Knowledge Engineering and Data Science (KEDS)
Vol 7, No 1, April 2024, pp. 1–12

pISSN 2597-4602
eISSN 2597-4637

Exploring LSTM-based Attention Mechanisms with PSO and Grid Search under Different Normalization Techniques for Energy demands Time Series Forecasting

Andri Pranolo [a,1,*], Xiaofeng Zhou [b,2], Yingchi Mao [b,3], Bambang Widi Pratolo [c,4], Aji Prasetya Wibawa [d,5], Agung Bella Putra Utama [d,6], Abdoul Fatakhou Ba [b,7], Abdullahi Uwaisu Muhammad [e,8]

[a] *Informatics Department, Faculty of Industrial Technology, Universitas Ahmad Dahlan*
Jl. Ringroad Selatan, Kragilan, Tamanan, Bantul 55191, Indonesia
[b] *College of Computer and Information, Hohai University*
Focheng West Road, Jiangning, Nanjing 211100, China
[c] *English Department, Faculty of Teacher Training and Education, Universitas Ahmad Dahlan*
Jl. Ringroad Selatan, Kragilan, Tamanan, Bantul 55191, Indonesia
[d] *Department of Electrical Engineering and Informatics, Faculty of Engineering, Universitas Negeri Malang*
Jl. Semarang no. 5, Malang 65145, Indonesia
[e] *Computer Science Department, Faculty Of Computing, Federal University Dutse*
Dutse, Jigawa 720101, Nigeria
[1] *andri.pranolo@tif.uad.ac.id**; [2] *zhouxf@hhu.edu.cn*; [3] *yingchimao@hhu.edu.cn*; [4] *bambang.pratolo@pbi.uad.ac.id*; [5] *aji.prasetya.ft@um.ac.id*; [6] *agungbpu02@gmail.com*; [7] *bafatakhou@gmail.com*; [8] *uwaisabdullahi87@yahoo.com*
* *corresponding author*

ARTICLE INFO

Article history:
Received 11 February 2024
Revised 03 April 2024
Accepted 08 April 2024
Published online 16 April 2024

Keywords:
Attention Mechanism
LSTM
PSO
Energy demands
Grid search

ABSTRACT

Advanced analytical approaches are required to accurately forecast the energy sector's rising complexity and volume of time series data. This research aims to forecast the energy demand utilizing sophisticated Long Short-Term Memory (LSTM) configurations with Attention mechanisms (Att), Grid search, and Particle Swarm Optimization (PSO). In addition, the study also examines the influence of Min-Max and Z-Score normalization approaches in the preprocessing stage on the accuracy performances of the baselines and the proposed models. PSO and Grid Search techniques are used to select the best hyperparameters for LSTM models, while the attention mechanism selects the important input for the LSTM. The research compares the performance of baselines (LSTM, Grid-search-LSTM, and PSO-LSTM) and proposes models (Att-LSTM, Att-Grid-search-LSTM, and Att-PSO-LSTM) based on MAPE, RMSE, and R2 metrics into two scenarios normalization: Min-Max, and Z-Score. The results show that all models with Min-Max normalization have better MAPE, RMSE, and R2 than those with Z-Score. The best model performance is shown in Att-PSO-LSTM MAPE 3.1135, RMSE 0.0551, and R2 0.9233, followed by Att-Grid-search-LSTM, Att-LSTM, PSO-LSTM, Grid-search-LSTM, and LSTM. These findings emphasize the effectiveness of attention mechanisms in improving model predictions and the influence of normalization methods on model performance. This study's novel approach provides valuable insights into time series forecasting in energy demands.

I. Introduction

Energy, especially electricity, is essential in developing and improving a country's welfare [1]. Studies show the important role of electricity consumption in stimulating economic growth and supporting the transition to more environmentally friendly production practices with lower carbon emissions. Estimating energy demand is crucial to the capacity and transmission system planning process, energy generation strategies, and determining pricing strategies [2]. In addition, despite continuous progress in technological development and global population growth, the detrimental impact of using renewable energy sources on the climate is becoming increasingly clear and cannot be ignored [3][4]. This problem demands serious attention and immediate action to reduce dependence

https://doi.org/10.17977/um018v7i12024p1-12

on energy sources that have the potential to damage the environment and accelerate the transition to cleaner and more sustainable energy sources. This process is important to ensure adequate energy supply and optimize operational and financial efficiency in energy resource management. Energy demand predictions vary by period, with capacity planning requirements focusing on long-term estimates [5] based on economic or demographic factors [2]. Daily habits and seasonal influences mainly cause short-term fluctuations, while extreme weather conditions [6] and special events [7], such as holidays or sporting events, can result in unexpected changes in trends. Long-range planning requires accurate and consistent estimates of hourly demand, which is essential in an evolving environment for effective and reliable resource allocation. To solve this problem, finding an accurate and consistent assessment of energy demands is important. Stating energy demand with precision can enable more accurate planning strategies and efficient distribution of energy to various sectors in need. This process helps optimize the accuracy of energy demand forecasting and greatly influences strategic and operational decision-making in the energy sector.

Intelligent algorithms based on Artificial Intelligence (AI) play a vital role in forecasting energy demand with high accuracy [8][9][10][11], which is crucial for effectively managing electrical energy consumption [12], generation, pricing and adapting to weather conditions to suit user needs [13]. Through the application of machine learning techniques, historical data is studied to form predictive models that optimize not only in projecting electrical energy needs but also in understanding the influence of variables such as weather on energy consumption and generation. The latest research explores in-depth simultaneous predictions of electricity demand, power generation composition and carbon emissions integrating the SDs model and PGMP model [5], and LSTM [14] was used to optimize the benefits obtained during the investment life cycle while reducing the period required for investors in energy storage to recover the capital spent. This approach is critical in driving wider adoption of energy storage technologies, strengthening the financial well-being of projects, and supporting the transition to cleaner, more efficient energy systems.

The study of multivariate time series data has traditionally been dependent on conventional statistical forecasting methods. As the size of datasets increases, accompanied by an increase in the number of variables and intricate non-linear relationships, the efficacy of conventional approaches diminishes. Deep learning techniques, specifically Recurrent Neural Networks (RNNs), present a promising prospect for achieving enhanced accuracy in the prediction of multivariate time series [15]. Long Short-Term Memory (LSTM) networks have been observed to exhibit superior performance in capturing long-term dependencies in sequential data compared to other RNN architectures [16]. Nevertheless, LSTM models have difficulties in effectively capturing the interdependencies present within a given dataset. The presence of this constraint is a hindrance to the accuracy of predicting.

A hybrid model combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) has been developed to make accurate predictions regarding sales of New Energy Vehicles (NEVs) as well as efficiently manage infrastructure ownership [17]. When compared with traditional models or other hybrid models, this CNN-LSTM combination shows superiority in various evaluation metrics. This approach stands out for its ability to effectively process spatial and temporal data with CNN processing spatial features and LSTM handling time dependencies, resulting in more accurate and efficient predictions. Further, the method combines Graph Convolutional Network (GCN) and Long Short-Term Memory (LSTM) that advantages of both networks to extract spatial and temporal information from extensive datasets, resulting in more precise short-term power load estimates compared to earlier methods that utilize CNN-LSTM [18].

These studies show that hybrid LSTM with other models can improve the accuracy of LSTM models. Attention processes are being emphasized as a means to address this fundamental issue [19]. By employing a strategic approach to selecting and prioritizing pertinent input data, these strategies enhance the model's ability to discern and evaluate various elements [20]. Attention mechanisms are utilized by models like LSTM to effectively capture intricate connections and interdependencies among several variables [21]. This is achieved by assigning diverse levels of importance or priority to different segments of the input sequence. Attention mechanisms play a crucial role in enhancing the ability of LSTM networks to comprehend intricate interdependencies within multivariate time series data. The enhancement is of utmost importance in predicting model correctness and durability when dealing with intricate and non-linear variable relationships. Therefore, the integration of this combination exhibits significant promise in enhancing the accuracy of multivariate time series forecasting by the incorporation of intricate and thorough modelling of data dynamics.

This research intends to analyze electrical energy consumption, generation, pricing, and the impact of weather. It proposes a predictive model based on data from relevant literature to enhance the accuracy of estimating these requirements. The primary advancements of this investigation, which set it apart from prior investigations, can be succinctly summarized as follows:

(1) To examine the impact of min-max and z-score normalization methods on the accuracy of predictions. This study examines the impact of two normalization methods on the prediction performance of the baseline and proposed models.
(2) To discover the optimal configuration that minimizes prediction error by introducing an LSTM model that incorporates attention mechanism, particle swarm optimization, and Grid search. The evaluation of this model is conducted using the Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and coefficient of determination (R^2) metrics.

This paper presents a fresh viewpoint on the modelling and forecasting of electrical energy demand, showcasing the potential of advanced methods in enhancing the comprehension and control of the fluctuations in energy use, generation, pricing, and weather effects. This research endeavours to employ cutting-edge artificial intelligence and data analysis approaches to offer superior and streamlined solutions for addressing present and future energy concerns.

II. Methods

The methodology of research to analyze the hourly energy demand time series dataset based on Attention, Particle Swarm Optimization (PSO), and Long Short-Term Memory (LSTM) is shown in Figure 1. Initially, the method entails normalizing the dataset through the utilization of two distinct statistical techniques: Min-Max normalization and Z-Score normalization. Min-Max normalization rescales the features in the dataset to a specific range, typically [0, 1] or [-1, 1], while Z-Score normalization standardizes the features by adjusting them according to the mean and standard deviation of the dataset. After normalizing, the datasets are divided into separate training and testing sets. This means that the model will be trained on one part of the data and evaluated on another to determine its ability to make accurate predictions.

The hyperparameter search space is established, encompassing the batch size, epochs, number of hidden layers, loss function, activation function, neurons, and optimizer. The hyperparameter tuning involves applying two separate processes: PSO and Grid Search. PSO is a computational technique that iteratively enhances a potential solution to a problem based on a specific measure of excellence. In contrast, Grid Search methodically explores a predetermined subset of the hyperparameter space. The experimental arrangement and configurations are compared. Baseline models, such as LSTM, are contrasted with models that have been optimized using Grid Search (Grid-search-LSTM) and PSO-LSTM. An attention mechanism (Att) is added to the baseline as an Att-LSTM, Att-Grid-search-LSTM, and Att-PSO-LSTM to aid the model in directing its attention toward select segments of the input sequence that hold greater relevance for the prediction task to increase the accuracy. This implies that the performance of the LSTMs and configured LSTM's baselines will be evaluated against LSTMs enhanced with the attention mechanism, attention mechanism when combined with the hyperparameter tuning outcomes from Grid Search and PSO (Att-Grid-search-LSTM and Att-PSO-LSTM). Lastly, the models are evaluated using three metrics: Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Coefficient of Determination (R^2).

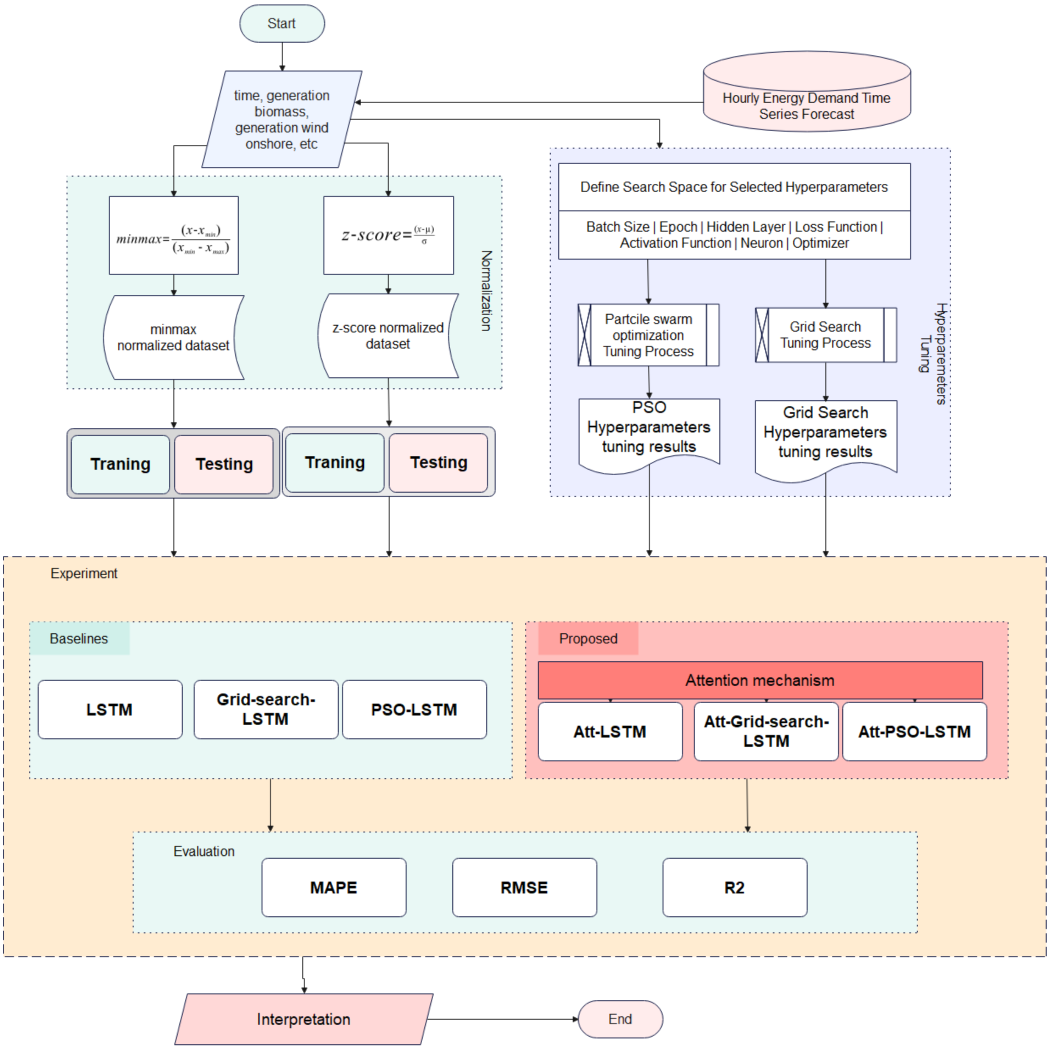


Fig. 1. Research methodology

A. Dataset

The dataset was obtained from the Kaggle.com website in the form of a multivariate time series titled 'Hourly Energy Demand Time Series Forecast', provided in CSV format [22]. This dataset encompasses hourly electricity consumption data from January 2015 to December 2018. It comprises 35,064 instances and 29 attributes. Within this dataset, the target attribute is the actual total load. However, the attribute for total load forecast is excluded from the analysis as it serves as a benchmark or comparison attribute to the target attribute. Consequently, the study focuses on 28 selected attributes.

B. Preprocessing

In pursuing rigorous analysis, this research significantly emphasizes preprocessing techniques to refine the raw data for subsequent testing [23]. Addressing missing values and ensuring uniformity through data normalization is paramount in this process. Missing data, a common occurrence in datasets, poses potential challenges to model performance. Thus, employing the deletion technique, which involves the removal of rows or columns containing missing values, becomes imperative [24]. By mitigating the presence of missing values, the dataset becomes more robust and conducive to accurate analysis.

Data normalization emerges as a crucial step in ensuring consistency and comparability across different features within the dataset. The normalization process entails transforming the values within the data to a standardized range. The selection of an appropriate normalization method significantly influences the accuracy and reliability of subsequent analyses. This study implements two widely used normalization methods: min-max as in (1) and z-score normalization as in (2). x represents the tested data, x_{min} represents the minimum attribute value of the tested data, x_{max} represents the maximum attribute value of the tested data. μ represents the mean value of the data, σ represents the standard deviation. Each method offers distinct advantages in adjusting the data distribution, contributing to more reliable analytical outcomes.

$$minmax = \frac{(x - x_{min})}{(x_{max} - x_{min})} \quad (1)$$

$$z - score = \frac{x - \mu}{\sigma} \quad (2)$$

In the context of this research, min-max normalization scales the data to a range between 0 and 1, effectively preserving the relative relationships between data points while ensuring uniformity [25]. Z-score normalization transforms the data distribution to have a mean of 0 and a standard deviation of 1, thereby standardizing the data within a range of -1 to 1 [26]. This normalization technique facilitates the identification of outliers and provides a clearer understanding of the data distribution. By meticulously implementing these preprocessing steps, the research aims to enhance the quality and reliability of subsequent analyses, yielding more robust findings and insights. The example normalization results of min-max and z-score are presented in Table 1.

Table 1. Example of normalization results

Actual Values	Min-max Normalization Value	Z-score Normalization
25385	0.319666	-0.723934
24382	0.276008	-0.943172
…	…	…
29735	0.509010	0.226902
28071	0.436580	-0.136820

C. *Forecasting Process*

In forecasting processes utilizing Long Short-Term Memory (LSTM) networks, the intricacies lie within the configuration of its parameters to optimize performance. LSTM, a type of Recurrent Neural Network (RNN), is renowned for effectively modelling sequential data by retaining information over extended periods [27]. However, determining the optimal parameters for LSTM poses a significant challenge due to the interplay between its various hyperparameters. Issues arise concerning selecting appropriate parameter settings, significantly impacting the model's predictive capabilities. To address this challenge, researchers turn to hyperparameter tuning techniques as a solution.

Hyperparameter tuning becomes imperative to fine-tune the LSTM model's parameters and enhance its forecasting accuracy. This study employs two prominent hyperparameter tuning methodologies: grid search and Particle Swarm Optimization (PSO). The grid search methodology is favoured for systematically exploring all conceivable parameter combinations [28]. By exhaustively evaluating different hyperparameter configurations, grid search facilitates comprehensive parameter optimization, addressing the challenge of selecting the most suitable parameter settings. Pseudocode 1 is the pseudocode for hyperparameter tuning using the grid search.

PSEUDOCODE 1. Hyperparameter tuning using grid search

```
Function grid_search_LSTM(parameters_range):
    best_params = None
    best_score = +infinity
    for each parameter combination in parameters_range:
        model = create_LSTM_model(parameter_combination)
        score = evaluate_model(model)
        if score < best_score:
            best_score = score
            best_params = parameter_combination
    return best_params
```

Inspired by nature's social behaviour, PSO iteratively updates candidate solutions based on each solution's historical performance and local and global influences. Researchers can navigate complex parameter spaces more efficiently by leveraging PSO's capacity for global exploration and exploitation [21]. This enables the identification of optimal hyperparameter configurations that may not be apparent through conventional grid search alone. Through the synergistic combination of grid search and PSO, researchers endeavour to unlock the full potential of LSTM in forecasting tasks, thereby advancing the field of predictive modelling. The pseudocode for hyperparameter tuning using PSO is in Pseudocode 2.

PSEUDOCODE 2. Hyperparameter tuning using PSO

```
Function PSO_LSTM(parameters_range):
    Initialize particles with random parameter values within parameters_range
    Initialize velocity for each particle
    Set global best_score = +infinity
    Set global best_params = None
    for each iteration:
        for each particle:
           Update velocity based on particle's current position and best
position
            Update particle's position using velocity
            Evaluate particle's position
            Update particle's best position if position is better than previous
best
            Update global best position if particle's best position is better
than global best
    return global best_params
```

The tuned LSTM parameters and the results of the hyperparameter tuning are shown in Table 2. It highlights the distinctions between Grid Search and PSO when choosing the hyperparameter space. Grid Search utilizes a methodical approach by systematically evaluating every combination within the specified grid in order to identify the optimal answer. On the other hand, PSO utilizes a dynamic approach by using a group of solutions that continuously adapt according to their performance. These hyperparameters include the batch size, the number of epochs, the number of hidden layers, the choice of the loss function, the activation function, the number of neurons per layer, and the optimizer.

Table 2. Hyperparameter tuning results

Parameter	Search Space	Grid Search	PSO
Batch Size	‘100’, ‘1000’	100	100
Epoch	’50’, ‘100’	50	100
Hidden Layer	‘2’, ‘5’, ‘10’	2	5
Loss Function	‘MSE’, ‘MAE’	MSE	MSE
Activation Function	‘Tanh’, ‘Sigmoid’	Tanh	Tanh
Neuron	‘32’, ‘64’	32	64
Optimizer	‘Adam’, ‘RMSprop’	RMSprop	RMSprop

Both Grid Search and PSO concur on a batch size of 100, which denotes the number of training instances that are processed prior to updating the model's parameters. However, they differ in their selection of epochs, with Grid Search choosing 50 and PSO optimal with a longer training period with 100 epochs. This implies that PSO expects to get advantages from a training cycle that is extended in duration. The intricacy of the network, as indicated by the number of hidden layers, Grid Search chooses a model with two layers that is relatively simpler, whereas PSO prefers a more intricate structure consisting of 5 layers. This could indicate that PSO can traverse through complex models, potentially leading to improved performance. However, both approaches lay the best parameters with the Mean Squared Error (MSE) loss function and the Tanh activation function. The number of neurons per layer is another factor that distinguishes the two systems. Grid Search opts for a conservative 32 neurons, while PSO favours a more densely populated layer with 64 neurons. This differentiation could be crucial in ascertaining the model's capacity to comprehend and represent the fundamental patterns of the data. Ultimately, both approaches unanimously choose RMSprop as the optimizer due to its effectiveness in managing noisy gradients and its capacity to adapt to various situations, highlighting its appropriateness for the given task.

D. LSTM-Based Attention Mechanism, PSO, and Grid Search

Figure 2 develops the proposed model: Attention LSTM, Att-PSO-LSTM, and Att-Grid-search-LSTM. In the illustration, the hyperparameter tuning process is not only carried out by PSO but also by Grid search. After the hyperparameter tuning process is complete, the optimized hyperparameter settings are then applied to the LSTM. The next step is integrating the attention mechanism into LSTM, which plays an important role in improving the capabilities and effectiveness of the LSTM model. This attention mechanism helps the model to focus on more relevant information during the learning process, thereby enabling the model to achieve better and more accurate performance in data processing.

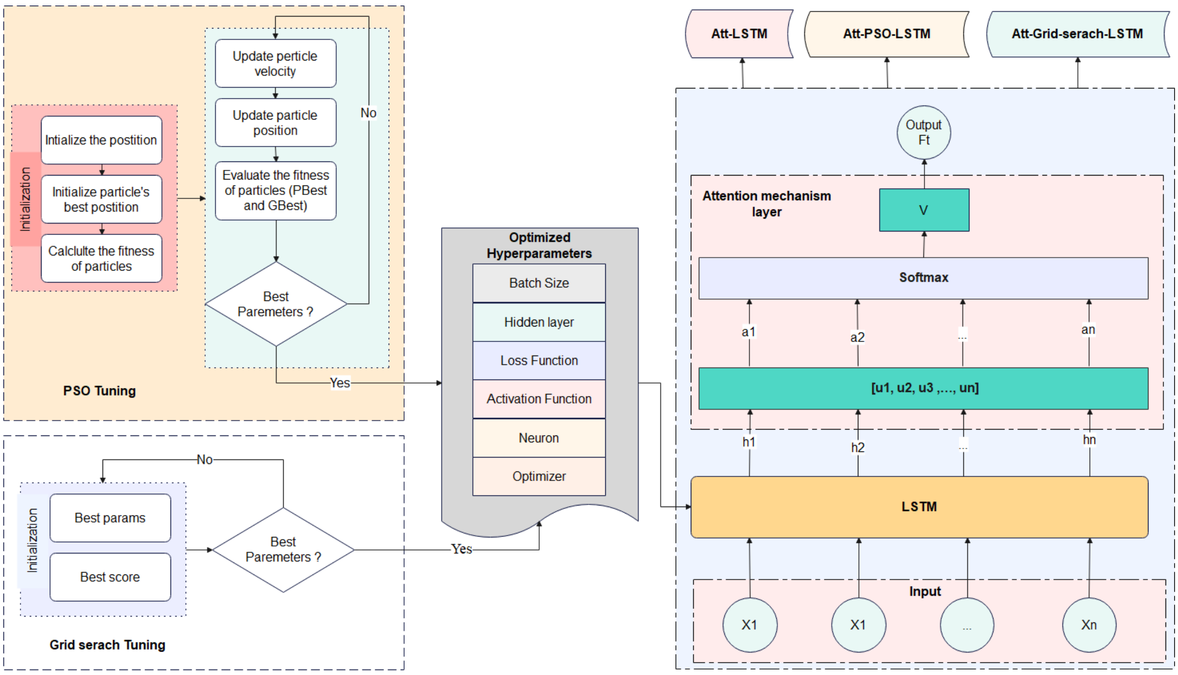


Fig. 2. LSTM Based Attention Mechanism, PSO, and Grid Search

The attention mechanism is implemented through a series of crucial phases [29]: At first, the LSTM outputs, represented as $[h_1, h_2, h_3, \ldots, h_n]$, are subjected to a nonlinear transformation resulting in $[u_1, u_2, u_3, \ldots, u_n]$. During shield tunneling operations, specific characteristics have a significant impact on the shield's direction and position, making it necessary to prioritize these parameters. In an attention weight matrix $[a_1, a_2, a_3, \ldots, a_n]$, a indicates the importance of each intermediate stage. Afterwards, the mechanism performs a weighted aggregation of the input parameters and their related weights to produce the encoding vector V. The encoding vector V is used to obtain the final output, ft, through a decoding process. The precise formula in question is as in (3) to (5).

$$u_k = \tanh(W_k h_k + b_k) \quad (3)$$

$$a_k = \frac{\exp\,(u_k^T u^s)}{\sum_{k=1}^{n} \exp\,(u_k^T u^s)} \quad (4)$$

$$V = \sum_k^n a_k h_k \quad (5)$$

Within this framework, W_k represents the weight matrices, b_k represents the bias term, a_k represents the normalized attention weights and u^s represents the randomly initialized time series attention matrix.

E. Evaluation

The evaluation stage constitutes a pivotal phase wherein the model's performance is assessed through error calculations to gauge its effectiveness. In this research, the evaluation employs various metrics, including the Mean Absolute Percentage Error (MAPE) as in (6), Root Mean Square Error (RMSE) as in (7), and R^2 as in (8). A_i is the actual value, F_i is the predicted value, n is the number of predictions, SS_{res} is the residual sum of squares and SS_{tot} is the total sum of squares.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \frac{|A_i - F_i|}{A_i} \tag{6}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(F_i - A_i)^2} \tag{7}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{8}$$

MAPE is a pivotal indicator of the model's accuracy, providing insights into its predictive capabilities [30]. RMSE assumes significance, particularly in outlier detection, shedding light on the model's susceptibility to extreme data points [31]. Meanwhile, R^2 facilitates correlations between the original and predicted data, elucidating the model's ability to capture underlying relationships [32].

III. Result and Discussion

In this section, we discuss the potential of LSTM models in predicting sequential data, focusing on Hourly Energy Demand Time Series Forecast data. This research explores various LSTM model configurations and tests two data normalization techniques, namely Min-Max and Z-Score, to identify a more optimal normalization approach. Furthermore, this research investigates the effect of parameter optimization techniques, such as Grid search and PSO, and the use of attention mechanisms in increasing the accuracy of model predictions. Specifically, this study focuses on three attention-based models, Attention LSTM, Attention Grid-search-LSTM, and Attention-PSO-LSTM, to compare their performance with three baseline models: LSTM, Grid-search-LSTM, and PSO-LSTM. Figure 3 and Figure 4 depict a graph of visualization for time series forecasting with actual data as a time series of real values, and a model generates the prediction line.

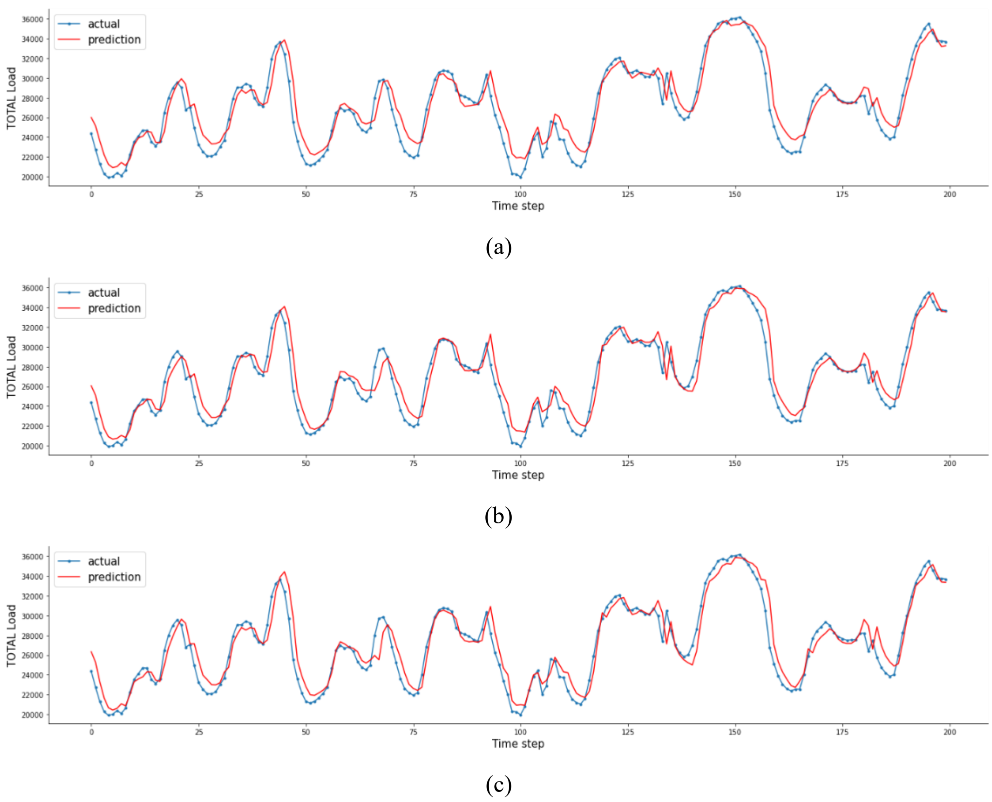
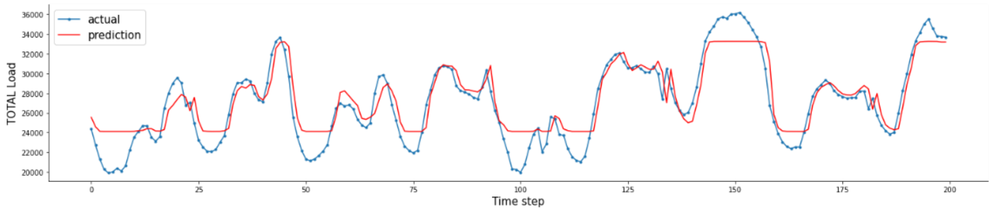


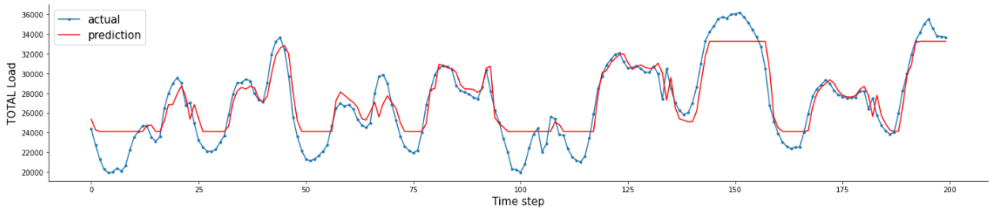
Fig. 3. Attention LSTM's configuration (Min-max normalization): (a) Att-LSTM, (b) Att-Grid Search-LSTM, and (c) Att-PSO-LSTM

This study tests the effectiveness of various model configurations using three main evaluation metrics: MAPE, RMSE, and R^2. MAPE measures prediction error as a percentage, allowing an intuitive understanding of model accuracy. RMSE provides a measure of prediction error in the same units as the prediction target, providing insight into the deviation of predictions from actual values. Meanwhile, R^2 shows the proportion of variability in the dataset that can be explained by the model, providing an idea of the model's fit to the data.

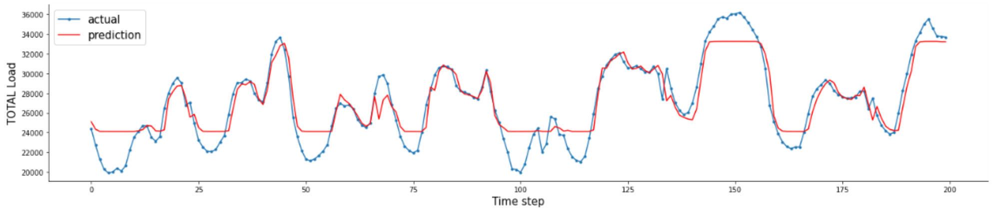
The model configurations with attention mechanism with min-max normalization: Att-LSTM (Figure 3a), Att-Grid-search-LSTM (Figure 3b), and Att-PSO-LSTM (Figure 3c) indicate that the models have been improved compared to the other baselines model without an attention mechanism and another model with z-score on Figure 4a to Figure 4c. The graph is presumably associated with the preceding discourse on the performance of the proposed model with an attention mechanism in forecasting the hourly energy consumption time series. The proximity of the lines suggests that the proposed model with the attention mechanism's predictions is highly accurate and closely aligned with the actual values. The peaks and troughs of both lines appear to be in close alignment, indicating that the model accurately represents the underlying trends and seasonality in the data.



(a)



(b)



(c)

Fig. 4. Attention LSTM's configuration (Z-Score normalization): (a) Att-LSTM, (b) Att-Grid Search-LSTM, and (c) Att-PSO-LSTM

Min-Max normalization demonstrates improved performance in comparison to Z-Score normalization, as indicated by the higher R^2 values observed in all models with Min-Max normalization (Table 3). This indicates a stronger correlation between forecasts and real data. As an illustration, the Attention-PSO-LSTM model, when combined with Min-Max normalization, has an outstanding R^2 value of 0.9233. In contrast, the standard LSTM model, which utilizes Z-Score normalization, only achieves an R^2 value of 0.4215.

Optimizing parameters is a crucial process for enhancing the performance of machine learning models. This study utilizes two parameter optimization strategies, namely Grid search and PSO, to identify the optimal parameter configuration that enhances the accuracy of model predictions. The impact of different parameter optimization strategies on model enhancement is evident through variations observed. Parameter optimization using Grid search enhances performance, as demonstrated by the Grid-search-LSTM model with Min-Max normalization, achieving a MAPE of

3.8666% and a R^2 of 0.9072, surpassing the performance of the regular LSTM model. PSO, with its superior parameter search capabilities, yields improved outcomes, as demonstrated in PSO-LSTM with Min-Max. This model achieves a MAPE of 3.7164% and a R^2 of 0.9118.

Table 3. Evaluation results

Normalization	Model	MAPE (%)	RMSE	R^2
Min-Max	LSTM	3.9658	0.0626	0.9010
	Grid-search-LSTM	3.8666	0.0606	0.9072
	PSO-LSTM	3.7164	0.0591	0.9118
	Attention LSTM	3.7067	0.0601	0.9088
	Attention Grid-search-LSTM	3.3658	0.0580	0.9150
	Attention PSO-LSTM	**3.1135**	**0.0551**	**0.9233**
Z-Score	LSTM	10.4299	0.7600	0.4215
	Grid-search-LSTM	4.8196	0.3894	0.8482
	PSO-LSTM	4.6450	0.3781	0.8568
	Attention LSTM	9.9980	0.4629	0.7681
	Attention Grid-search-LSTM	4.2946	0.3850	0.8516
	Attention PSO-LSTM	**4.1820**	**0.3827**	**0.8534**

Once put into practice, the attention approach significantly improves the precision of forecasts. Models that incorporate attention mechanisms, such as Attention LSTM with Min-Max, demonstrate higher performance than models that do not utilize attention. The MAPE reduces to 3.7067% and the R^2 value increases to 0.9088. This confirms that the model's ability to select important components of sequential data significantly improves accuracy. This in-depth analysis highlights the significance of carefully selecting data pre-processing strategies, rigorously optimizing parameters, and applying innovative technologies such as attention mechanisms in order to improve the performance of LSTM models for sequential data prediction tasks. Other factors that are taken into consideration include the importance of ensuring that the parameters are optimized. These are two necessary aspects.

IV. Conclusion

This research introduces a promising method for evaluating the Hourly Energy Demand Time Series dataset. This article aims to determine the most efficient methods for predicting energy demand by analyzing different model setups. LSTM combines attention mechanisms, grid search, and PSO. The dataset is normalized using both Min-Max and Z-Score approaches in order to maintain data consistency. The study subsequently investigates the enhancement of LSTM models by employing PSO and Grid Search techniques to select the hyperparameter, while the Attention mechanism is introduced to improve model performance by giving priority to important regions of the input sequence. The results show that all models utilizing Min-Max normalization exhibit superior MAPE, RMSE, and R^2 values compared to those using Z-Score normalization. The model with the highest performance is observed in Att-PSO-LSTM and Att-Grid-search-LSTM, Att-LSTM, PSO-LSTM, Grid-search-LSTM, and LSTM then follow it. This research enables a methodical evaluation of various model configurations based on LSTM with PSO and Grid search hyperparameters optimization under min-max and z-score normalization to achieve better effectiveness, perhaps resulting in an optimized solution for the current forecasting issue.

Declarations

Author contribution

All authors contributed equally as the main contributor of this paper. All authors read and approved the final paper.

Funding statement

The authors are grateful for the support provided by Universitas Ahmad Dahlan under grant number PD-119/SP3/LPPM-UAD/XII/2023.

Conflict of interest

The authors declare no known conflict of financial interest or personal relationships that could have appeared to influence the work reported in this paper.

Ethical and informed consent for data used

This article does not contain any studies with human participants or animals performed by the author, and therefore, ethical and informed consent from data subjects is not required.

Data availability and access

The data that support the findings of this study are openly available on Kaggle at https://www.kaggle.com/code/varanr/hourly-energy-demand-time-series-forecast.

Additional information

Reprints and permission information are available at http://journal2.um.ac.id/index.php/keds.

Publisher's Note: Department of Electrical Engineering and Informatics - Universitas Negeri Malang remains neutral with regard to jurisdictional claims and institutional affiliations.

References

[1] T. González Grandón, J. Schwenzer, T. Steens, and J. Breuing, "Electricity demand forecasting with hybrid classical statistical and machine learning algorithms: Case study of Ukraine," Appl. Energy, vol. 355, p. 122249, 2024.

[2] E. Yukseltan, A. Yucekaya, and A. H. Bilge, "Hourly electricity demand forecasting using Fourier analysis with feedback," Energy Strateg. Rev., vol. 31, p. 100524, 2020.

[3] M. Aldarraji, B. Vega-Márquez, B. Pontes, B. Mahmood, and J. C. Riquelme, "Addressing energy challenges in Iraq: Forecasting power supply and demand using artificial intelligence models," Heliyon, vol. 10, no. 4, p. e25821, 2024.

[4] S. O. Effiom, P. C. O. Effiom, R. Akwagiobe, and P. O. Odu, "Technical and economic appraisal for harnessing a proposed hybrid energy system nexus for power generation and CO2 mitigation in Cross River State, Nigeria," Appl. Eng. Technol., vol. 2, no. 2, pp. 153–175, 2023.

[5] H. Jin, J. Guo, L. Tang, and P. Du, "Long-term electricity demand forecasting under low-carbon energy transition: Based on the bidirectional feedback between power demand and generation mix," Energy, vol. 286, p. 129435, 2024.

[6] S. Eggimann, W. Usher, N. Eyre, and J. W. Hall, "How weather affects energy demand variability in the transition towards sustainable heating," Energy, vol. 195, p. 116947, 2020.

[7] A. Dietrich and C. Melville, "Energy demand characteristics and the potential for energy efficiency in sports statium and arenas," Duke Univ., 2011.

[8] A. Nikseresht and H. Amindavar, "Energy demand forecasting using adaptive ARFIMA based on a novel dynamic structural break detection framework," Appl. Energy, vol. 353, p. 122069, Jan. 2024.

[9] S. Chaturvedi, E. Rajasekar, S. Natarajan, and N. McCullen, "A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India," Energy Policy, vol. 168, p. 113097, 2022.

[10] P. Liu, F. Quan, Y. Gao, B. Alotaibi, T. R. Alsenani, and M. Abuhussain, "Green energy forecasting using multiheaded convolutional LSTM model for sustainable life," Sustain. Energy Technol. Assessments, vol. 63, p. 103609, 2024.

[11] Q. Wang, R. Suo, and Q. Han, "A study on natural gas consumption forecasting in China using the LMDI-PSO-LSTM model: Factor decomposition and scenario analysis," Energy, vol. 292, p. 130435, 2024.

[12] M. Bilgili and E. Pinar, "Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye," Energy, vol. 284, p. 128575, 2023.

[13] N. Mounir, H. Ouadi, and I. Jrhilifa, "Short-term electric load forecasting using an EMD-BI-LSTM approach for smart grid energy management system," Energy Build., vol. 288, p. 113022, 2023.

[14] L. Wang et al., "Optimal allocation of customer energy storage based on power big data and improved LSTM load forecasting," Energy Reports, vol. 11, pp. 3902–3913, 2024.

[15] J. Liu and S. Chen, "Non-stationary Multivariate Time Series Prediction with Selective Recurrent Neural Networks," in Lecture Notes in Computer Science, 2019, pp. 636–649.

[16] A. W. Saputra, A. P. Wibawa, U. Pujianto, A. B. P. Utama, and A. Nafalski, "LSTM-based Multivariate Time-Series Analysis : A Case of Journal Visitors Forecasting," Ilk. J. Ilm., vol. 14, no. 1, pp. 57–62, 2022.

[17] B. Li, X. Lv, and J. Chen, "Demand and supply gap analysis of Chinese new energy vehicle charging infrastructure: Based on CNN-LSTM prediction model," Renew. Energy, vol. 220, p. 119618, 2024.

[18] H. Chen, M. Zhu, X. Hu, J. Wang, Y. Sun, and J. Yang, "Research on short-term load forecasting of new-type power system based on GCN-LSTM considering multiple influencing factors," Energy Reports, vol. 9, pp. 1022–1031, 2023.

[19] Z. Niu, G. Zhong, and H. Yu, "A review on the attention mechanism of deep learning," Neurocomputing, vol. 452, pp. 48–62, Sep. 2021.

[20] H. Abbasimehr and R. Paki, "Improving time series forecasting using LSTM and attention models," J. Ambient Intell. Humaniz. Comput., Jan. 2021.

[21] A. Pranolo, Y. Mao, A. P. Wibawa, A. B. P. Utama, and F. A. Dwiyanto, "Robust LSTM With Tuned-PSO and Bifold-Attention Mechanism for Analyzing Multivariate Time-Series," IEEE Access, vol. 10, pp. 78423–78434, 2022.

[22] A. P. Wibawa et al., "Bidirectional Long Short-Term Memory (Bi-LSTM) Hourly Energy Forecasting," in E3S Web of Conferences, 2024, vol. 501, p. 1023.

[23] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, "DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model," Procedia CIRP, vol. 79, pp. 403–408, 2019.

[24] A. Mirzaei, S. R. Carter, A. E. Patanwala, and C. R. Schneider, “Missing data in surveys: Key concepts, approaches, and applications,” Res. Soc. Adm. Pharm., vol. 18, no. 2, pp. 2308–2316, Feb. 2022.

[25] A. Abayomi-Alli, M. O. Odusami, O. Abayomi-Alli, S. Misra, and G. F. Ibeh, “Long Short-Term Memory Model for Time Series Prediction and Forecast of Solar Radiation and other Weather Parameters,” in 2019 19th International Conference on Computational Science and Its Applications (ICCSA), Jul. 2019, pp. 82–92.

[26] N. Passalis, J. Kanniainen, M. Gabbouj, A. Iosifidis, and A. Tefas, “Forecasting Financial Time Series Using Robust Deep Adaptive Input Normalization,” J. Signal Process. Syst., vol. 93, no. 10, pp. 1235–1251, Oct. 2021.

[27] M. Ehteram, M. Afshari Nia, F. Panahi, and A. Farrokhi, “Read-First LSTM model: A new variant of long short term memory neural network for predicting solar radiation data,” Energy Convers. Manag., vol. 305, p. 118267, Apr. 2024.

[28] B. H. Shekar and G. Dagnew, “Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data,” in 2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP), Feb. 2019, pp. 1–8.

[29] Q. Kang, E. J. Chen, Z.-C. Li, H.-B. Luo, and Y. Liu, “Attention-based LSTM predictive model for the attitude and position of shield machine in tunneling,” Undergr. Sp., vol. 13, pp. 335–350, 2023.

[30] E. Vivas, H. Allende-Cid, and R. Salas, “A Systematic Review of Statistical and Machine Learning Methods for Electrical Power Forecasting with Reported MAPE Score,” Entropy, vol. 22, no. 12, p. 1412, Dec. 2020.

[31] F. Baig, L. Ali, M. A. Faiz, H. Chen, and M. Sherif, “How accurate are the machine learning models in improving monthly rainfall prediction in hyper arid environment?,” J. Hydrol., vol. 633, p. 131040, Apr. 2024.

[32] W. Sun and C. Huang, “A novel carbon price prediction model combines the secondary decomposition algorithm and the long short-term memory network,” Energy, vol. 207, p. 118294, Sep. 2020.