SimpleRNN layers with fewer hidden units (128 → 64 → 32), combined with Dropout layers applied after each layer at a rate of 0.2. This structure allows the model to learn features over time at multiple levels, in which the first layers learn the overall trend while the later layers refine the information. The fewer hidden units help the model avoid overloading the last layer with too many parameters, reducing complexity and the risk of overfitting. This design ensures the model is deep enough to capture complex relationships in energy consumption data while maintaining stability and generalizability during training.

3.2.4. A combination of GA and DL models

To improve the predictive accuracy of energy consumption and reduce the dependence on manual hyperparameter selection, this study proposes a unified GA-based hyperparameter optimization framework for CNN, LSTM, and DRNN models in smart factory energy consumption forecasting. In this approach, GA is used as a global optimization mechanism to automatically search for optimal hyperparameter configurations for each model based on multivariate time-series data of the production process in a smart factory. Each individual in the GA population represents a specific DL model configuration, where the chromosomes encode the key hyperparameters of the model. The fitness function is determined based on the predictive performance on the validation dataset, using error measures such as RMSE or MAPE.

In this study, GA is used to optimize key hyperparameters of DL models. For CNN, these include the number of filters, kernel size, number of dense neurons, dropout rate, and learning rate. For the LSTM, the optimized hyperparameters are the number of LSTM layers, dropout rate, learning rate, and batch size. For DRNN, GA optimizes the number of recurrent layers, dropout rate, batch size, and learning rate.

The integration of GA with DL models is illustrated in Figure 4, which can be described as follows:

Step 1: Initialize the population with randomly generated individuals (population size of 80) based on a hyperparameter configuration of CNN, LSTM, or DRNN.

Step 2: Build and train a DL model tailored to each individual on the training dataset.

Step 3: Evaluate the model's performance on the validation set to compute the fitness function value.

Step 4: Select individuals using the roulette wheel method for genetic operation. The crossover and mutation processes employ the real-coded GA as follows:

Crossover:

$$O_1 = \alpha \times P_1 + (1-\alpha) \times P_2 \qquad (2)$$

$$O_2 = \beta \times P_1 + (1-\beta) \times P_2 \qquad (3)$$

Mutation:

$$O_i = \beta O_i + s_i \times r_i \times q_i \qquad (4)$$

where, P is the population, O is the offspring, and α and β are two random numbers in the range of (0,1). Parameters for the mutation process, such as s_i, r_i, q_i is set based on previous research [19].

Step 5: Repeat the process until the maximum number of generations, which is set at 100, is reached or until the optimization process converges.

Step 6: Predict energy consumption using CNN, LSTM, and DRNN models based on the optimal parameters obtained from GA.

Integrating GA with various DL architectures allows for a systematic assessment of the role of hyperparameter optimization in industrial energy consumption forecasting. This approach not only enhances model flexibility but also reduces dependence on manual experience in network design, particularly suitable for highly nonlinear multivariate time series data.

4. Experimental Results

4.1. Experimental Setup and Evaluation Metrics

The experiments were conducted on a real-world multivariate time-series dataset collected from a smart manufacturing process, in which multiple operational sensor variables were used to predict energy consumption.

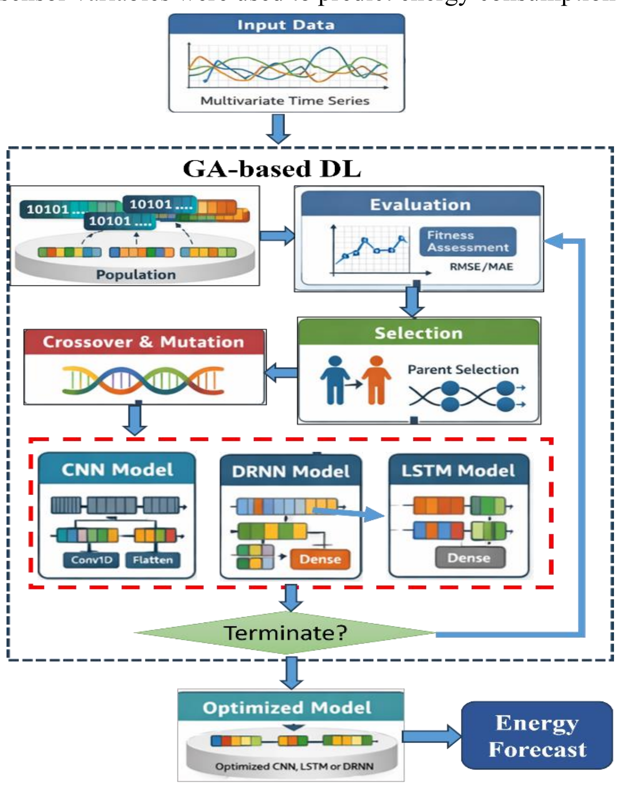

Figure 4. *Overall framework of the proposed GA-based DL approach*

To evaluate the forecasting performance of the proposed GA-based DL models, three widely used regression metrics - RMSE, MAE, and MAPE were adopted. RMSE penalizes large prediction errors and reflects overall model stability, while MAE measures the average magnitude of prediction deviations. MAPE expresses the relative prediction error in percentage form and is particularly suitable for energy consumption forecasting under varying operating conditions.

The GA was applied to optimize the temporal window size and feature importance weights for each DL architecture. Table 1 reports the GA-optimized parameter configurations obtained for the CNN, LSTM, and DRNN models after convergence. Differences in the selected window sizes indicate that each architecture captures temporal dependencies at different scales, while the optimized feature weights reflect the varying influence of process variables on energy consumption prediction.

Table 1. *GA-optimized parameter configurations for the CNN, LSTM, and DRNN*

Category	Parameter	CNN	LSTM	DRNN
Architecture	Num filters	80	–	–
	Kernel size	5	–	–
	LSTM units	–	128	128
	LSTM units	–	64	64
	LSTM units	–	32	32
Regularization	Dropout rate	0.3	0.3	0.3
Training	Learning rate	0.001	0.001	0.001
	Batch size	80	64	64

4.2. Prediction performance comparison

Table 2-3 presents the forecasting performance of the baseline DL models and their GA-optimized counterparts on both training and testing datasets, evaluated using RMSE, MAE, and MAPE.

Regarding Table 2, for the baseline models, CNN achieves the lowest training errors among the three conventional DL approaches, with RMSE, MAE, and MAPE values of 0.278, 0.238, and 16.576%, respectively. In contrast, DRNN and LSTM exhibit slightly higher training errors to indicate a comparable but less effective fitting capability when trained without hyperparameter optimization.

After applying GA optimization, a substantial improvement is observed for all models. GA-CNN achieves the best training performance, reducing RMSE to 0.213, MAE to 0.183, and MAPE to 8.692%. Compared to the baseline CNN, this corresponds to a reduction of approximately 47.5% in MAPE. This result demonstrates the effectiveness of GA in enhancing the model's learning capacity. Similarly, GA-DRNN and GA-LSTM also show notable error reductions, achieving training MAPE values of 12.597% and 12.248%, respectively. However, their training errors remain higher than those of GA-CNN.

Table 2. *Experimental result on the training dataset*

	RMSE	MAE	MAPE (%)
CNN	0.278 ± 0.001	0.238 ± 0.002	16.576± 0.001
DRNN	0.287 ± 0.001	0.248 ± 0.003	17.401 ± 0.003
LSTM	0.287 ± 0.002	0.247 ± 0.002	17.433 ± 0.003
GA-CNN	0.213 ± 0.001	0.183 ± 0.001	8.692 ± 0.001
GA-DRNN	0.238 ± 0.002	0.206 ± 0.001	12.597 ± 0.002
GA-LSTM	0.234 ± 0.001	0.198 ± 0.002	12.248 ± 0.001

Table 3. *Experimental result on the testing dataset*

	RMSE	MAE	MAPE (%)
CNN	0.295 ± 0.001	0.254 ± 0.001	17.625 ± 0.001
DRNN	0.292 ± 0.002	0.253 ± 0.001	17.704 ± 0.002
LSTM	0.292 ± 0.003	0.253 ± 0.002	17.774 ± 0.002
GA-CNN	0.226 ± 0.001	0.211 ± 0.001	9.225 ± 0.001
GA-DRNN	0.245 ± 0.001	0.235 ± 0.002	13.284 ± 0.002
GA-LSTM	0.241 ± 0.002	0.228 ± 0.002	13.016 ± 0.001

Table 3 summarizes the prediction results on the testing dataset to reflect the generalization ability of the evaluated models. The result is quite similar to the training dataset. Among the baseline approaches, CNN, DRNN, and LSTM exhibit nearly identical testing performance, with MAPE values ranging from 17.625% to 17.774%. Incorporating GA optimization leads to significant performance gains across all models. GA-CNN again outperforms the other approaches, achieving the lowest testing RMSE (0.226), MAE (0.211), and MAPE (9.225%). Compared to the baseline CNN, the testing MAPE is reduced by approximately 47.7%. GA-DRNN and GA-LSTM also demonstrate improved predictive accuracy, with testing MAPE values of 13.284% and 13.016%, respectively. Although these results confirm the effectiveness of GA-based optimization for recurrent models, their predictive performance is still lower than that of the GA-CNN model. This indicates that the convolutional architecture, after GA optimization, is more effective in learning the dominant temporal patterns in the testing data. As a result, GA-CNN demonstrates better generalization capability for the considered forecasting task.

Besides, the convergence behavior in Figure 5 indicates that all models reach stable training performance after approximately 80 epochs, beyond which the training error remains nearly unchanged. GA-optimized models converge noticeably faster than their baseline counterparts, reflecting the effectiveness of GA in selecting suitable hyperparameters and temporal configurations. In particular, GA-CNN achieves the lowest training RMSE and exhibits a smoother convergence trajectory, suggesting improved training stability. These observations further confirm the role of GA in enhancing both convergence speed and optimization efficiency for DL–based energy forecasting models.

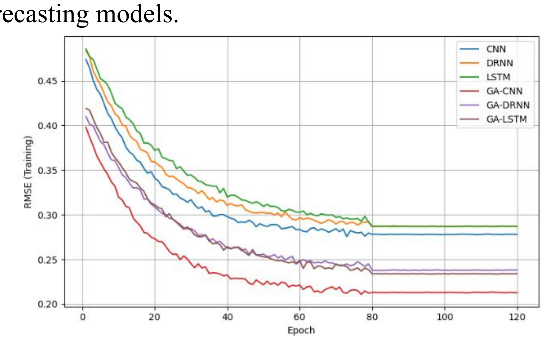

Figure 5. *Comparison of convergence plots*

5. Conclusion

This study developed a GA-optimized DL framework for energy consumption prediction in smart factories using multivariate manufacturing process data. The experimental results show that GA significantly improves the forecasting performance of CNN, LSTM, and DRNN models, with GA-CNN achieving the best accuracy across RMSE, MAE, and MAPE metrics. The convergence analysis further confirms that GA enhances training stability and accelerates model convergence, particularly for convolution-based architectures.

Future work will focus on multi-objective GA optimization to jointly minimize prediction error and computational cost, and on integrating hybrid and attention-based models to improve robustness under more complex industrial operating conditions. Besides, the GA model can explore more advanced selection and search strategies to better balance exploration and exploitation during the optimization process.

Acknowledgments: This research is funded by Funds for Science and Technology Development of the University of Danang under project number B2024-DN02-23. This support is much appreciated.

REFERENCES

[1] M. Soori, B. Arezoo, and R. Dastres, "Internet of things for smart factories in industry 4.0, a review," *Internet of Things and Cyber-Physical Systems,* vol. 3, pp. 192-204, 2023.

[2] Y. Hu and Y. Man, "Energy consumption and carbon emissions forecasting for industrial processes: Status, challenges and perspectives," *Renewable and Sustainable Energy Reviews,* vol. 182, p. 113405, 2023.

[3] Y. Dou, S. Tan, and D. Xie, "Comparison of machine learning and statistical methods in the field of renewable energy power generation forecasting: a mini review," *Frontiers in Energy Research,* vol. 11, p. 1218603, 2023.

[4] C. Fan, J. Wang, W. Gang, and S. Li, "Assessment of deep recurrent neural network-based strategies for short-term building energy predictions," *Applied energy,* vol. 236, pp. 700-710, 2019.

[5] C. Zhou, Z. Fang, X. Xu, X. Zhang, Y. Ding, and X. Jiang, "Using long short-term memory networks to predict energy consumption of air-conditioning systems," *Sustainable Cities and Society,* vol. 55, p. 102000, 2020.

[6] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems,* vol. 33, no. 12, pp. 6999-7019, 2021.

[7] N.-Q. Nguyen, P-T-N Nguyen, and Q.-C. Truong, "Short-term prediction of regional energy consumption by metaheuristic optimized deep learning models," *The University of Danang - Journal of Science and Technology,* Vol. 22, No. 11C, pp. 109-114, 2024.

[8] I. K. Nti, M. Teimeh, O. Nyarko-Boateng, and A. F. Adekoya, "Electricity load forecasting: a systematic review," *Journal of Electrical Systems and Information Technology,* vol. 7, no. 1, p. 13, 2020.

[9] N. Majeske *et al.*, "Industrial energy forecasting using dynamic attention neural networks," *Energy AI,* vol. 20, p. 100504, 2025.

[10] S. Noureen, S. Atique, V. Roy, and S. Bayne, "A comparative forecasting analysis of arima model vs random forest algorithm for a case study of small-scale industrial load," *International Research Journal of Engineering and Technology,* vol. 6, no. 09, pp. 1812-1821, 2019.

[11] A. Nanjar, R. E. Saputro, and B. Berlilana, "Machine Learning and Deep Learning Approaches for Energy Prediction: A Systematic Literature Review," *Sinkron: jurnal dan penelitian teknik informatika,* vol. 8, no. 4, pp. 2603-2614, 2024.

[12] S. Navale, N. Mishra, and S. Borhade, "Deep learning approaches for energy consumption forecasting: analyzing stress factors and optimizing models for future demand," *Discover Applied Sciences,* vol. 7, no. 10, p. 1092, 2025.

[13] N. Oliveira, N. Sousa, and I. Praça, "Deep Learning for Short-Term Instant Energy Consumption Forecasting in the Manufacturing Sector," in *International Symposium on Distributed Computing and Artificial Intelligence*, 2022, pp. 165-175: Springer.

[14] K. Kerdprasop, N. Kerdprasop, and P. Chuaybamroong, "Deep Learning and Machine Learning Models to Predict Energy Consumption in Steel Industry," *Int. J. Mach. Learn,* vol. 13, pp. 142-145, 2023.

[15] X. Song, L. Deng, H. Wang, Y. Zhang, Y. He, and W. Cao, "Deep learning-based time series forecasting," *Artificial Intelligence Review,* vol. 58, no. 1, p. 23, 2024.

[16] T. Alam, S. Qamar, A. Dixit, and M. Benaida, "Genetic algorithm: Reviews, implementations, and applications," *arXiv preprint arXiv:2007.12673,* 2020.

[17] M. M. Al Haromainy, D. A. Prasetya, and A. P. Sari, "Improving performance of RNN-based models with genetic algorithm optimization for time series data," *TIERS Information Technology Journal,* vol. 4, no. 1, pp. 16-24, 2023.

[18] Y. Cao, J. Yu, R. Zhong, and M. Munetomo, "Forecasting Renewable energy and electricity consumption using evolutionary hyperheuristic algorithm," *Scientific Reports,* vol. 15, no. 1, p. 2565, 2025.

[19] P. Developer, "Smart Manufacturing Process Data," ed. Kaggle Competition, 2025, [Online]. Available: https://www.kaggle.com/datasets/programmer3/smart-manufacturing-process-data/data, [Accessed: January 15, 2026].

[20] T. P. Q. Nguyen, R. Kuo, M. D. Le, T. C. Nguyen, and T. H. A. Le, "Local search genetic algorithm-based possibilistic weighted fuzzy c-means for clustering mixed numerical and categorical data," *Neural Computing and Applications,* vol. 34, no. 20, pp. 18059-18074, 2022.