Journal of Energy Research and Reviews

Volume 16, Issue 6, Page 13-26, 2024; Article no.JENRR.117020
ISSN: 2581-8368

Forecasting U.S. Fossil Energy Consumption: Advancing Accuracy with Multilayer Perceptron and Residual Learning

Yiwu Hao [a*] and Qingping He [a]

[a] *School of Science, Southwest University of Science and Technology, Mianyang, China.*

Authors' contributions

This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information

DOI: https://doi.org/10.9734/jenrr/2024/v16i6354

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here:
http://www.sdiarticle5.com/review-history/117020

Received: 10/03/2024
Accepted: 13/05/2024
Published: 18/05/2024

Original Research Article

ABSTRACT

Accurate midterm forecasts of fossil fuel consumption are essential for effective energy planning, economic management, and resource allocation. While machine learning models have demonstrated their efficacy in handling large-scale nonlinear datasets, many, including Multilayer Perceptrons (MLPs), suffer from performance degradation with increased depth. Fortunately, recent studies have revealed that Residual Networks (ResNets) can mitigate or even overcome this challenge. In this paper, we propose a Weighted Residual Network based on MLP to enhance predictive performance. We employ the Adam algorithm for model training and utilize the Gridsearch algorithm for hyperparameter tuning. In the application section, we develop predictive models using three case studies: natural gas, petroleum, and total fossil fuel consumption.

**Corresponding author: Email: cn_h_math@mails.swust.edu.cn;*

Cite as: *Hao, Y., He, Q. (2024). Forecasting U.S. Fossil Energy Consumption: Advancing Accuracy with Multilayer Perceptron and Residual Learning. Journal of Energy Research and Reviews, 16(6), 13–26. https://doi.org/10.9734/jenrr/2024/v16i6354*

We validate the effectiveness of our proposed model and compare it with ten other machine learning models. Our findings demonstrate that our proposed model consistently outperforms others in all three cases, underscoring its superior performance in midterm forecasting of fossil fuel consumption.

Keywords: Weighted residual network; multilayer perceptron; adaptive moment estimation; grid Search; fossil energy consumption.

1 INTRODUCTION

Accurate mid-term forecasts of fossil energy consumption are crucial for energy planning, economic management, and resource allocation. Numerous studies have shown the effectiveness of machine learning models in handling large-scale nonlinear datasets, such as NOA-LSTM[1], frequency-domain MLPs[2], CNN-GRU[3].

Multilayer Perceptron (MLP) is a fundamental neural network architecture in machine learning models. Frank Rosenblatt introduced the concept of a layered perceptron network in his seminal work "Perceptron" [4][5][6]. His perceptrons comprised an input layer, a hidden layer with static weights, and an output layer with trainable connections. However, this model was initially perceived as an extreme learning machine rather than a deep learning network [7]. Despite this classification, Rosenblatt's work laid the foundation for subsequent advancements in neural network architectures. In 1965, Alexey Grigorevich Ivakhnenko and Valentin Lapa published the pioneering Group Method of Data Handling, representing the first instance of a deep-learning feedforward network. Notably, this method did not incorporate stochastic gradient descent [8][9]. Two years later, Shun'ichi Amari introduced a deep-learning network capable of classifying non-linearly separable pattern classes, employing stochastic gradient descent for the first time in such networks [10]. Amari's team also constructed a five-layered feedforward network, underscoring the viability of deep learning architectures. The modern backpropagation method, a pivotal aspect of MLP training, was introduced by Seppo Linnainmaa in 1970 [11]. This application of chain-rule-based supervised learning revolutionized neural network training by facilitating error propagation for parameter updates. Subsequent enhancements to the backpropagation algorithm, such as its formalization by Paul Werbos in 1982 [12], and experimental investigations by David E. Rumelhart et al. in 1985 [13], further underscored its significance in the realm of deep learning. To date, MLP has undergone significant development and finds widespread application in energy consumption prediction[17][18][19].

The depth of the MLP model is constrained by the vanishing gradient problem, posing challenges in training deeper networks. In addressing this issue, Sepp Hochreiter introduced skip connections, also known as residual connections, within the long short-term memory (LSTM) recurrent neural network architecture in 1991 [14]. These connections allow gradients to flow more effectively through the network, mitigating the vanishing gradient problem. Subsequently, in 2015, the concept of Highway Networks emerged as a further refinement. Inspired by the forget gates utilized in LSTM networks, Highway Networks integrate similar mechanisms into feedforward neural networks [15]. By enabling information to propagate more freely through the network, Highway Networks alleviate the challenges associated with vanishing gradients. Building upon the principles of Highway Networks, ResNet (Residual Network) simplifies the architecture by eliminating forget gates and directly employing simple skip connections [16]. This approach enables signals to bypass certain layers without the need for gating mechanisms, facilitating the training of exceedingly deep neural networks. The effectiveness of this structure has been empirically demonstrated across various domains. In recent years, numerous machine learning models based on the ResNet architecture have been extensively introduced, including ResNet-LSTM [20], ResNet-ARIMA [21], attention-ResNet-LSTM[36], Attention-ResNet-AR-LSTM[37], and ResNet-LightGBM [22]. However, there remains a paucity of ResNet-based models tailored specifically for studying energy consumption time series forecasting, particularly within the domain of U.S. fossil energy consumption, with virtually no examples to date.

In this study, we propose a novel approach termed Weighted ResNet with MLP, with the objective of enhancing feature extraction from time series data and improving forecast accuracy.

The subsequent sections of the paper will delineate the data collection process in Section 2, expound upon the theory and solution methodology of WResNet-MLP in Section 3, explore its application across three cases involving natural gas, petroleum, and total fossil fuel consumption in Section 4, and conclude with a summary in Section 5.

2 DATA COLLECTION

Utilizing consumption data of natural gas, petroleum, and total fossil fuels in the United States as the contextual foundation for constructing and analyzing time series forecasting models holds paramount importance. Firstly, energy consumption stands as a pivotal metric within a nation's economic and societal framework, with accurate forecasting serving as a cornerstone for governmental departments in formulating energy policies. Secondly, constructing and analyzing time series forecasting models facilitates the elucidation of inherent trends and cyclical variations within energy consumption, thereby fostering a deeper comprehension of the interplay between energy utilization and economic development. Furthermore, the application of forecasting models aids in assessing factors such as energy supply-demand equilibrium, resource efficiency, and environmental conservation, thus furnishing a scientific basis for sustainable energy development and environmental protection initiatives. Lastly, by applying advanced time series forecasting methodologies to actual energy consumption datasets, there exists the potential to propel the advancement and application of time series forecasting techniques, thereby fostering progress in related academic research endeavors. Consequently, employing the consumption metrics of natural gas, petroleum, and total fossil fuels in the United States as a backdrop for time series forecasting model construction and analysis bears profound significance in the realms of energy management, environmental conservation, and sustainable development.

The data collection originates from the U.S. Energy Information Administration (EIA), encompassing monthly records spanning from January 1973 to December 2023. These data are visually depicted in Fig. 1.

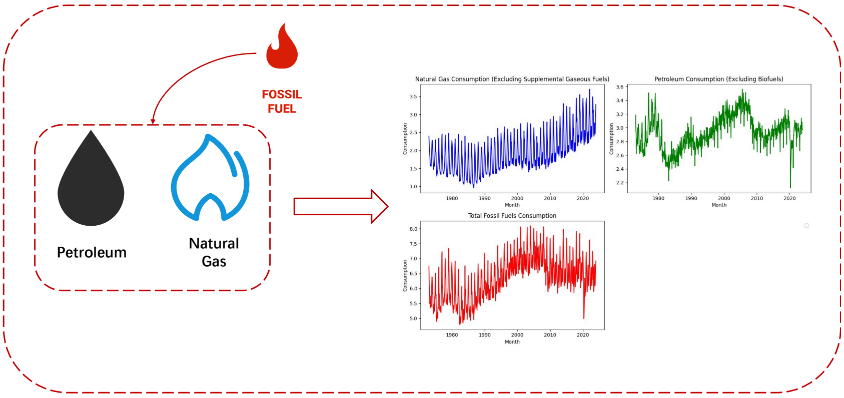

Fig. 1. Natural Gas, Petroleum, Total Fossil Fuels Consumption

3 RESEARCH METHODOLOGY

3.1 The Proposed Model and Its Mathematical Principles

At the heart of ResNet lies the skip connection, a structural feature that facilitates the transfer of information from earlier layers directly to subsequent layers. This mechanism enables the network to effectively capture key characteristics within the data, enhancing its ability to learn and generalize.

The Weighted ResNet with MLP (WResNet-MLP) model proposed in this study comprises a ResNet input layer and an output layer. Positioned between these layers are multiple blocks consisting of MLP units, the quantity of which is determined by the chosen depth, denoted as d, of the ResNet architecture. The schematic representation of the WResNet-MLP structure, with an input denoted as Z, is depicted in Fig. 2.

Let (Z, T) denote the input of the model, comprising n instances (Z_i, T_i) for $(i = 1, 2, \cdots, n)$. The output value s_0 of the input layer is computed as follows:

$$s_0 = W^{(1)}Z + \beta^{(1)}, \tag{3.1}$$

Subsequently, the output of the first block is determined as:

$$\begin{cases} s_1 = W^{(3)}\delta(W^{(2)}s_0 + \beta^{(2)}) + \beta^{(3)}, \\ s_2 = s_0 + W_1^{(4)}s_1, \end{cases} \tag{3.2}$$

where s_1 represents the output of the MLP, and s_2 denotes the output of the first block. The symbol $\delta(\cdot)$ represents the activation function, with various options available. In this paper, we adopt the sigmoid function as $\delta(\cdot)$ due to its favorable mathematical characteristics, defined as:

$$\sigma(\cdot) = \frac{1}{1 + e^{-\cdot}} \tag{3.3}$$

Likewise, the output of the second block is obtained as follows:

$$\begin{cases} s_3 = W^{(3)}\delta(W^{(2)}s_2 + \beta^{(2)}) + \beta^{(3)}, \\ s_4 = s_2 + W_2^{(4)}s_3, \end{cases} \tag{3.4}$$

Following the d layers of blocks, the output is represented as:

$$s_{2d} = s_{2d-2} + W_d^{(4)}s_{2d-1}, \tag{3.5}$$

Consequently, the output of the WResNet-MLP model is expressed as:

$$\hat{T} = W^{(5)}h_{2d} + \beta^{(5)}. \tag{3.6}$$

In the preceding equation, $W^{(k)}$ (for $k = 1, 2, 3, 4, 5$) and $\beta^{(k)}$ (for $k = 1, 2, 3, 5$) denote the parameters of the WResNet-MLP model.

3.2 Optimizing Model Training

Neural networks, including the WResNet-MLP model in this study, typically lack analytical solutions. Hence, optimization algorithms like Gradient Descent[23], Stochastic Gradient Descent[24], and Adam[25] are utilized. Here, we employ the Adam algorithm for its efficiency, robustness, and adaptability.

Firstly, we define the training error e_i at each point (Z_i, T_i) as follows:

$$e_i = T_i - W^{(5)}h_{2d} + \beta^{(5)}, \tag{3.7}$$

This yields the sum of training errors:

$$E(\boldsymbol{W}, \boldsymbol{\beta}) = \frac{1}{n}\sum_{i=1}^{n} e_i^2 = e^T e, \tag{3.8}$$

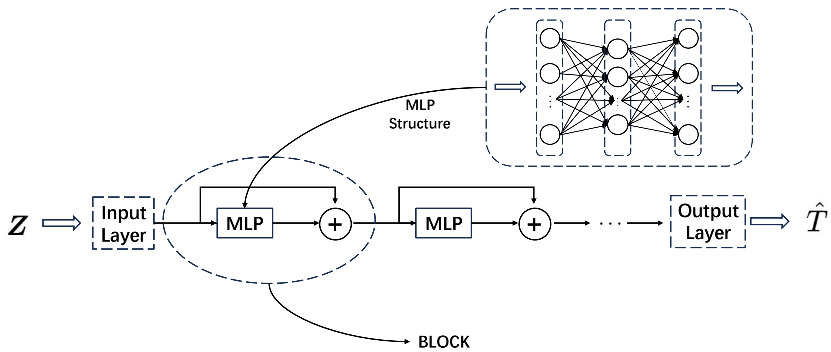

Fig. 2. The structure of WResNet-MLP with the input Z

where $\boldsymbol{W}$ comprises $W^{(k)}$ and $\boldsymbol{\beta}$ comprises $\beta^{(k)}$. These parameters $\boldsymbol{W}$ and $\boldsymbol{\beta}$ are to be solved using the Adam algorithm.

Next, to complete the Adam optimization process, we must compute the gradient of $E(\boldsymbol{W}, \boldsymbol{\beta})$:

$$\nabla E = [\frac{\partial E}{\partial \boldsymbol{W}}, \frac{\partial E}{\partial \boldsymbol{\beta}}], \tag{3.9}$$

where ∇E represents the gradient.

Unlike traditional gradient descent methods, Adam introduces the notion of modified bias-corrected first moment estimate $\hat{m}_i$ and bias-corrected second moment estimate $\hat{v}_i$ to accelerate convergence. Before computing these, we require the values of the first-order moment estimate m_i and second-order moment estimate v_i, which are expressed as follows:

$$m_i = \mu_1 \cdot m_{i-1} + (1 - \mu_1) \cdot \nabla E, \tag{3.10}$$

$$v_i = \mu_2 \cdot v_{i-1} + (1 - \mu_2) \cdot \nabla E^2, \tag{3.11}$$

where μ_1 and μ_2 denote decay rates controlling the speed of decay for the first and second moments of the gradients, respectively.
Then, the $\hat{m}_i$ and $\hat{v}_i$ could be obtained:

$$\hat{m}_i = \frac{m_i}{1 - \mu_1^i}, \tag{3.12}$$

$$\hat{v}_i = \frac{v_i}{1 - \mu_2^i}, \tag{3.13}$$

Finally, the iterative formula is derived as:

$$\begin{bmatrix} \boldsymbol{W}^{i+1} \\ \boldsymbol{\beta}^{i+1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{W}^{i} \\ \boldsymbol{\beta}^{i} \end{bmatrix} - r \cdot \frac{\hat{m}_i}{\sqrt{\hat{v}_i} + \epsilon}. \tag{3.14}$$

Here, r represents the learning rate of Adam, and ϵ is a small constant.

Algorithm 1: Adam Training for WResNet-MLP

Input: $E(\boldsymbol{W}, \boldsymbol{\beta})$ (Eq.(3.8)), Learning rate r, max_iterations
Output: $[\boldsymbol{W}, \boldsymbol{\beta}] \leftarrow random()$;
$\mu_1 \leftarrow 0.9$; $\mu_2 \leftarrow 0.999$;
$m_0 \leftarrow 0$; $v_0 \leftarrow 0$;
$iteration \leftarrow 0$;
while *iteration ¡ max_iterations* **do**
    iteration = iteration + 1 ;
    $\nabla E \leftarrow$ Eq.(3.9);
    $\widehat{m}_i, \widehat{v}_i \leftarrow$ Eq.(3.12)(3.13);
    $\begin{bmatrix} \boldsymbol{W}^{i+1} \\ \boldsymbol{\beta}^{i+1} \end{bmatrix} \leftarrow$ Eq.(3.14);
end
return $[\boldsymbol{W}, \boldsymbol{\beta}]$

Following Adam training, parameters such as depth (d), learning rate (r), and the number of neurons in the MLP still require tuning. This is accomplished using the Grid Search algorithm.

4 APPLICATIONS

In this section, we employ the proposed model to develop prediction models for three distinct datasets: natural gas consumption, oil consumption, and overall fossil fuel consumption. Each of the three datasets comprises 612 data points. The initial 80% of the data is allocated for training purposes, while the remaining 20% is reserved for testing. To enhance stability during model training, we preprocess the data using min-max scaling.

Furthermore, this study utilizes the Mean Absolute Percentage Error (MAPE) as a key metric to assess the model's performance. The represention of the MAPE is shown in Eq.(4.1). Additionally, we conduct comparative analyses with 10 machine learning models to validate the efficacy of our proposed approach, and their specific information is shown in Table 1.

$$\text{MAPE} = \frac{1}{s} \sum_{k \in U} \frac{|\hat{T}(k) - T(k)|}{|T(k)|} \tag{4.1}$$

where s is the length of U, and U is the training dataset or testing dataset.

Table 1. The specific information of comparsion machine learning models

Full Name	Abbreviation	Reference	Year
Gated Recurrent Unit	gru	[26]	2017
Random Forest Regression	rf	[27]	2001
Extreme Gradient Boosting	xgb	[28]	2015
Long Short-Term Memory	lstm	[29]	2000
Support Vector Regression	svr	[30]	1996
Convolution Neural Network	cnn	[31]	2015
Multilayer Perceptron	mlp	[32]	2009
CNN-LSTM	cnnlstm	[33]	2019
Convolutional LSTM	convlstm	[34]	2017
General Regression Neural Network	grnn	[35]	2004

4.1 Case 1:Natural Gas Consumption

Natural gas is pivotal in shaping the United States' energy landscape, influencing its energy supply, economic growth, and environmental stewardship. Accurately forecasting future natural gas consumption trends is instrumental in optimizing energy resource utilization and fostering sustainable national economic development.

The MAPE evaluation for training and testing across all models in Case-1 is presented in Table 2, while the prediction curves for both training and testing are depicted in Figs. 3 and 4, respectively.

Upon examination of Table2, it becomes evident that the proposed WResNet-MLP model exhibits the most favorable test MAPE value. While the rf model displays the best train MAPE value, its test MAPE value is among the least satisfactory. Although the cnn model's MAPE closely resembles that of the proposed model, its train MAPE is inferior to that of WResNet-MLP.

Furthermore, upon reviewing Figs. 3 and 4, it is apparent that the prediction curve of the proposed model closely aligns with the true curve in both training and testing phases. Conversely, the testing forecasting curves of models such as gru, rf, xgb, and grnn exhibit an overall upward shift when compared to the original curve.

4.2 Case 2:Petroleum Consumption

Petroleum is a pivotal energy resource in modern industrial society, extensively utilized in transportation, energy production, chemical engineering, and other sectors. Therefore, accurately forecasting future trends in petroleum consumption holds significant implications for national energy security, economic development, and environmental conservation.

Table 3 displays the MAPE evaluation results for training and testing across all models in Case-1. Additionally, Figs. 5 and 6 illustrate the prediction curves for both training and testing phases, respectively.

Upon examining Table 3, it's evident that the proposed WResNet-MLP model demonstrates the most promising test MAPE value. While the rf model boasts the best train MAPE value, its test MAPE value falls short in comparison. Upon reviewing Figs. 5 and 6, it becomes apparent that in scenarios with significant fluctuations between adjacent data points, the proposed model exhibits robust performance. Additionally, models such as gru, lstm, cnn, and convlstm also perform admirably under similar conditions.

Table 2. MAPE Evaluation for Training and Testing Across All Models: Case-1

Model	WResNet-MLP	gru	rf	xgb	lstm	svr	cnn	mlp	cnnlstm	convlstm	grnn
Train	10.572	7.412	**2.960**	4.051	5.600	8.699	10.699	8.375	5.017	6.862	5.023
Test	**8.447**	12.914	17.704	20.203	14.717	14.033	8.609	16.063	14.316	16.273	18.105

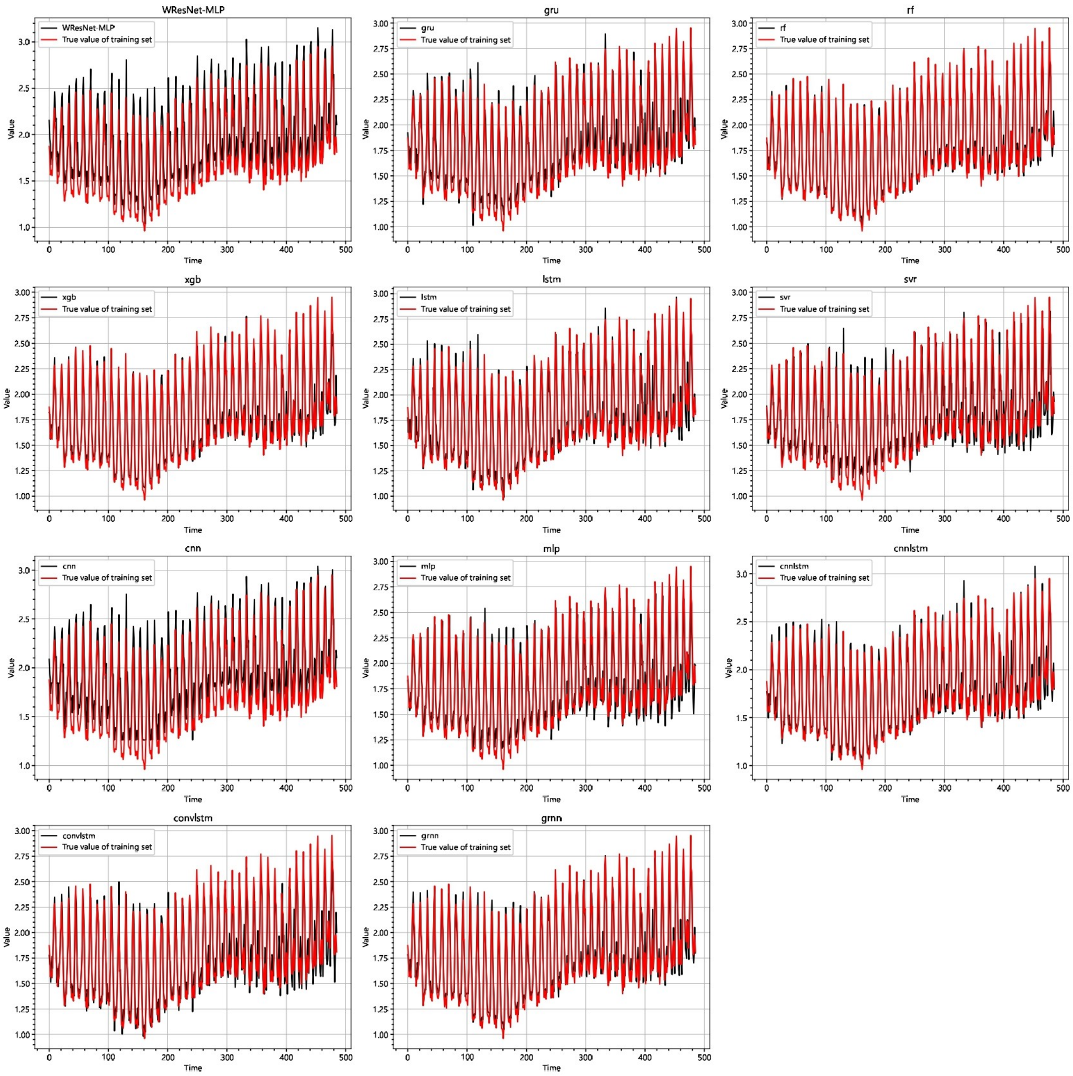

Fig. 3. Forecasted Natural Gas Consumption Values: Training Set

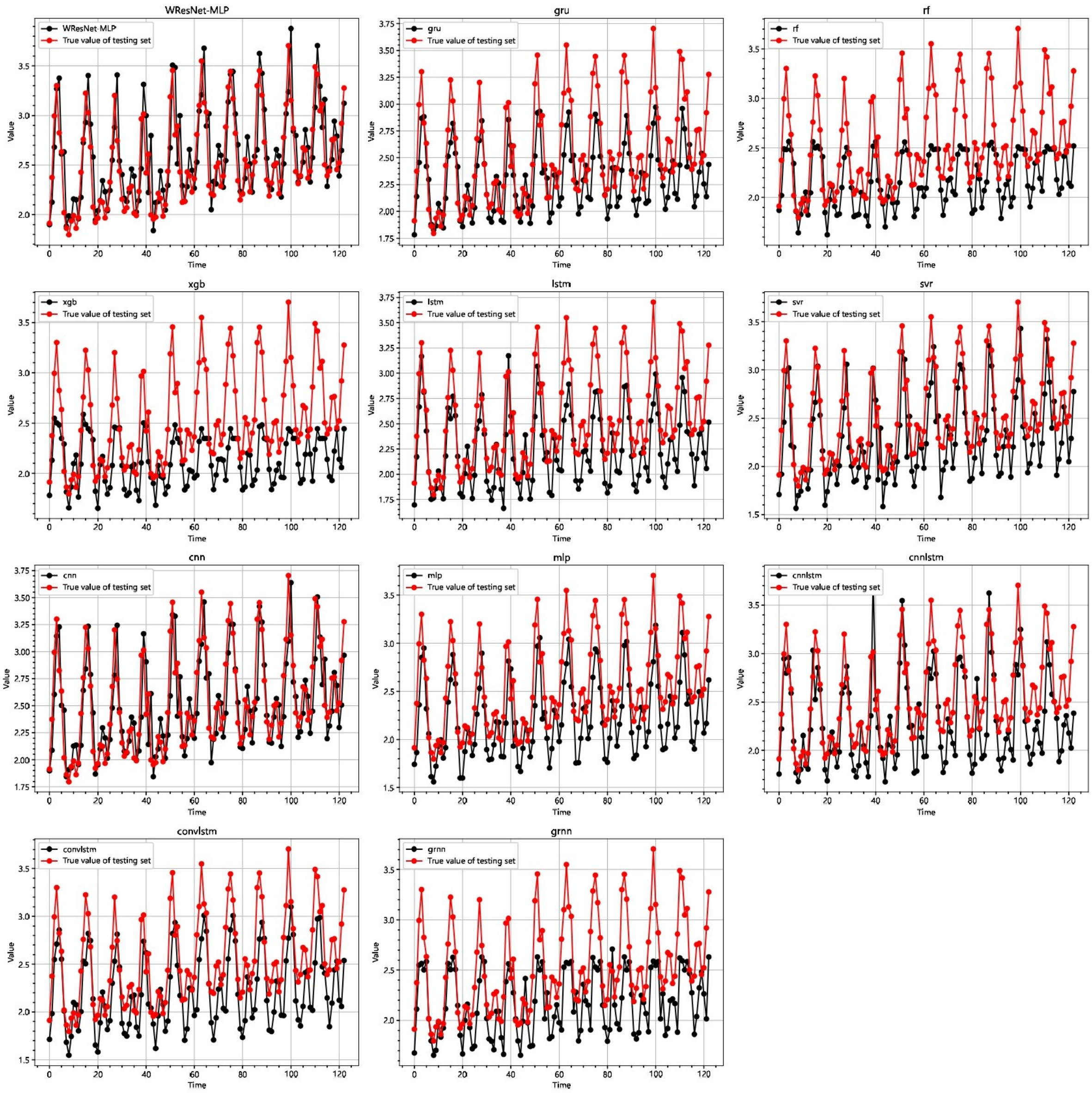

Fig. 4. Forecasted Natural Gas Consumption Values: Testing Set

Table 3. MAPE Evaluation for Training and Testing Across All Models: Case-2

Model	WResNet-MLP	gru	rf	xgb	lstm	svr	cnn	mlp	cnnlstm	convlstm	grnn
Train	3.638	3.624	**2.481**	3.091	3.615	3.623	3.499	3.610	3.141	3.559	3.069
Test	**3.460**	3.740	3.761	3.677	3.652	3.539	3.699	3.507	3.796	3.764	3.540

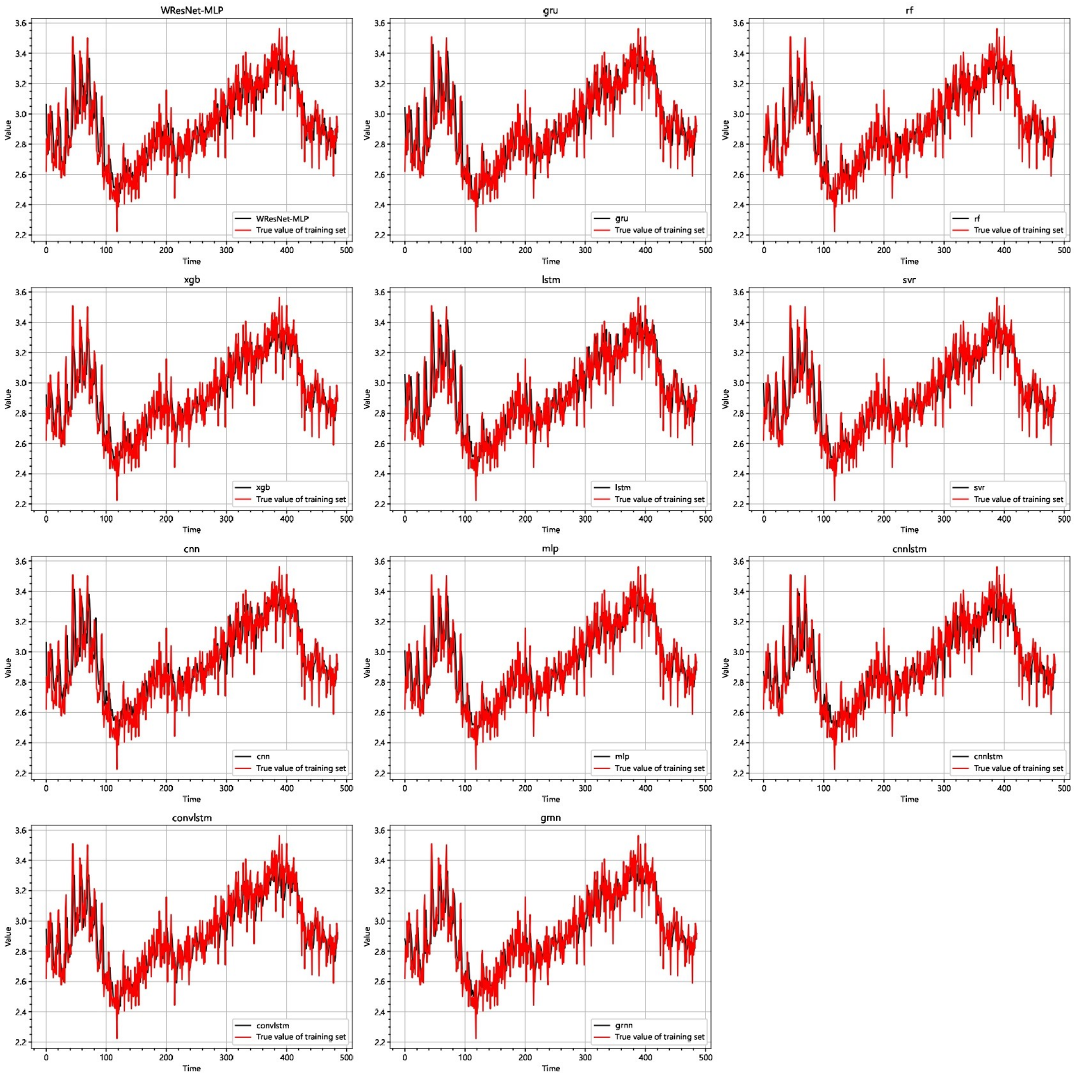

Fig. 5. Forecasted Petroleum Consumption Values: Training Set

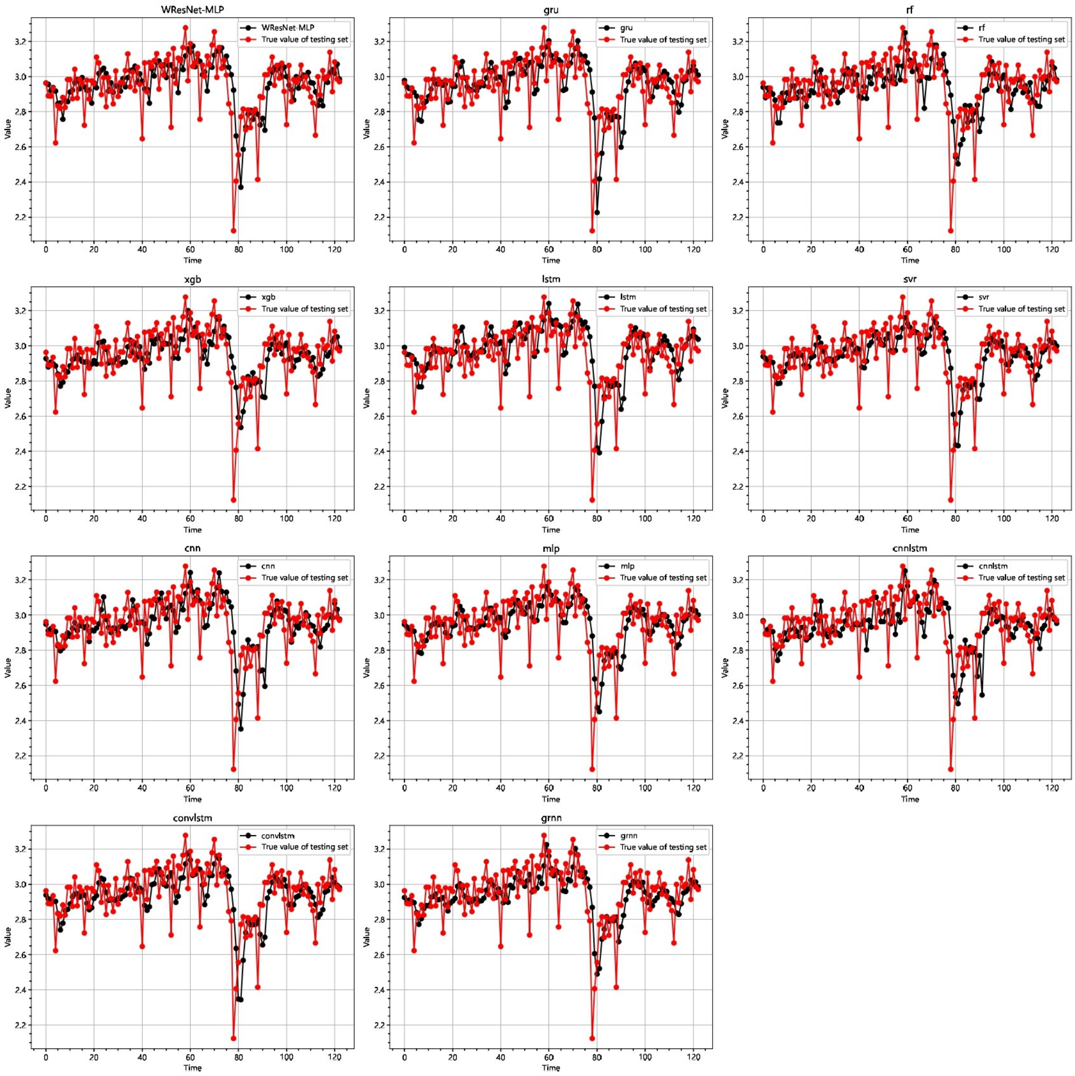

Fig. 6. Forecasted Natural Gas Consumption Values: Testing Set

4.3 Case 3:Total Fossil Fuels Consumption

Fossil fuels, as one of the primary energy sources, are extensively utilized in electricity generation, industrial production, transportation, and other sectors. Hence, accurately forecasting future trends in total fossil fuels consumption is crucial for national energy policy formulation, economic planning, and environmental conservation efforts.

The table below (see Table 4) presents the MAPE evaluation results for training and testing across all models in Case-1. Additionally, Figs. 7 and 8 depict the prediction curves for both training and testing phases, respectively. In line with the preceding cases, WResNet-MLP continues to exhibit the optimal test MAPE value, while RF demonstrates the superior train MAPE value. Furthermore, the proposed model's prediction curve aligns closely with the true curve in both the training and testing phases.

Table 4. MAPE Evaluation for Training and Testing Across All Models: Case-3

Model	WResNet-MLP	gru	rf	xgb	lstm	svr	cnn	mlp	cnnlstm	convlstm	grnn
Train	4.434	3.713	**2.136**	2.732	3.919	4.277	4.104	5.235	3.360	4.143	3.220
Test	**4.464**	4.816	4.900	4.848	4.822	4.621	4.466	5.730	4.801	4.959	4.742

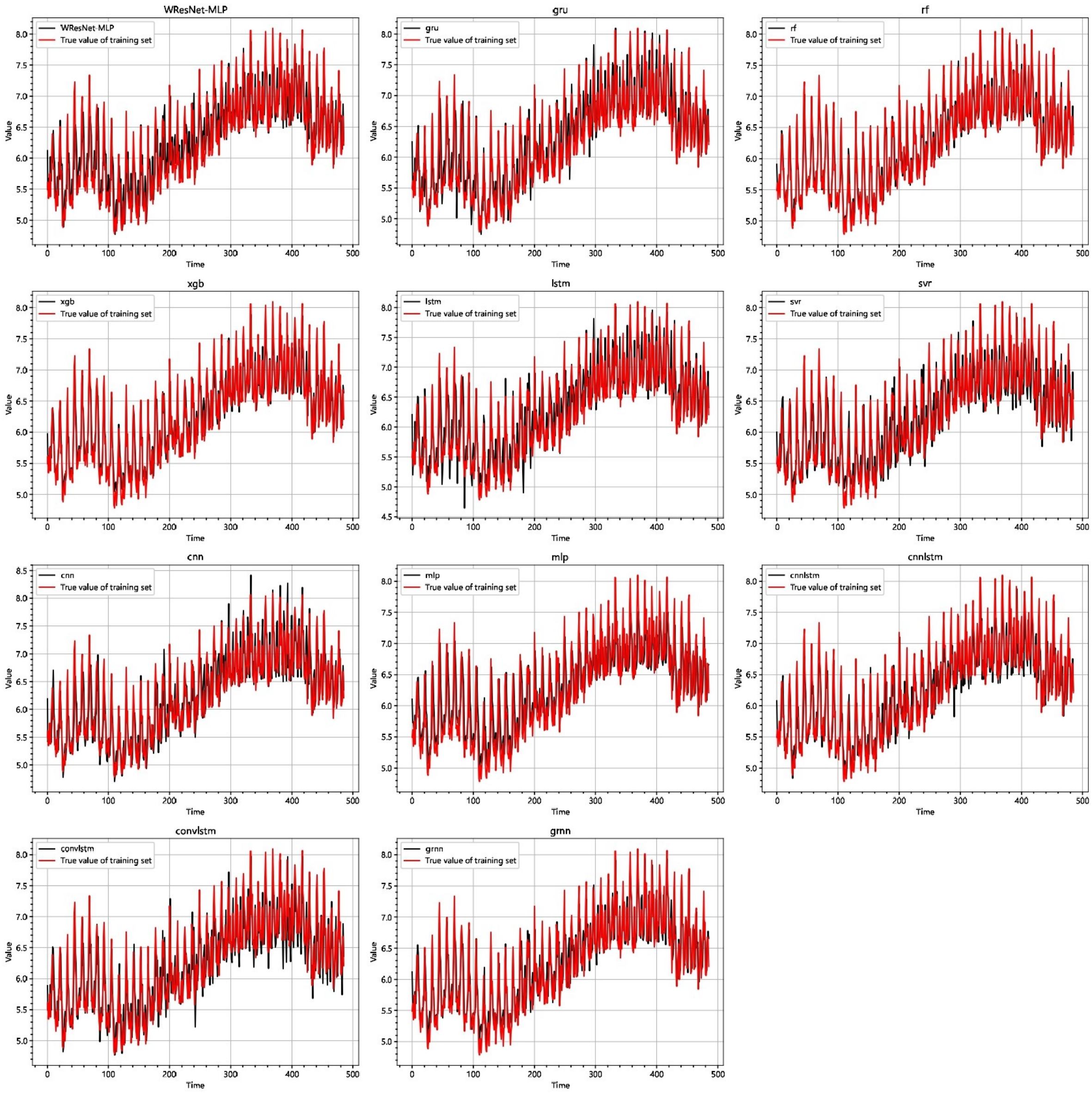

Fig. 7. Forecasted Total Fossil Fuels Consumption Values: Training Set

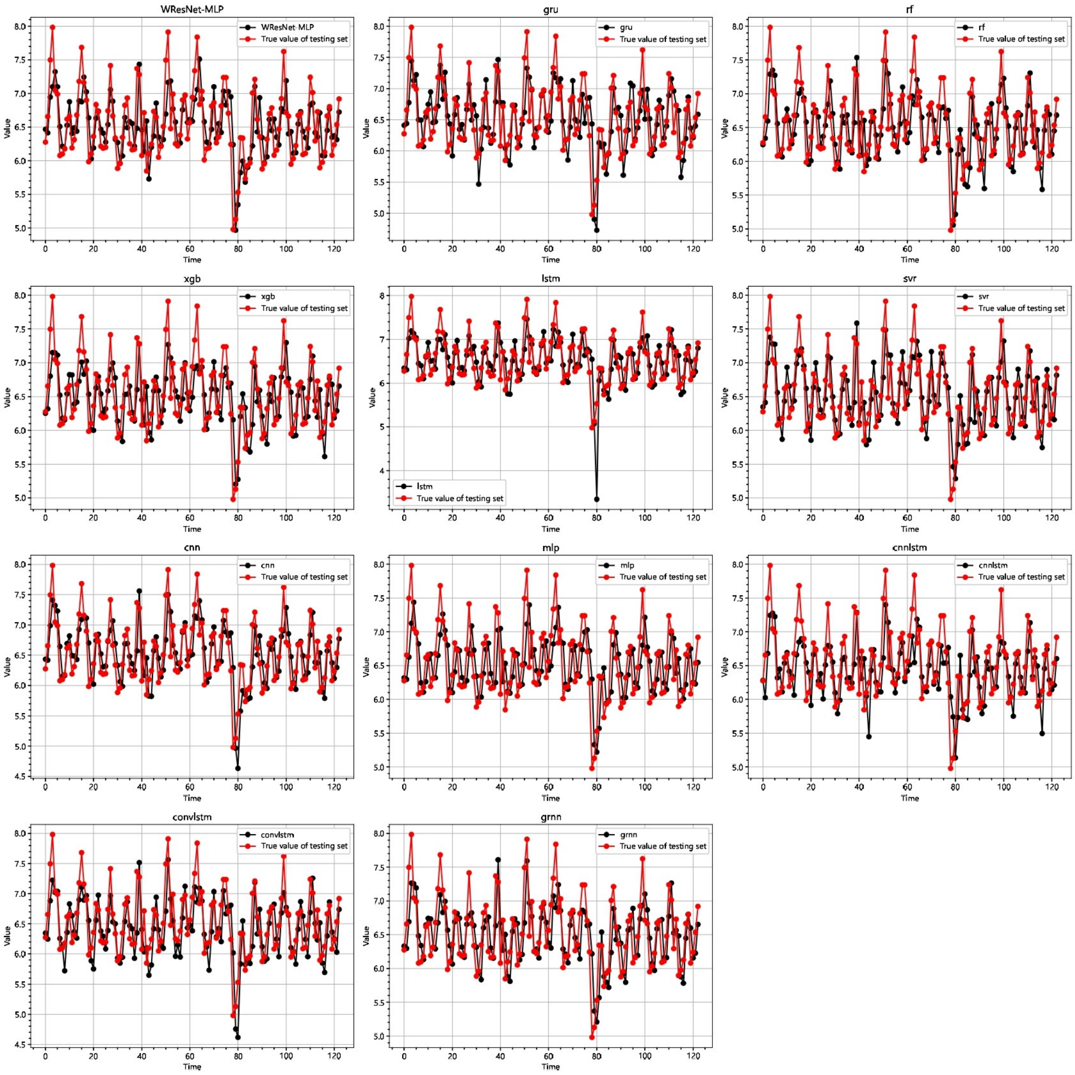

Fig. 8. Forecasted Total Fossil Fuels Consumption Values: Testing Set

4.4 Disscusion

The WResNet-MLP model, as proposed, demonstrates a notable performance divergence between its performance on the training set and its efficacy during testing. While its performance on the training set may not be particularly outstanding, across three distinct cases, it emerges as the top performer during testing, closely aligning with the raw data through its predictive curve. This notable performance in testing highlights the model's robust generalization capability and its accuracy in predictions, thereby demonstrating its reliability and stability when applied to real-world scenarios.

Conversely, the Random Forest (RF) model, despite achieving the highest Mean Absolute Percentage Error (MAPE) value during training, frequently displays suboptimal performance during testing phases. This discrepancy could potentially stem from either the model's insufficient complexity or its susceptibility to overfitting, wherein the model excessively fits to the

training data and struggles to generalize well to unseen data points.
mm]

5 CONCLUSIONS

This paper presents the WResNet-MLP model alongside its comprehensive theoretical framework and model training methodologies. Additionally, we applied this model to construct models for total natural gas, oil, and fossil fuel consumption in the United States. While the model's performance on the training set was moderate compared to 10 other machine learning models, it exhibited superior predictive performance on the test set.

The research demonstrates that the WResNet-MLP model holds promise as a dependable decision support tool for future energy forecasting. The integration of ResNet and MLP notably enhances prediction accuracy and applicability. We firmly believe that this approach holds significant potential for further exploration in future research endeavors.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

[1] Yadav H, Thakkar A. NOA-LSTM: An efficient LSTM cell architecture for time series forecasting[J]. Expert Systems with Applications, 2024;238:122333.

[2] Yi K, Zhang Q, Fan W, et al. Frequency-domain MLPs are more effective learners in time series forecasting[J]. Advances in Neural Information Processing Systems. 2024;36.

[3] Sajjad M, Khan Z A, Ullah A, et al. A novel CNN-GRU-based hybrid approach for short-term residential load forecasting[J]. Ieee Access, 2020;8:143759-143768.

[4] Rosenblatt F. The perceptron: A probabilistic model for information storage and organization in the brain[J]. Psychological Review. 1958;65(6):386.

[5] Rosenblatt F. Principles of neurodynamics: Perceptrons and the theory of brain mechanisms[M]. Washington, DC: Spartan books; 1962.

[6] Schmidhuber J. Annotated history of modern AI and Deep learning[J]. arXiv preprint ar. 2022;Xiv:2212.11279.

[7] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: Theory and applications[J]. Neurocomputing, 2006;70(1-3):489-501.

[8] Ivakhnenko AG, Lapa VG. Cybernetic predicting devices[J]. (No Title); 1965.

[9] Ivakhnenko AG, Lapa VG. Cybernetics and forecasting techniques[J]. (No Title); 1967.

[10] Amari S. A theory of adaptive pattern classifiers[J]. IEEE Transactions on Electronic Computers. 1967;(3):299-307.

[11] Rodríguez O H, Lopez Fernandez J M. A semiotic reflection on the didactics of the chain rule[J]. The Mathematics Enthusiast. 2010;7(2):321-332.

[12] Werbos PJ. Applications of advances in nonlinear sensitivity analysis[C]//System Modeling and Optimization: Proceedings of the 10th IFIP Conference New York City, USA, August 31–September 4, 1981. Berlin, Heidelberg: Springer Berlin Heidelberg. 2005;762-770.

[13] Rumelhart DE, Hinton GE, Williams RJ. Learning internal representations by error propagation[J]; 1985.

[14] Hochreiter S. Untersuchungen zu dynamischen neuronalen Netzen[J]. Diploma, Technische Universität München. 1991;91(1):31.

[15] Srivastava R K, Greff K, Schmidhuber J. Highway networks[J]. arXiv preprint ar. 2015;Xiv:1505.00387.

[16] He K, Zhang X, Ren S, et al. Identity mappings in deep residual networks[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14. Springer International Publishing. 2016;630-645.

[17] Azadeh A, Ghaderi SF, Sohrabkhani S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors[J]. Energy Conversion and management. 2008;49(8):2272-2278.

[18] Ekonomou L. Greek long-term energy consumption prediction using artificial neural networks[J]. Energy. 2010;35(2):512-517.

[19] Trejo-Perea M, Herrera-Ruiz G, Rios-Moreno J, et al. Greenhouse energy consumption prediction using neural networks models[J]. Training. 2009;1(1):2.

[20] Albelwi S. A Robust Energy Consumption Forecasting Model using ResNet-LSTM with Huber Loss[J]. International journal of computer science and network security: IJCSNS. 2022;22(7):301-307.

[21] Khanarsa P, Luangsodsai A, Sinapiromsaran K. Self-Identification ResNet-ARIMA Forecasting Model[J]. WSEAS transactions on systems and control. 2020;15(21): 196-211.

[22] Zhao Y, Khushi M. Wavelet Denoised-ResNet CNN and LightGBM method to predict forex rate of change[C]//2020 International Conference on Data Mining Workshops (ICDMW). IEEE. 2020;385-391.

[23] Andrychowicz M, Denil M, Gomez S, et al. Learning to learn by gradient descent by gradient descent[J]. Advances in neural information processing systems. 2016;29.

[24] Amari S. Backpropagation and stochastic gradient descent method[J]. Neurocom puting, 1993;5(4-5):185-196.

[25] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint. 2014ar;Xiv:1412.6980.

[26] Dey R, Salem F M. Gate-variants of gated recurrent unit (GRU) neural networks[C]//2017 IEEE 60th international midwest symposium on circuits and systems (MWSCAS). IEEE. 2017;1597-1600.

[27] Breiman L. Random forests[J]. Machine learning, 2001;45:5-32.

[28] Chen T, He T, Benesty M, et al. Xgboost: extreme gradient boosting[J]. R package version 0.4-2, 2015;1(4):1-4.

[29] Gers F A, Schmidhuber J, Cummins F. Learning to forget: Continual prediction with LSTM[J]. Neural computation, 2000;12(10):2451-2471.

[30] Drucker H, Burges CJ, Kaufman L, et al. Support vector regression machines[J]. Advances in neural information processing systems. 1996;9.

[31] O'Shea K, Nash R. An introduction to convolutional neural networks[J]. arXiv preprint 2015;arXiv:1511.08458.

[32] Popescu M C, Balas V E, Perescu-Popescu L, et al. Multilayer perceptron and neural networks[J]. WSEAS Transactions on Circuits and Systems. 2009;8(7):579-588.

[33] Kim TY, Cho SB. Predicting residential energy consumption using CNN-LSTM neural networks[J]. Energy. 2019;182:72-81.

[34] Kim S, Hong S, Joh M, et al. Deeprain: Convlstm network for precipitation prediction using multichannel radar data[J]. arXiv preprint. 2017;arXiv:1711.02316.

[35] Kulkarni SG, Chaudhary AK, Nandi S, et al. Modeling and monitoring of batch processes using principal component analysis (PCA) assisted generalized regression neural networks (GRNN)[J]. Biochemical Engineering Journal. 2004;18(3):193-210.

[36] Yu S, Zhang Z, Wang S, et al. A performance-based hybrid deep learning model for predicting TBM advance rate using attention-ResNet-LSTM[J]. Journal of Rock Mechanics and Geotechnical Engineering. 2024;16(1):65-80.

[37] Zhang Y, Ma M, Li Y, et al. Attention-ResNet-AR-LSTM: An intelligent method for PCCP deformation prediction via structure monitoring based on distributed fiber optics[J]. Engineering Failure Analysis. 2024;108157.

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle5.com/review-history/117020