Journal of Operation and Automation in Power Engineering

Vol. XX, No. XX, Dec. 2022, Pages: XXXX (Proofed)

http://joape.uma.ac.ir

University of Mohaghegh Ardabili 1978

Research Paper

Predicting Electrical Load Demand Using Bagging Ensemble of Multi-Layer Perceptron and Adjusted Long Short-Term Memory with Metaheuristic Methods

Somayeh Talebzadeh , *Reza Radfar* *, and *Abbas Toloie Eshlaghy*

Department of Information Technology Management, Science and Research Branch, Islamic Azad University, Tehran, Iran.

Abstract— *Effective prediction of electric power demand is critical for maintaining the stability and reliability of the energy supply in both residential and industrial sectors. Accurate energy demand forecasting is essential for balancing consumption needs with grid stability. However, the complexity of energy consumption data, influenced by a variety of factors, makes this forecasting challenging. Traditional methods often struggle to capture the intricacies of such complex data, highlighting the need for more advanced and adaptable approaches. In this research, we propose a novel solution based on a Bagging ensemble of Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM) networks, combined through a voting mechanism to improve the accuracy and generalization ability of the model. Metaheuristic methods, including Particle Swarm Optimization (PSO) and the Genetic Algorithm (GA), are employed for optimal hyperparameter tuning of the LSTM. Unlike many existing studies that rely on proprietary or limited datasets, this approach uses publicly available data from the Electric Power Consumption dataset of Tetouan city (01-01-2017 to 12-31-2017), making it more accessible and applicable to broader contexts. It also enhances prediction performance by combining the results of multiple models, allowing for a more robust and accurate prediction of energy consumption. Experimental results demonstrate that the proposed approach significantly outperforms existing machine learning and deep learning methods.*

Keywords—Multi-layer perceptron, long short-term memory, bagging regressor, electrical load demand.

Nomenclature

GA	Genetic Algorithm
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MLP	Multi-Layer Perceptron
MSE	Mean Absolute Error
MSLE	Mean Squared Logarithmic Error
PSO	Particle Swarm Optimization
r2	Coefficient of Determination
RMSE	Root Mean Squared Error

1. Introduction

In today's world, with the development of human society, economic growth, and population growth, the need for various amenities in the domestic, industrial, and transportation sectors has increased in societies. In this regard, the vital need for energy and meeting the energy demand of various sectors has become very important [1]. Generally, the supply and demand are managed by scheduling on the generation side [2]. Accurate forecasting of demand is a vital issue for any organization so that it can respond to existing needs at the right time with proper planning and achieve maximum efficiency. Demand depends on many factors such as price, time, and place. If the supply of energy exceeds demand, the maintenance and storage costs lead to an increase in costs. On the other hand, supply falling short of demand when the energy is in high demand causes economic problems and creates a lot of dissatisfaction [3]. Accurate forecasts of total electricity demand can contribute to grid stability, system performance, reliability and safety through the detection of irregular events, leading to the possibility of online planning at higher levels and minimizing consumer dissatisfaction caused by unmet demand. Conducting studies focused on forecasting energy consumption using different techniques for different countries, especially developing countries, is significant. Accurate forecasting of this issue depends on several parameters that are related to issues such as economic and political conditions of the country, weather and market fluctuations. Accurate identification of these features has a strong impact on energy consumption prediction. The next issue that is very important to solve this problem is choosing a suitable modeling method. The main challenge in this field is the non-linear relationship between most of the input and output variables in such a way that it is not possible to find a precise mathematical relationship between the variables. The modeling method should be able to predict future events with the available data and perform well in the performance criteria [4]. Simple linear models and many existing statistical models cannot interpret complex nonlinear relationships, and the influence of

Received: 17 Jul. 2024
Revised: 30 Mar. 2025
Accepted: 30 Apr. 2025
*Corresponding author:
E-mail: radfar@gmail.com (R. Radfar)
DOI: 10.22098/joape.2025.15471.2187

outliers in the data can hinder the accuracy of predictions in these models. Research in this area has shown that artificial intelligence and machine learning have significantly contributed to the accurate estimation of aggregate demand [5]. Machine learning methods are valuable techniques that have had a profound impact on a wide range of applications and system automation and have been used in the modernization of industrial systems. Building machine learning models that incorporate existing recorded data allows accurate energy consumption estimation based on reality [6]. In fact, machine learning makes it possible to feed a huge amount of data to a computer algorithm and force the computer to make predictions based on the input data alone. This research presents an innovative approach using an ensemble method that combines Bagging models with MLP and LSTM, which are adjusted with metaheuristic algorithms. The goal of this approach is to improve prediction accuracy and model generalization. By leveraging publicly available data, the proposed method compares prediction performance with existing machine learning and deep learning methods. The novelty of this approach lies in the integration of multiple prediction models using an ensemble technique, combined with hyperparameter optimization through metaheuristics, which together provide a stronger and more accurate forecasting solution. Next, in the Section 2, a number of methods that have recently been used to solve this problem are examined. Section 3 introduces the dataset used in the research. In Sectioin 4, the proposed system of this research and the details of its different parts are described. In Section 5, the simulation environment, implementation details, and computational resources used in this study are described. In Sectioin 6, the results obtained from the proposed system and its comparison with other methods are presented. Finally, conclusions are reported in Section 7.

2. Literature review

In research [7] a method was proposed that uses Stacked Auto Encoders (SAE) to extract building energy consumption characteristics and Extreme Learning Machine (ELM) to predict energy consumption. The results of this research were analyzed on data from a retail building in Fremont, California, and showed that this method had the best predictive performance compared to the popular machine learning method. In the research [8], a new method based on the kNN (k-nearest neighbor) algorithm was proposed to predict energy demand and tested using real data. In this method, the user can interact with the system by analyzing the prediction and identifying the input parameters. The results obtained from the scenarios investigated in this method show its acceptable accuracy. In the article [9], a hybrid method based on faster k-medoids clustering, support vector machine and artificial neural network is proposed for device consumption forecasting and peak demand forecasting. This method was able to achieve 99.2% accuracy in predicting the consumption of electrical appliances, which is a very good result. In the article [10], a method was proposed to predict future electricity consumption for residential households, in which Gaussian mixture clustering is used to identify behavior clusters and XGBoost method is used to predict the behavior pattern of the future day. This method reached a value of 0.633 in the EDA metric based on the Euclidean distance, which is a good output.In the research [11] proposed ARIMAX, BOA-SVR, and BOA-NARX models to forecast annual electricity consumption in Saudi Arabia. In this research, the pre-processing operation was done by determining the important features and the Bayesian optimization algorithm (BOA) was used to improve the meta-parameters of the model. Among the investigated methods, the BOA-NARX method has the best performance with a MAPE value of 0.3219. In research [12], forecasting techniques based on time series, machine learning and hybrid models were implemented for load forecasting in Korea. SARIMAX time series model, ANN, SVR and LSTM machine learning models and SARIMAX-ANN, SARIMAX-SVR and SARIMAX-LSTM hybrid models were investigated. The obtained results showed that the combined methods work better than the time series and machine learning approaches. LSTM-based methods have also performed best among their group. In the paper [13], JLSTM model was proposed to predict electric load and price in big data. In the pre-processing stage, various techniques such as z-score method, Jaya optimization method, and normalization were applied on the data. The comparison of this method with other techniques such as LSTM and SVM showed the superiority of this method. In the article [14], a method based on Convolutional Neural Networks (CNN) and Long Short Term Memory (LSTM) was proposed to predict electricity consumption in smart homes. In this method, gray wolf optimization (GWO) is used to improve the performance of CNN-LSTM model. The implementation results showed that the proposed method of this research obtains fewer errors compared to the basic models. In the study [15], a total of 19 machine learning models were examined for the initial selection of models and a model based on stacking was proposed. Feature selection was done in the investigated data set by achieving two effective features. The obtained results showed that the proposed method based on stacking with all available features performs better than other implemented methods and the combination of other features. In the article, deep learning algorithms including long short-term memory (LSTM), gated recurrent units (GRU) and recurrent neural networks (RNN) were used to build prediction models for accurate estimation of electric load. The implementation results showed that the GRU model, which is actually a type of RNN, achieved the best performance in terms of accuracy and the least error. The paper [16] introduced a new method for electricity demand forecasting by combining PSO for feature extraction and CNN-MRMR for model training. The model achieved an accuracy of 97.20%.

In the paper [17], the ICEEMDAN-LSTM-TCN-Bagging model was proposed for short-term load forecasting methods. This model combines ICEEMDAN for data decomposition, LSTM-TCN for feature extraction. Experimental results show that the proposed model outperforms traditional methods and achieves higher forecasting accuracy with a root mean square error of 31.47 kW. In the study [18] a hybrid building load prediction method for office buildings was proposed. The approach uses EnergyPlus to generate a comprehensive building load database and uses the LightGBM algorithm to identify key feature variables for load prediction. Validation in real office buildings shows promising results, with a MAPE of 12.42% and 7.97% for cooling and heating load prediction, respectively. In the study [19], a hybrid forecasting model combining GEP and ANFIS is proposed for predicting electrical load demand in industries, which offers high accuracy with reduced errors and lower computation time compared to standalone models. This model was tested using real-time electrical load data from Uganda and achieved an RMSE of 0.0007.

Table 1 summarizes the information and results obtained from the reviewed methods. In this table, in addition to the characteristics of the research, the method used in it, the data used in the research and the results recorded in it are also specified.

Recent research advancements in energy load forecasting have focused on improving prediction accuracy, computational efficiency, and model robustness. Studies such as those by [16] and [17] have explored new hybrid models by combining optimization techniques with deep learning methods to enhance load forecasting performance. These advancements highlight the increasing reliance on data-driven approaches to manage the growing electricity demand in industrial and residential sectors. Despite the impressive results of these newer models, challenges remain, such as the need for high-quality, real-time data and the complexity of model integration. Additionally, many studies rely on proprietary or limited-access datasets, which may not be easily replicable or applicable to all regions or industries.

In this context, the present study proposes a new model that aims to improve prediction accuracy and generalization

Table 1. Comparison of previous works.

Research	Method	Data collection	Evaluation
[7]	SAE-ELM	A retail building in Fremont, California	RMSE / 59.1812
[8]	KNN	The German municipal utility in the SIT4Energy project	MAPE / 5.77%
[9]	Aster k-medoids clustering + SVM + ANN	Information of 550 households	Accuracy / 99.2%
[10]	Gaussian mixture clustering + XGBoost	500 residential users in a region of southeastern Spain	EDA / 0.633
[11]	BOA–NARX	TEC data of the Kingdom of Saudi Arabia	MAPE / 0.3219
[12]	SARIMAX + LSTM	Data of Korea was collected from KPX	RMSE / 3093.37
[13]	JLSTM	Big data from different resources (microgrids, smart buildings, smart meters, EVs, factories)	RMSE / 0.02
[14]	CNN-LSTM-GWO	Smart home energy consumption dataset	RMSE / 0.6213
[15]	Stacking	Historic data from 1975–2019 in Turkey	R-squared / 0.99
[20]	GRU	Tubas electricity company—palestine	MSE / 0.00215
[16]	PSO + CNN-MRMR	Power plant units data	Accuracy / 97.20%
[17]	ICEEMDAN-LSTM-TCN-bagging	Different load forecasting datasets	Root mean square error / 31.47
[18]	LightGBM	Real office buildings data	MAPE / 12.42% and 7.97% (cooling and heating load)
[19]	GEP-ANFIS-LTLF	Real-time electrical load data from Uganda	RMSE / 0.0007

capability by combining multiple prediction results. Furthermore, by using publicly accessible data, it provides a pathway for future research to refine and expand existing methods, ensuring continuous improvement and broader application. By leveraging publicly available data, this model bridges a significant gap in energy forecasting for developing regions and demonstrates its practical viability in real-world industrial settings.

3. Data set and data preprocessing

In this research, the Electric Power Consumption dataset is used. This dataset is widely utilized in the field of energy consumption analysis and prediction. It is particularly valuable for researchers and data scientists working on developing algorithms and models to forecast energy consumption patterns. The dataset contains energy consumption data from the city of Tetouan, Morocco, which is located in the northern part of the country, near the Mediterranean Sea. The city experiences a mild and rainy climate in winter and a hot, dry climate in summer. The dataset consists of 52,416 observations, each representing energy consumption over a 10-minute interval. This data was collected between 01-01-2017 and 31-12-2017. The dataset includes 9 features recorded for each sample, which can help in understanding the factors influencing energy consumption. These features may include variables such as temperature, humidity, wind speed, and other relevant metrics. By analyzing these features, researchers can identify patterns and relationships between them and energy consumption, which can be used to build more accurate predictive models.

This dataset has been widely used in numerous studies and research papers to develop and evaluate energy consumption forecasting models, which have significant implications for energy management, resource allocation, and overall sustainability efforts. Some examples of this dataset are presented in Table 2.

To utilize this dataset in research, a preprocessing and feature engineering phase is conducted. In this stage, attributes such as hour, day of the week, quarter, and month are extracted from the date and time. Additionally, simple moving averages for 10-day, 15-day, and 30-day periods are calculated from the data. The final characteristics obtained in the dataset used for this research are shown in Table 3. In this table, some of the examples in the dataset are given.

4. Proposed system

Since artificial intelligence techniques have demonstrated significant performance in the field of energy consumption prediction, this research investigates the efficiency of basic machine learning and deep learning methods to address this problem. The proposed method combines both machine learning and deep learning models to create a more accurate and robust prediction system. The final prediction results are obtained through an ensemble approach that leverages the performance of multiple models. The method involves an ensemble technique where the predictions from various models are combined using a weighted average approach to generate the final prediction. The general structure of the proposed method is presented in Fig. 1, which outlines the main steps of data preparation, model training, and final prediction. The system is structured as follows:

Data pre-processing: After obtaining the raw data, preprocessing steps are performed, including normalization, handling missing values, and transforming the data into a suitable format for training.

Data splitting: The dataset is split into two categories: training and testing data. Model Training: The LSTM model, adjusted with PSO and GA, is trained on the training data. Additionally, two Bagging regression models based on MLP are applied to the training dataset.

Model aggregation: After obtaining predictions from all models, a final result is computed using an averaging-based voting method. This helps to combine the strengths of each model and minimize the error.

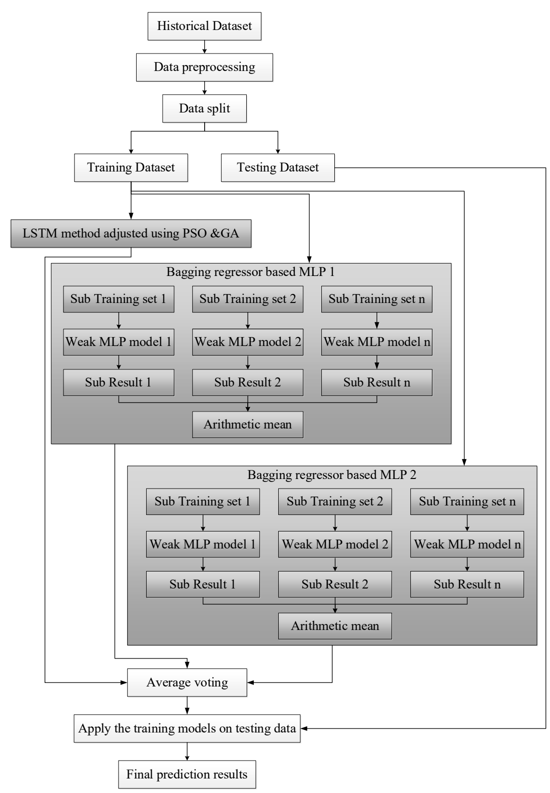

Fig. 1. Flowchart of proposed system.

4.1. Bagging regressor models based on MLP

A Multilayer Perceptron (MLP) is a type of fully connected artificial neural network that includes an input layer, one or more hidden layers, and an output layer. The MLP performs a series of

Table 2. Examples of data in the electric power consumption dataset.

Date time	**1/1/2017 0:00**	**1/1/2017 0:10**	**1/1/2017 0:20**	**1/1/2017 0:30**	**1/1/2017 0:40**
Temperature ($°C$)	6.559	6.414	6.313	6.121	5.921
Humidity (%)	73.8	74.5	74.5	75	75.7
Wind speed (m/s)	0.083	0.083	0.08	0.083	0.081
General diffuse flows	0.051	0.07	0.062	0.091	0.048
Diffuse flows	0.119	0.085	0.1	0.096	0.085
Power consumption_Zone1 (W)	34055.6962	29814.68354	29128.10127	28228.86076	27335.6962
Power consumption_Zone2 (W)	16128.87538	19375.07599	19006.68693	18361.09422	17872.34043
Power consumption_Zone3 (W)	20240.96386	20131.08434	19668.43373	18899.27711	18442.40964

Table 3. New features in the dataset after feature engineering.

Date time	**1/1/2017 0:00**	**1/1/2017 0:10**	**1/1/2017 0:20**	**1/1/2017 0:30**	**1/1/2017 0:40**
Temperature	6.559	6.414	6.313	6.121	5.921
Humidity	73.8	74.5	74.5	75	75.7
Wind speed	0.083	0.083	0.08	0.083	0.081
General diffuse flows	0.051	0.07	0.062	0.091	0.048
Diffuse flows	0.119	0.085	0.1	0.096	0.085
Power consumption_Zone1	34055.6962	29814.68354	29128.10127	28228.86076	27335.6962
Power consumption_Zone2	16128.87538	19375.07599	19006.68693	18361.09422	17872.34043
Power consumption_Zone3	20240.96386	20131.08434	19668.43373	18899.27711	18442.40964
Hour	0	0	0	0	0
Day Of week	6	6	6	6	6
Quarter	1	1	1	1	1
Month	1	1	1	1	1
Year	2017	2017	2017	2017	2017
Day Of year	1	1	1	1	1
Day Of month	1	1	1	1	1
Week Of year	52	52	52	52	52
SMA10	NaN	NaN	NaN	NaN	NaN
SMA15	NaN	NaN	NaN	NaN	NaN
SMA30D	NaN	NaN	NaN	NaN	NaN

non-linear transformations on the input feature set, which allows it to capture complex patterns and relationships in the data, making it highly effective for regression tasks [21].

The method of Bagging regression models based on MLP involves several MLP regressors, each trained on a random subset of the data. In Bagging, sampling is performed with replacement to create different training subsets. The predictions from each regressor are then aggregated to generate the final prediction. This process helps reduce the variance of the model and improve prediction stability.

In this study, the number of MLP sub-regressors is denoted by N. A larger N increases the model's potential to capture more diverse patterns, leading to improved accuracy. However, the model complexity and training time also increase with a larger N. Based on empirical testing, the value of N was set to 10 in the first model and 20 in the second model. The MLP regressors in both models were trained using 10 and 20 different data sets, respectively, achieving satisfactory results.

4.2. LSTM model adjusted with PSO and GA

The Long Short-Term Memory (LSTM) network is specifically chosen for its proficiency in handling sequential data due to its unique architecture, which includes memory cells that capture and retain temporal dependencies. LSTM networks are especially well-suited for time-series forecasting because of their ability to model long-term dependencies in sequential data [22].

In machine learning, hyperparameters are predefined settings that are not learned during the training process but must be set by the user before training. The performance of the LSTM model is highly dependent on the selection of appropriate hyperparameters. Key hyperparameters in the LSTM model include the number of hidden layers, the number of LSTM units per layer, and other settings that directly impact model performance.

The primary objective of hyperparameter tuning is to identify the optimal hyperparameter values to enhance model performance. To achieve this, PSO and GA are employed in the proposed system for hyperparameter optimization. As shown in Fig. 2, both PSO and GA independently search for a set of optimal hyperparameters and generate models that minimize the error when validated against independent data. The final model is selected based on the highest prediction accuracy and the lowest error, which is then used to fine-tune the LSTM model's hyperparameters.

5. Simulation environment

The implementation was conducted in Python, utilizing libraries such as pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, and tensorflow.keras. Dataset details are provided in Section 3, with 70% of the data used for training and 30% for testing.

For neural networks, we evaluated different solvers, including lbfgs, sgd, and adam, and ultimately selected Adam as the optimizer due to its superior performance. The experiments were executed on a system with an Intel Core i7 processor, 16GB RAM, and a Linux/Windows operating system.

Additionally, we conducted some tests using "Google Colab", which provides cloud-based computational resources. The hardware configuration of Google Colab includes:

- **CPU:** Intel Xeon (2 vCPUs, 2.2 GHz)
- **RAM:** Up to 12GB (depending on runtime allocation)
- **GPU:** NVIDIA Tesla T4 (16GB VRAM) or NVIDIA Tesla K80 (12GB VRAM), depending on session availability

Using Google Colab allowed us to leverage GPU acceleration for training deep learning models, improving computational efficiency.

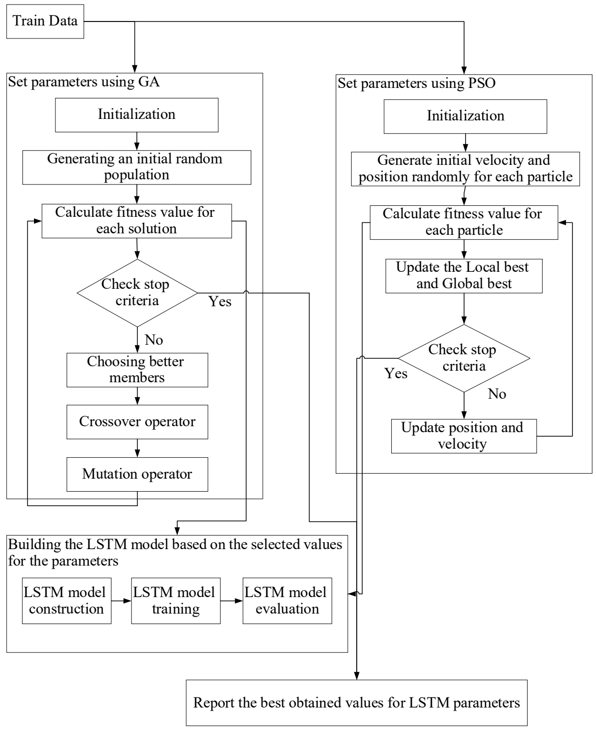

Fig. 2. Adjusting LSTM parameters using PSO and GA.

Table 4. Obtained values for adjusting LSTM parameters based on PSO and GA methods.

Layer (type)	Output shape	Param #
lstm (LSTM)	(None, 16)	1152
dense_3 (Dense)	(None, 10)	170
dense_4 (Dense)	(None, 8)	88
dense_5 (Dense)	(None, 3)	27
dense_6 (Dense)	(None, 1)	4
Total params: 1,441		
Trainable params: 1,441		
Non-trainable params: 0		

6. Results

In this section, the results obtained from the proposed research system and its comparison with the basic methods reviewed are presented. The performance of the proposed system has been evaluated using a range of criteria, including explained variance, MSLE, R^2, MAE, MSE, RMSE, and MAPE, which provide a comprehensive overview of the model's accuracy, robustness, and ability to generalize. Below is a brief definition of each of the criteria used. In the following equations, Y are the actual values, $\hat{Y}$ are the predicted values, and n is the number of samples.

Explained variance: This metric measures the difference between the target variance (the variation in the actual data) and the prediction error variance (the variation in the differences between the predicted and actual values). An upward trend in explained variance indicates that the model is improving in explaining the variation in the data.

$$EV = 1 - \frac{\mathrm{Var}(Y - \hat{Y})}{\mathrm{Var}(Y)} \tag{1}$$

MSLE (Logarithmic Mean Square Error): MSLE is used to evaluate the performance of a model, especially for regression problems. It calculates the mean square error after taking the natural logarithm of each predicted value. A lower MSLE value suggests better model performance.

$$\mathrm{MSLE} = \frac{1}{n}\sum_{i=1}^{n}\left(\log(y_i + 1) - \log(\hat{y}_i)\right)^2 \tag{2}$$

R^2 (Coefficient of Determination): R^2 measures how well the regression model fits the data A higher R^2 value (ranging from 0 to 1) means the model explains more of the variation in the data, leading to a better fit.

$$R^2(Y, \hat{Y}) = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2} \tag{3}$$

MAE (Mean Absolute Error): MAE calculates the average absolute difference between the predicted and actual values. It measures the average magnitude of the errors made by the model. Lower MAE values indicate better model performance.

$$\mathrm{MAE}(Y, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right| \tag{4}$$

MSE (Mean Squared Error): MSE is the average of the squared differences between the true and estimated values. Like MAE, it measures the average magnitude of the errors made by the model. Lower MSE values suggest better model accuracy.

$$\mathrm{MSE}(Y, \hat{Y}) = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{5}$$

RMSE (Root Mean Square Error): RMSE is the square root of MSE. It provides a measure of the average difference between the model's predicted values and the actual values. Lower RMSE values indicate better model predictions.

$$\mathrm{RMSE}(Y, \hat{Y}) = \sqrt{\mathrm{MSE}(Y, \hat{Y})} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{6}$$

MAPE (Mean Absolute Percentage Error): MAPE calculates the average absolute percentage difference between the predicted and actual values. It expresses the error as a percentage, making it useful for comparing models across different scales. A lower MAPE value signifies better forecast accuracy.

Table 5. The results of LSTM adjusted with GA and PSO.

Evaluation criteria	Train results	Test results
Explained_variance	0.9631	0.9596
MSLE	0.002	0.0017
r2	0.9569	0.9596
MAE	1011.8132	856.1678
MSE	2294596.6789	1761456.6761
RMSE	1514.7926	1327.1988
MAPE	0.9695	0.972

Table 6. The results of proposed system LSTM adjusted with GA and PSO.

Evaluation criteria	LSTM tuned with GA and PSO	Proposed system
Explained_variance	0.9596	0.9728
MSLE	0.0017	0.0013
r2	0.9596	0.9718
MAE	856.1678	760.9908
MSE	1761456.6761	1229701.8233
RMSE	1327.1988	1108.9192
MAPE	0.972	0.9302

Table 7. Results of CNN, XGBRegressor, SVR, RandomForestRegressor and proposed system.

Evaluation criteria	CNN	XGB regressor	SVR	Random forest regressor	Proposed system
Explained variance	0.9183	0.4306	0.8464	0.9728	0.948
MSLE	0.0033	0.031	0.0069	0.0013	0.003
$\mathbf{r^2}$	0.9158	0.3827	0.846	0.9718	0.932
MAE	1336.7707	4276.7629	2086.55	760.9908	1368.9875
MSE	3673252.7289	26915794.5338	6714329.5133	1229701.8233	2963509.9815
RMSE	1916.5732	5188.0434	2591.2023	1108.9192	1721.4848
MAPE	0.9573	0.8475	0.9315	0.9302	0.9553

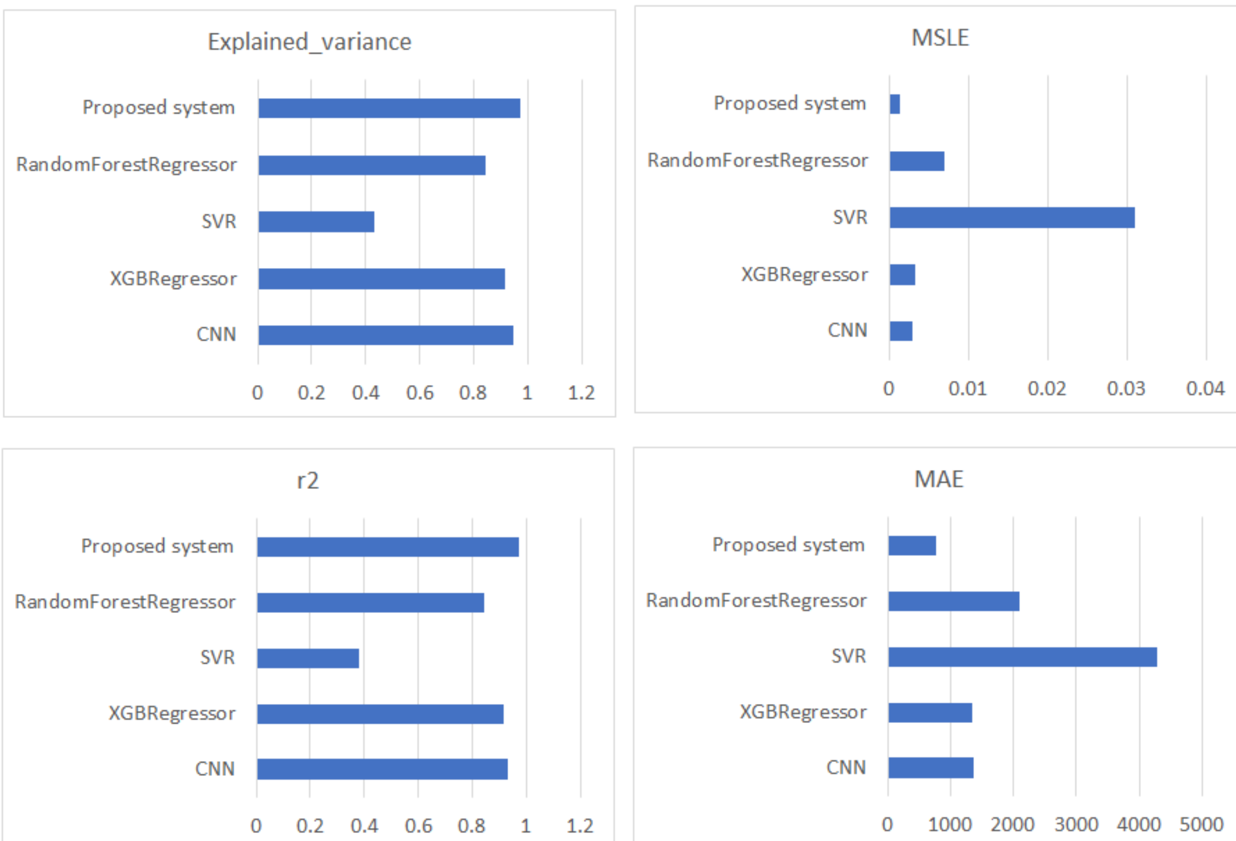

Fig. 3. Comparison charts of methods in explained_variance, MSLE, r2, and MAE criteria.

$$\mathrm{MAPE}(Y, \hat{Y}) = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{Y_i - \hat{Y}_i}{Y_i} \right| \tag{7}$$

6.1. Results of LSTM tuned with GA and PSO

In this section, the results obtained from the adjusted LSTM method using PSO and GA methods are presented. By running the PSO and GA methods to obtain the best values for the hyperparameters, the same results were obtained that were used to adjust the LSTM. The Table 4 shows the values obtained for the hyperparameters of the LSTM method. The number of layers and the number of neurons in them are among those whose optimal values are determined by PSO and GA methods. By adjusting the LSTM parameters based on these values, the results presented in the Table 5 have been obtained using the LSTM method. In this table, different criteria are presented in the training and test data. Obtaining good results in different criteria shows the good performance of LSTM method. However, in order to obtain more suitable results in the proposed system, a method based on voting has been proposed, in which this LSTM method adjusted with PSO and GA methods acts as a part of it. It is expected that the final results will improve significantly.

In this section, the results obtained from the adjusted LSTM method using PSO and GA methods are presented. By running the PSO and GA methods to obtain the best values for the hyperparameters, the same results were obtained that were used to adjust the LSTM.

The Table 4 presents the values obtained for the hyperparameters of the LSTM method, including the number of layers and neurons. These were optimized using PSO and GA methods to achieve the best performance. The fine-tuned LSTM model, using these methods, demonstrated strong predictive power as shown in Table 5. Compared to basic methods, the LSTM method achieved optimal values in several evaluation metrics, such as explained variance and R^2, which are indicative of the model's effectiveness in capturing the underlying patterns in the data.

6.2. Results of proposed system

In this section, the results obtained from the proposed system are presented. Table 6 compares the performance of the proposed system to the LSTM model adjusted with PSO and GA methods. It is evident that the proposed system significantly outperforms the individual LSTM model across all the evaluation criteria. Specifically, the proposed method achieves higher values in explained variance and R^2, and lower values in MSLE, MAE, MSE, RMSE, and MAPE. These improvements suggest that the proposed ensemble approach, which integrates the LSTM model, leads to more accurate and stable predictions by reducing the error in forecasting.

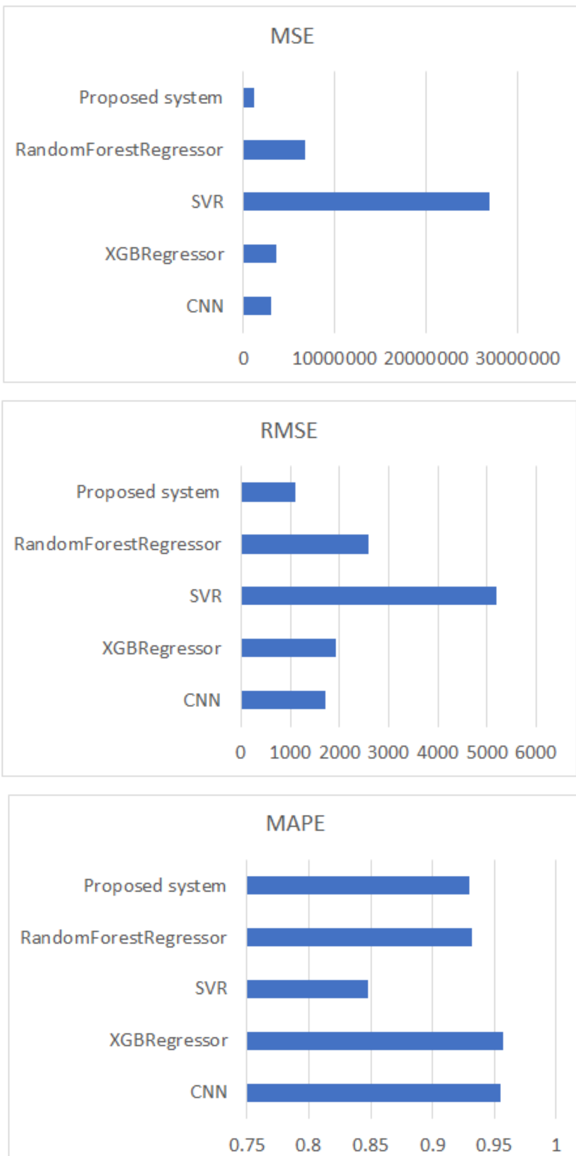

Fig. 4. Comparison charts of methods in MSE, RMSE, and MAPE criteria.

6.3. Comparison with basic methods

To compare the proposed system, methods such as CNN, XGBRegressor, SVR, and RandomForestRegressor were implemented. Table 7 clearly shows that the proposed method significantly outperforms these basic methods in most of the criteria, indicating its superiority in making accurate predictions. The only exception is the MAPE metric, where SVR outperforms the proposed system. However, this should not overshadow the overall performance of the proposed method, which excels in other important criteria, such as R^2 and explained variance.

The comparison charts in Figs. 3 and 4 provide a visual representation of how the proposed system compares with other methods. These charts illustrate the trends and differences in performance across various evaluation criteria, providing clear evidence of the proposed method's strengths.

The results obtained from this study highlight the superiority of the proposed system in comparison to basic machine learning models. The proposed system integrates three models: the LSTM model fine-tuned using PSO and GA, and two Bagging-based MLP regression models. These models are combined using an averaging-based voting method, which enables the system to leverage the strengths of each individual model for improved overall performance.

The results show that the proposed ensemble approach outperforms the individual LSTM model in all key evaluation metrics, including explained variance, R^2, MAE, MSE, RMSE, and MSLE. The increase in explained variance and R^2 indicates that the proposed system is better able to capture the underlying patterns in the data, leading to more accurate predictions. Furthermore, the reduction in error metrics such as MAE, MSE, RMSE, and MSLE confirms the enhanced accuracy and robustness of the ensemble system.

Compared to basic methods like RandomForestRegressor, XGBRegressor, and SVR, the proposed system demonstrates a clear advantage. While the SVR method achieves a lower MAPE, it performs significantly worse in the other criteria, particularly explained variance and R^2, where the proposed system excels. This highlights the importance of using multiple evaluation metrics to assess model performance comprehensively. The use of the averaging-based voting method helps to mitigate the weaknesses of individual models. The key strength of the proposed system lies in the ensemble approach, where the combination of multiple models, each capturing different aspects of the data, leads to a more reliable and stable prediction.

7. Conclusion

In this research, we propose a novel method for predicting energy consumption using a voting-based ensemble approach that integrates multiple machine learning techniques. The combination of different models reduces errors, improves prediction accuracy, and enhances the robustness of the system. The dataset was pre-processed by extracting time-related features (e.g., hour, day of the week, month) and computing rolling averages (10-day, 15-day, 30-day) from the original data. We employed an LSTM model optimized through PSO and GA, along with two Bagging Regressor models based on MLP with different architectures. The results of these three models were combined using an averaging method to produce the final predictions. The experimental results demonstrated that the proposed ensemble method significantly outperforms individual machine learning techniques, proving its effectiveness in energy consumption forecasting. This approach is especially useful in real-world scenarios where data varies and demand forecasting must be adaptable. Future work can focus on incorporating additional external features and extending the dataset to multiple regions. While our dataset includes essential variables such as temperature, humidity, wind speed, and solar radiation, there are other factors that could potentially improve prediction accuracy. Data on seasonal trends, public holidays, and events that influence energy consumption patterns could be incorporated. These features, combined with real-time data from smart sensors or IoT devices, would provide richer insights into consumption behavior and improve the robustness of the predictive model. Furthermore, the current model relies on data from a single city, Tetouan. Expanding the dataset to include data from multiple cities or regions with different demographic and environmental characteristics would make the model more generalizable.

REFERENCES

[1] M. M. Forootan, I. Larki, R. Zahedi, and A. Ahmadi, "Machine learning and deep learning in energy systems: A review," *Sustainability*, vol. 14, no. 8, p. 4832, 2022.

[2] N. Kumar, S. Dahiya, and K. Singh Parmar, "Multi-objective economic emission dispatch optimization strategy considering battery energy storage system in islanded microgrid," *J. Oper. Autom. Power Eng.*, vol. 12, no. 4, pp. 296–311, 2024.

[3] R. Tugay and S. G. Oguducu, "Demand prediction using machine learning methods and stacked generalization," *arXiv preprint*, 2020.

[4] D. Kontogiannis, D. Bargiotas, A. Daskalopulu, A. I. Arvanitidis, and L. H. Tsoukalas, "Structural ensemble regression for cluster-based aggregate electricity demand forecasting," *Electr.*, vol. 3, no. 4, pp. 480–504, 2022.

[5] D. Mhlanga, "Artificial intelligence and machine learning for energy consumption and production in emerging markets: A review," *Energies*, vol. 16, no. 2, p. 745, 2023.

[6] S. Kapp, J.-K. Choi, and T. Hong, "Predicting industrial building energy consumption with statistical and machine-learning models informed by physical system parameters," *Renew. Sustain. Energy Rev.*, vol. 172, p. 113045, 2023.

[7] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, "Building energy consumption prediction: An extreme deep learning approach," *Energies*, vol. 10, no. 10, p. 1525, 2017.

[8] A. I. Grimaldo and J. Novak, "Combining machine learning with visual analytics for explainable forecasting of energy demand in prosumer scenarios," *Procedia Comput. Sci.*, vol. 175, pp. 525–532, 2020.

[9] E. U. Haq, X. Lyu, Y. Jia, M. Hua, and F. Ahmad, "Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach," *Energy Rep.*, vol. 6, pp. 1099–1105, 2020.

[10] F. Lazzari, G. Mor, J. Cipriano, E. Gabaldon, B. Grillone, D. Chemisana, *et al.*, "User behaviour models to forecast electricity consumption of residential customers based on smart metering data," *Energy Rep.*, vol. 8, pp. 3680–3691, 2022.

[11] S. H. Almuhaini and N. Sultana, "Forecasting long-term electricity consumption in saudi arabia based on statistical and machine learning algorithms to enhance electric power supply management," *Energies*, vol. 16, no. 4, p. 2035, 2023.

[12] J. Lee and Y. Cho, "National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model?," *Energy*, vol. 239, p. 122366, 2022.

[13] R. Khalid, N. Javaid, F. A. Al-Zahrani, K. Aurangzeb, E.-u.-H. Qazi, and T. Ashfaq, "Electricity load and price forecasting using jaya-long short term memory (jlstm) in smart grids," *Entropy*, vol. 22, no. 1, p. 10, 2019.

[14] T. Singh, A. Solanki, S. K. Sharma, N. Jhanjhi, and R. M. Ghoniem, "Grey wolf optimization based cnn-lstm network for the prediction of energy consumption in smart home environment," *IEEE Access*, 2023.

[15] J. Hoxha, M. Y. Çodur, E. Mustafaraj, H. Kanj, and A. El Masri, "Prediction of transportation energy demand in türkiye using stacking ensemble models: Methodology and comparative analysis," *Appl. Energy*, vol. 350, p. 121765, 2023.

[16] T. Balachander, S. A. Khot, R. Huseyn, S. Garg, S. Vijay, and V. Pandey, "An innovative method for short term electrical load forecasting based on adaptive cnn-mrmr model," in *Proc. Int. Conf. Electron., Comput., Commun. Control Technol.*, IEEE, 2024.

[17] G.-Q. Zheng, L.-R. Kong, Z.-E. Su, M.-S. Hu, and G.-D. Wang, "Approach for short-term power load prediction utilizing the iceemdan–lstm–tcn–bagging model," *J. Electr. Eng. Technol.*, vol. 20, no. 1, pp. 231–243, 2025.

[18] H. Lian, Y. Ji, M. Niu, J. Gu, J. Xie, and J. Liu, "A hybrid load prediction method of office buildings based on physical simulation database and lightgbm algorithm," *Appl. Energy*, vol. 377, p. 124620, 2025.

[19] M. S. Bakare, A. Abdulkarim, A. N. Shuaibu, and M. M. Muhamad, "A hybrid long-term industrial electrical load forecasting model using optimized anfis with gene expression programming," *Energy Rep.*, vol. 11, pp. 5831–5844, 2024.

[20] M. Abumohsen, A. Y. Owda, and M. Owda, "Electrical load forecasting using lstm, gru, and rnn algorithms," *Energies*, vol. 16, no. 5, p. 2283, 2023.

[21] M. Shahriyari and H. Khoshkhoo, "A deep learning-based approach for comprehensive rotor angle stability assessment," *J. Oper. Autom. Power Eng.*, vol. 10, no. 2, pp. 105–112, 2022.

[22] N. Almuratova, M. Mustafin, K. Gali, M. Zharkymbekova, D. Chnybayeva, and M. Sakitzhanov, "Enhancing microgrid resilience with lstm and fuzzy logic for predictive maintenance," *J. Oper. Autom. Power Eng.*, vol. 12, pp. 1–20, 2024.