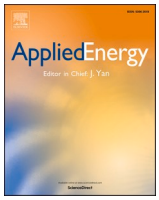
Applied Energy 343 (2023) 121230

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy



TLIA: Time-series forecasting model using long short-term memory integrated with artificial neural networks for volatile energy markets

Dalal AL-Alimi [a], Ayman Mutahar AlRassas [b], Mohammed A.A. Al-qaness [c,*], Zhihua Cai [a], Ahmad O. Aseeri [d,*], Mohamed Abd Elaziz [e,h,i,j], Ahmed A. Ewees [f,g]

[a] School of Computer Science, China University of Geosciences, Wuhan, 430074, China
[b] School of Petroleum Engineering, China University of Petroleum (East China), Qingdao, 266555, China
[c] College of Physics and Electronic Information Engineering, Zhejiang Normal University, Jinhua, 321004, China
[d] Department of Computer Science, College of Computer Engineering and Sciences, Prince Sattam Bin Abdulaziz University, Al-Kharj, 11942, Saudi Arabia
[e] Department of Mathematics, Faculty of Science, Zagazig University, Zagazig, 44519, Egypt
[f] College of Computing and Information Technology, University of Bisha, Bisha, 61922, Saudi Arabia
[g] Department of Computer, Damietta University, Damietta, 34517, Egypt
[h] Artificial Intelligence Research Center (AIRC), Ajman University, Ajman, 346, United Arab Emirates
[i] Department of Artificial Intelligence Science and Engineering, Galala University, Suze, 435611, Egypt
[j] Department of Electrical and Computer Engineering, Lebanese American University, Byblos, 13-5053, Lebanon

HIGHLIGHTS

- Use the ETR method to increase the data stationariness.
- Propose a new hybrid model for energy forecasting.
- Use six different energy time series datasets.
- The hybrid model with the ETR method provides the best results.
- The proposed methods yield the best results for lengthy and short data periods.

ARTICLE INFO

Keywords:
ETR
Time series forecasting
Hybrid model
ANN
LSTM
Energy

ABSTRACT

Due to weather and political fluctuations that significantly impact the production and price of energy sources, enhancing data distribution and reducing data complexity is crucial to achieving accurate forecasting. Additionally, it is essential to provide a flexible forecasting model capable of handling rapid changes in the energy market and effectively anticipating energy supplies and demands. This study introduces a novel method to deal with energy market fluctuations in the long and short term and provide highly accurate forecasts for various energy data. It uses the Enhancing Transformation Reduction (ETR) method to improve the stationarity of the data, reduce seasonality and trend, and resolve rapid fluctuations. The output of ETR is then passed into a hybrid forecasting model referred to as " Time-Series Forecasting Model using Long Short-Term Memory integrated with Artificial Neural Networks" (TLIA). The TLIA model benefits from transfer learning, which transmits the output of the LSTM layers into the ANN layers, enabling TLIA to base its work on the best performance and continue improving it. The study evaluates and tests its methods using six different datasets, including the electricity dataset of Victoria State, the oil price for the West Texas Intermediate, the Elia Grid load dataset, and wind power production. In addition to its characteristics, ETR accelerates and enhances the TLIA processing to achieve the

Abbreviations: ML, Machine Learning; DL, Deep Learning; PCA, Principal Component Analysis; ANN, Artificial Neural Network; LSTM, Long Short-Term Memory; CNN, Convolutional Neural Network; TLIA, Transferring Long Short-Term Memory into an Artificial Neural Network Model; BN, Batch Normalization; ETR, Enhancing Transformation Reduction; TR, Transfer Learning; MD, Morphological Dilation; K, Mask; R, Marker; RNN, Recurrent Neural Network; MAE, Mean Absolute Error; RMSE, Root Mean Square Error; RMSPE, Root Mean Square Percentage Error; ADF, Augmented Dickey-Fuller Test; GRU, Gated Recurrent Units; MM, MinMaxScaler.

* Corresponding authors.
E-mail addresses: dalal@cug.edu.cn (D. AL-Alimi), alrassas1989@gmail.com (A.M. AlRassas), alqaness@zjnu.edu.cn (M.A.A. Al-qaness), zhcai@cug.edu.cn (Z. Cai), a.aseeri@psau.edu.sa (A.O. Aseeri), abd_el_aziz_m@yahoo.com (M. Abd Elaziz), ewees@du.edu.eg (A.A. Ewees).

https://doi.org/10.1016/j.apenergy.2023.121230
Received 5 April 2023; Accepted 25 April 2023
Available online 16 May 2023

highest accuracy compared to seven forecasting models in all six datasets. The TLIA is often 40 times or more superior to competing models. Compared to another model, the Mean Absolute Error (MAE) results of TLIA range between (0.008 and 0.088) versus (0.77 and 4318.544).

1. Introduction

Renewable and non-renewable energy needs have grown to affect many parts of our society, making this one of the most difficult issues of the 20th century. Therefore, protecting energy availability is a crucial target on a global scale [1]. Energy is imperative for sustaining social welfare and economic growth, raising living standards, and achieving social security [2]. Energy resources are considered a pivotal factor in economic and sustainable development [3,4]. Hydrocarbon resources and renewable energy are the primary supply energy sources worldwide [5]. Renewable energy and fossil fuels are used in every part of contemporary society [1,5]. The remarkable demand for energy has escalated significantly due to the technological revolution in different industrial sectors and infrastructure development. Thus, it is important to find a technique that has the capability to forecast energy supplies/demand accurately. In the 20th century, machine learning (ML), including deep learning (DL) methods, has emerged as the most cutting-edge technology that could provide an accurate forecast in various aspects.

Energy supply, demand, and pricing fluctuate continuously during the day, month, and year. In this context, forecasting energy fluctuations is a precise way to avoid energy sector uncertainties [6]. The reliability of forecasting in the energy market sectors is [1] regularly a main predictor in planning and making decisions for companies, investors, and governments [7,8]. Numerous studies have been conducted to forecast the energy sectors, including oil price [8,9], oil production and consumption [10,11], wind energy [12–14], electricity price [7,15,16], electricity load [16–18], and solar energy [19]. Moreover, due to the volatile, nonlinear, complicated, and chaotic properties of energy generation and price, it has always been challenging to correctly anticipate the direction of the energy market time series, particularly in light of recent weather and energy price fluctuations [12]. As a result, previous studies demonstrated the significant role of forecasting techniques in the energy sector [18]. The role of machine learning and deep learning in the energy sector has blossomed significantly as the optimal approaches for predicting energy prices, supplies, and demands.

Machine learning (ML) has emerged as an alternative approach for conducting energy forecasting and surveillance. Its main target is to use automatic learning to build a reliable energy forecasting model [20]. Machine learning has attained remarkable performance when forecasting renewable and fossil fuel energy. Furthermore, several studies were conducted to establish ML algorithms that could forecast energy supply and demand. Previously, traditional approaches for forecasting, such as the autoregressive integral moving average, random walk, generalized autoregressive conditional heteroscedasticity, and vectorial autoregression, have been utilized [21]. Traditional approaches have acceptable predictive effects for variables with linear correlation, but they cannot identify the nonlinear properties of time series data. Consequently, with the growth of machine learning, several improved approaches for processing nonlinear data have been developed [21].

For nonstationary and nonlinear carbon price forecasting, [22] suggested a new multiscale nonlinear ensemble learning paradigm that uses empirical mode decomposition (EMD) to simplify the data and a least square support vector machine (LSSVM) with a prototype kernel function. In [23], the authors divided data processing into two stages. First, the carbon price network transferred the carbon price data to map the price based on the coarse granulation method. The second stage effectively extracted information fed from the first stage to reconstruct the original carbon price data using the topology structure. Principal component analysis (PCA) was utilized by [24] to transform the input dataset into a subset containing the most informative factors. The PCA output was input into various ML prediction methods to forecast the carbon price; the PCA results sped up the prediction process but decreased its precision. Alrassas et al. [25] developed a time series forecasting model for oil production using oil production data. They integrated an adaptive neuro-fuzzy inference system and a slime mould algorithm to develop the model, and the results show a remarkable performance for forecasting oil production. Lee J. et al. [18] compared traditional forecasting, machine learning, and the hydride model to identify the optimal peak load prediction models for electricity load in Korea. Machine learning demonstrated high performance in forecasting the peak load forecasting model for electricity. Al-qaness et al. [13] developed an efficient machine-learning model to forecast wind power energy. The authors employed an optimized dendritic neural regression (DNR) with an integrated seagull optimization algorithm and Aquila optimizer to train and optimize the DNR parameters. The performance of the developed forecasting model reveals excellent results. Yang W et al. [26] developed an advanced forecasting model for electricity prices using an adaptive data preprocessing algorithm, a kernel-based extreme learning machine, and a chaotic sine cosine algorithm. The integrated model demonstrated high performance with excellent results. Ultimately, the application of machine learning to fossil and renewable energy has gained more attention due to its capability to tackle the issues of energy models and predict them effectively. Although the ML approaches have offered adequate performance and continue to be utilized, they cannot handle more complicated or large datasets. Numerous researchers have converted to deep learning techniques for this and other reasons.

The numerous innovations in the domain of neural networks (NNs) have given rise to deep learning [27]. Deep learning techniques were developed and have been employed to tackle various predicting issues for several years [19]. The application of deep learning for forecasting the energy sector has dramatically increased due to its high efficiency and ability to achieve optimal results [19,28,29]. DL models were developed for use in computer science domains, e.g., hyperspectral image classification [30], carbon trapping [31], and natural language processing [32]. Recently, DL has expanded to several applications in the field of energy sectors. Because the artificial neural network (ANN) model provides high precision and rapid processing, it was utilized to forecast oil production or extraction [33]. Lago et al. [27] developed an efficient DL framework to predict electricity prices. The results indicate that the deep learning models achieve excellent performance. Chen et al. [34] applied data augmentation to predict wind power, considering the physics-oriented and data-oriented time-series wind data augmentation approaches. The advanced augmentation method and predicting algorithm are engaged over five turbines. The results point out that the data-oriented outperformed the physics-oriented. Yazici et al. [35] introduced an advanced deep-learning approach to forecasting the daily and weekly crude oil price (COP) by integrating an optimized variation mode decomposition to decompose COP data into multiple modes to augment the performance of predicting and an AdaBoost Random Forest to carry out forecasting modeling on high-frequency modes. The results show that deep learning forecasting is a powerful framework for building forecasting models, making it an attractive option for enabling sustainable strategic planning for oil prices. Although deep learning has introduced powerful approaches for improving prediction operations, each method focuses on extracting specific types of characteristics and possesses its own constraints.

Due to the data disparities and the necessity to establish a flexible forecasting model that can analyze and cope with the varied datasets

while still giving high accuracy, several researchers have begun adopting and developing hybrid models. The hybrid model is the model that incorporates many models; its data undergoes a variety of approaches and processes to extract more features and provide better accuracy than the non-hybrid (single) models [36]. In [18], the authors demonstrated the efficacy of time series, machine learning, and hybrid approaches for forecasting Korea's peak load dataset. They integrated SARIMAX output with SVR, long short-term memory (LSTM), and ANN techniques to create hybrid models. They found that the hybrid models outperformed the other methods because the hybrid model performed better by merging the time series and machine learning models into a single model. The feasibility and applicability of energy price prediction were demonstrated in [37] by using the convolutional neural network (CNN) with batch normalization (BN) and LSTM to anticipate day-ahead electricity prices for the electric power markets' (PJM) regulation zone preliminary billing data. The authors in [38] created an adaptive hybrid model based on variational mode decomposition, SARIMA, self-adaptive particle swarm optimization, and a deep belief network to predict short-term electricity prices; the hybrid model outperformed single models in many electricity markets. Moreover, many papers also concluded that hybrid models provide superior processing and results [39–41]. On the other hand, developing a complex model or combining many different methods in one model to address the complex data resulted in "overfitting," a problem that required additional techniques such as "dropout" and "regularization" to resolve [36,42,43]. In the end, this caused the model to operate more slowly.

1.1. Main paper contributions

Despite the growth of time series forecasting methods, an appropriate forecasting process is still required to deal with abrupt changes in data sequence caused by natural, climatic, or political developments in the energy sector. This study provides a novel way to process time series datasets. The input data goes through two stages. The first stage uses a new technique, called enhancing transformation reduction (ETR). The ETR technique minimizes the data's complexity by eliminating seasonality and trends from the input data and boosting its stationarity. The ETR method was utilized for the first time in [44] to improve the data distribution of hyperspectral images. The output of the ETR is sent into the second stage, a hybrid forecasting model that contains LSTM and ANN models, called "transferring LSTM into ANN" (TLIA). In many instances in DL, using deep networks or combining different subnetworks does not improve forecasting because these end-to-end approaches adjust the total produced weights in each training epoch. In the TLIA, between the LSTM and ANN layers, transfer learning techniques (TL) are employed to prevent recalculating the entire set of weights. The TLIA model processes and forecasts the time series data for the ETR method's output. In order to evaluate the efficacy of the presented methodologies, the ETR method is fed six distinct energy market time series datasets. The six datasets have distinct data distributions, sizes, and degrees of complexity. Briefly, the main contribution of this work can be recapped as follows:

- We present a new forecasting model based on advanced deep learning techniques for the volatile energy market.
- An efficient preprocessing method, enhancing transformation reduction (ETR), was developed. The ETR can minimize the complexity of the input dataset due to its high capability of eliminating seasonality and trends from the input data and boosting its stationarity. Thus, the enhanced ETR can smooth the preprocessing stage of the proposed TLIA forecasting model.
- We developed a new forecasting model called TLIA, which combines the processing and characteristics of the LSTM and ANN models using the hybrid concept. This combination helps extract more features and provides more flexibility to deal with any complex datasets.
- The TLIA model incorporated the concept of transfer learning between the two utilized models. The transfer learning layer functions as a valve, allowing processing to move in only one direction (forward) and preventing processing from returning to the LSTM's layers. Utilizing the transfer learning layer expedites the procedure and improves the results.
- Comprehensive and extensive evaluation experiments were carried out using six different energy datasets with comparisons to the state-of-art deep learning methods.

2. Methods

This section describes in detail the study's two main methods: the enhancing transformation reduction method and the transferring of the LSTM into an ANN model.

2.1. Enhancing transformation reduction method

In this study, all the datasets are unsupervised and include only one column (univariate datasets). The sequence input data is divided into multiple input patterns called samples, where 60-time steps (s) are used as input, and a one-time step is used as output for the one-step prediction being learned. Thus, during the time (t) for the input data (x), the (x_{t+1}) is the prediction of (x_{t-1}); it can be presented as follows:

$$X = \left\{ \left(x_{1,1}, x_{1,2}, x_{1,3}, \cdots, x_{1,s}\right), \cdots, \left(x_{t,1}, x_{t,2}, x_{t,3}, \cdots, x_{t,s}\right) \right\} \text{and} f(X) = x_{s+1} \quad (1)$$

where s is the time step for the input data, which is equal to 60 in this study (s = 60), and the predicted step is one. (s) represents the time steps (t), $xands \in Randt \in T$.

Enhancing transformation reduction (ETR) is a new feature extraction method. It is used for many purposes: to enhance the data distribution, reduce outliers and noise, increase the classes' variation, and speed up the processing, according to [44]. This study employs the ETR to improve feature extraction for the time series forecasting model by improving data distribution, increasing the data stationary, and decreasing data complexity. In the ETR method, the data goes through two parts: enhancing the features and transformation. The first part scales and highlights the informative data, and the second part reduces the noise and removes any seasonality from the data.

2.1.1. Enhancing the features

The first part is to enhance the data scaling and highlight the informative values, which will help smooth the training, increase the data stationary, and speed up the training.

The input data (x) has the d and h dimensions. The covariance is taken firstly for $x_{(t,s)}$ to enhance the variation between the time series values to be unique for recognition, representing features after creating the sequences features.

$$C_x = \left(\sum_{i=1}^{s} (\mathscr{X}_i - \overline{\mathscr{X}})(\mathscr{X}_i - \overline{\mathscr{X}})^T \right) \Big/ s - 1 \quad (2)$$

where $\overline{\mathscr{X}}$ is the mean of $\mathscr{X}$. The dimension of C is $(s \times s)$.

As is known, the serial values in time series datasets are typically quite near. Hence it is important to increase the variation and correlation values of the features. By removing the coefficient correlation matrix of $x_{(t,s)}$ from the covariance matrix $C_{(s,s)}$, the following equation solved the issue.

$$\widehat{C}_x = C_x - \varepsilon_x, \quad (3)$$

the range values of ε is between 1 and −1.

Next, the eigenvectors of $\widehat{C}$ are used to generate the weight matrix (W) with a specified size s, which is then multiplied by X, the primary data, to obtain the new distribution without trend.

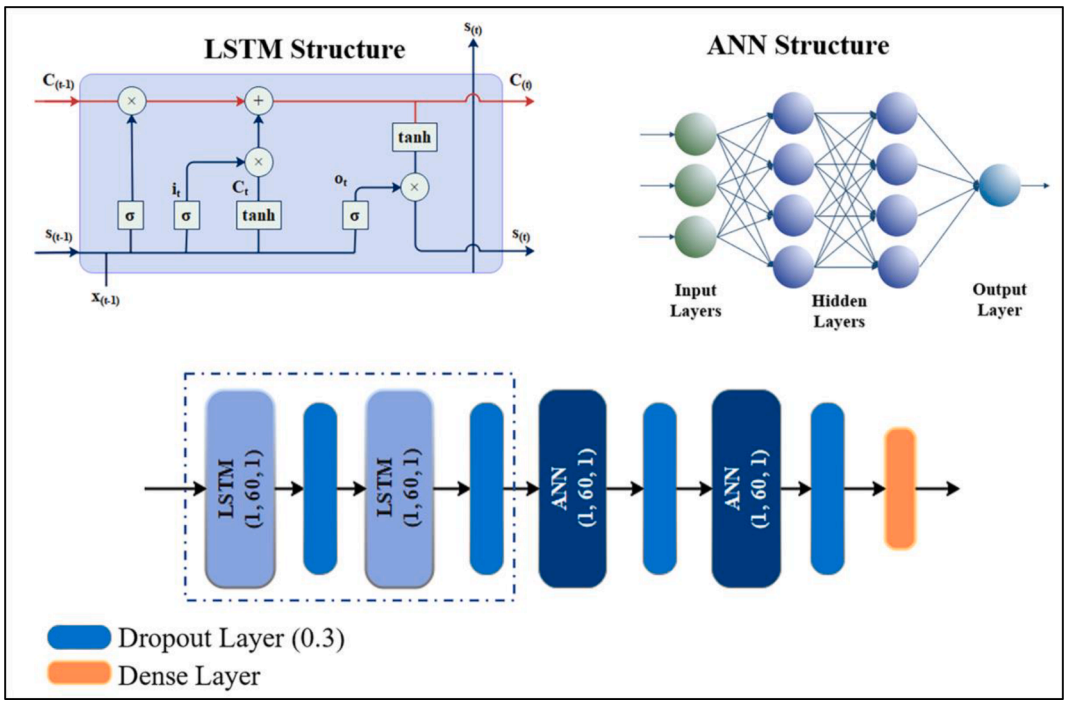


Fig. 1. The structure of the TLIA model.

$$\mathscr{T}_{t\times s} = \mathscr{X}_{t\times s}\times(\mathscr{W}_{s\times 1})^T \tag{4}$$

The new data distribution ($\mathscr{T}_{t\times s}$) is fed into the second part of the ETR method to remove any seasonality and reduce the gap between the serial values and the noise and outliers.

2.1.2. Enhancing the transformation

This part works to reduce the noise and change the position of the values by passing the newly transferred dataset ($\mathscr{T}$) through the morphological dilation method (MD). The MD method works to highlight small values and limit large values, and thus this will help reduce the seasonality phenomenon.

The MD operation depends on two matrixes, mask (K) and marker (R). The mask matrix is already generated from the first part of the ETR, $K_{t\times s} = \mathscr{T}_{t\times s}$. The marker matrix will also be generated from ($\mathscr{T}_{t\times s}$) and the values of the R matrix must be less or equal ($\mathscr{T}$) values. So, firstly to create the R matrix, the maximum values of each instance in ($\mathscr{T}_{t\times s}$) is obtained:

$$\mathscr{M}_t = argmax(\mathscr{T}_t)\times \tilde{\mathrm{A}}^{\circ} \tag{5}$$

where $\tilde{\mathrm{A}}^{\circ}$ is a constant value, it should be very small between zero and one. $\tilde{\mathrm{A}}^{\circ}$ helps to reduce the gap between values as much as possible, which will reduce extremist values.

The $\mathscr{M}_t$ vector is converted into a matrix by repeating this vector s times, $\mathscr{M}_{t\times s}$. Now, the two matrixes (K and R) are used in the MD process to obtain the new unseasonal matrix.

$$D_K^{\delta}(\mathscr{R}) = \delta_K^n(\mathscr{R}) = \delta_K^1\left(\delta_K^{n-1}(\mathscr{R})\right) \wedge K \tag{6}$$

where $K \geq \mathscr{R}$, $n \geq 1$, δ denotes the MD process, and $\wedge$ indicates the pointwise minimum.

The mask matrix controls the new matrix generated by the MD method. The peak values in the R image act as seed values that spread out to fill in the K image, so the distance between the large and small values will be smaller. Thus, the MD reduced the outliers and the skewed distribution.

The final step in the ETR is normalizing the output of the MD method using the Gaussian distribution. So, the second part of the ETR is to remove any seasonality and make the data more stationary.

2.2. Transferring the LSTM into ANN model

This study's forecasting model is a hybrid model that uses the TL method to extract the features of the input data from two different types of layers: LSTM and ANN. The TL transfers the output of the LSTM into the ANN layers and prevents backpropagation modification in the LSTM layers. This section describes the LSTM and ANN layers and the TLIA model.

2.2.1. LSTM

LSTM-based modeling methods are widely used nowadays in various applications, including energy [45,46], medicine [47,48], pollution [49,50], and business [51,52]. Because the recurrent neural network (RNN) has a limited memory, which reduces its learning performance, the LSTM was developed to compensate for this shortcoming. By augmenting the usual RNN's hidden state with a cell state, the LSTM can handle long-term dependencies between features. In LSTM, three gates are available: forget (f), input (i), and output(o). The following are the formulations of the LSTM structure:

$$f_t = \sigma\left(W_f.h_{t-1} + x_t.U_f + b_f\right) \tag{7}$$

$$i_t = \sigma(W_i.h_{t-1} + x_t.U_i + b_i) \tag{8}$$

$$\widetilde{C}_t = tanh(W_c.h_{t-1} + x_t.U_c + b_c) \tag{9}$$

$$C_t = (C_{t-1}\otimes f_t)\oplus\left(i_t\otimes\widetilde{C}_t\right) \tag{10}$$

$$o_t = \sigma(W_o.h_{t-1} + x_t.U_o + b_o) \tag{11}$$

$$h_t = o_t\otimes tanh(C_t) \tag{12}$$

where x represents the input data at a time (t), W exemplifies trainable weight, candidate values ($\widetilde{C}$), C is to present the cell state, and h represents the hidden state. Each gate has its own recurrent weight (W) and

Table 1
The description summary for the six datasets used in the study.

#	Dataset	Record	Date: dd/mm/yyyy	Collection	Mean	Std	Min	Max
1	**Demand**	2106	01/01/2015–06/10/2020	Daily	120035.48	13747.99	85094.38	170653.84
2	**RRP**	2106	01/01/2015–06/10/2020	Daily	76.08	130.25	−6.08	4549.65
3	**Solar Exposure**	2105	01/01/2015–06/10/2020	Daily	14.74	7.95	0.70	33.30
4	**Elia Grid Load**	277,396	01/01/2015–29/11/2022	Daily, each 15-minute	8383.97	1410.18	2494.10	12869.52
5	**Oil Price**	475	30/03/1983–01/09/2022	Monthly	44.72	28.79	10.42	140.00
6	**Wind Power**	1536	03/12/2022–10/12/2022	Daily, two times each 15 min	75.66	78.97	2.50	448.60

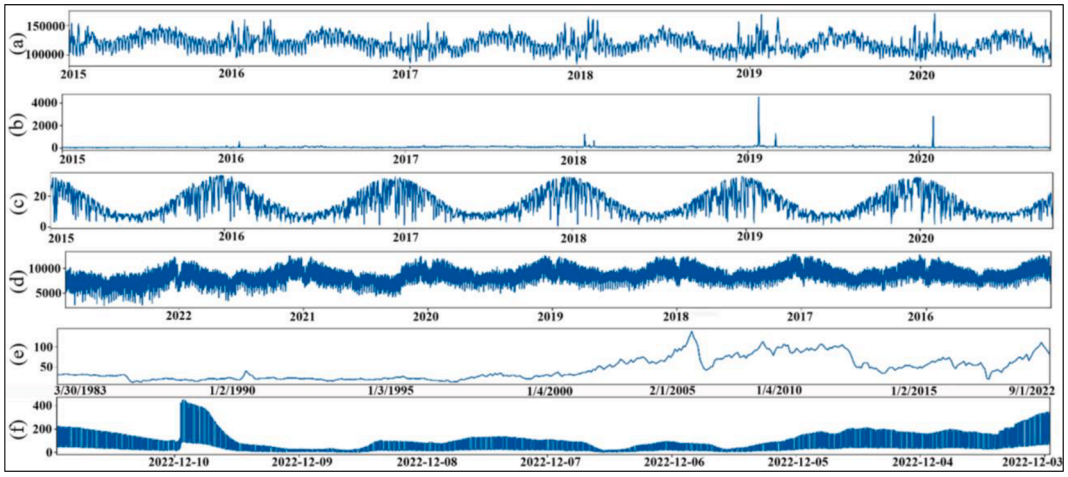


Fig. 2. The data distribution of the six datasets: (a) Demand, (b) RRP, (c) Solar Exposure, (d) Elia Grid Load, (e) Oil Price, and (f) Wind Power dataset.

input weight (U). Moreover, σ stands for the sigmoid function, which can be represented as:

$$\sigma(x) = 1/(1+\exp(-x)) \tag{13}$$

tanh is the activation function, and it is represented as follows:

$$\tanh(x) = (\exp(x) - \exp(-x))/(\exp(x) + \exp(-x)) \tag{14}$$

2.2.2. ANN

The ANN model resembles the neuron-based nervous system of the brain. Like the natural nervous system, these neurons can learn and make judgments. The three layers of this model are input, hidden, and output. Each of these layers has its unique function. The input layer is responsible for receiving simulation-required data. The hidden layer is in charge of data processing and simulation. The output layer offers the ultimate outcome. The general structure of the ANN model is depicted in Fig. 1. Multilayer perceptron learns by adding hidden layers and a backpropagation technique between the input and output layers. The output value of each neuron in the hidden layer is determined by the following formula, which is based on the sum of all input data weights (ω) and the node activation function (Φ):

$$\widehat{y}_j = \Phi\left(\sum_{i=1}^{i} \omega_{ij} x_i + b_i\right) \tag{15}$$

where x is the input data into the ANN node, i is the number of the input data, j is the layer number, and b represents the bias. The activation function (Φ) is ReLu used:

$$\Phi = \max(0, x) \tag{16}$$

The transfer LSTM into ANN (TLTA) model consists of two LSTM layers, two ANN layers, a dropout layer between each LSTM and ANN layer, and a dense output layer. ReLu is the activation function in ANN layers, Fig. 1. In DL techniques, the processing goes in forward and backward paths; "forward propagation" and "backward propagation/backpropagation." During forward propagation, the input data travels from the first to the last layer, generating the weights and prediction values. The loss function evaluates the forward propagation operation's predictions. The loss function in the TLIA model is the mean squared error:

$$MSE = \left(\sum \left(y_j - \widehat{y}_j\right)^2\right)\Big/ n \tag{17}$$

where n is the number of observations, and y is the actual value.

The number of neurons in the TLIA model equals the generated time step (60) from the input data, as seen in Fig. 1 and the 2.1 section. The task of backpropagation is to improve the previously created values of weight and threshold based on the loss function's findings. These two orientations are continued until the highest degree of accuracy is attained. In the TLIA model, the backward propagation procedure is halted in order to change the LSTM layers' output, and it only works in the ANN layers. Training and produced weights in LSTM are propagated into ANN layers without update or modification during training. The ANN layers undergo both forward and backward training. Consequently, the produced weights of ANN layers are modified throughout training.

3. Experiments

This study evaluated the proposed methods, the "enhancing transformation reduction (ETR)" method and the "transferring long short-term memory (LSTM) into an artificial neural network (ANN) (TLIA)" model, using six distinct datasets. In addition, the TLIA model was compared to many well-known forecasting models. Therefore, this section describes the datasets used. In addition, it covers the operation of the ETR method with the TLIA model and exhibits its performance in several ways.

3.1. Datasets

This study used six different datasets with different data distributions

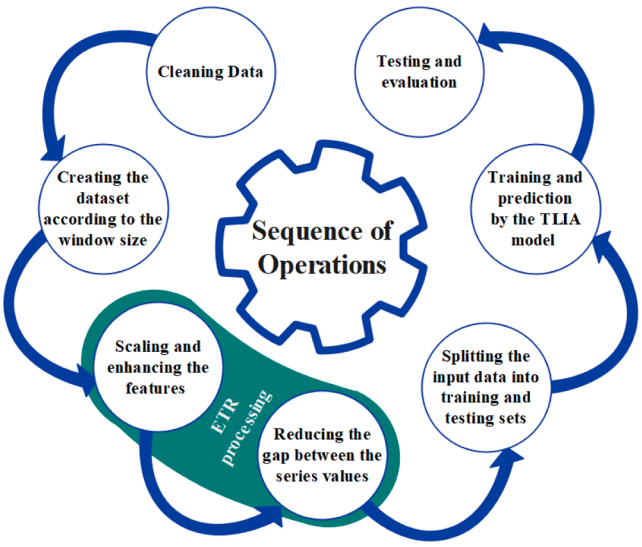


Fig. 3. The processing steps of the study.

underlying commodity and one of the most important global oil benchmarks.

- **Elia Grid Load Dataset:** Historical electric load data from the Belgian company Elia Grid; it is open data from the OpenDataElia website[3]. It is a daily collection of the load every 15 min.
- **Wind power production Dataset:** It is also from the OpenDataElia website, which was collected two times each 15 min for the Flanders region, Belgium.

All datasets are float numerical data, each with different data distributions, as seen in Table 1 and Fig. 2. The summary description of these datasets is in Table 1.

Each dataset was split into 80% for training and 20% for testing. The number of epochs was five, and the batch size was one. The value of the Ã° constant in the ETR was 0.3 for the datasets.

3.2. Evaluation

Mean absolute error (MAE), root mean square error (RMSE), and root mean square percentage error (RMSPE) were applied to assess the model efficiency. The average error magnitude of the model is known as MAE,

Table 2
The results of EDF for the whole six datasets used.

Evaluation	Dataset	Demand	RRP	Solar Exposure	Elia Grid Load	Oil Price	Wind Power
ADF Statistic:	**Main**	−3.953	−11.040	−2.515	−10.646	−2.230	−2.142
	ETR	**−13.479**	**−14.624**	**−17.824**	**−114.046**	**−10.754**	**−9.672**
P-value:	**Main**	0.002	**0**	0.112	**0**	0.195	0.228
	ETR	**0**	**0**	**0**	**0**	**0**	**0**
Critical Values:	**1%:**	−3.433	−3.433	−3.433	−3.430	−3.444	−3.435
	5%:	−2.863	−2.863	−2.863	−2.862	−2.868	−2.863
	10%:	−2.568	−2.568	−2.568	−2.567	−2.570	−2.568

Table 3
The EDF results for a pried time for each dataset.

Evaluation	Dataset	Demand	RRP	Solar Exposure	Elia Grid Load	Oil Price	Wind Power
ADF Statistic:	**Main**	−2.113	−2.770	−0.396	−1.422	−0.933	−1.221
	ETR	**−6.285**	**−6.934**	**−10.288**	**−3.393**	**−3.723**	**−7.462**
P-value:	**Main**	0.239	0.063	0.911	0.571	0.777	0.664
	ETR	**0**	**0**	**0**	**0.011**	**0.004**	**0**
Critical Values:	**1%:**	−3.449	−3.449	−3.449	−3.534	−4.332	−3.468
	5%:	−2.870	−2.870	−2.870	−2.906	−3.233	−2.878
	10%:	−2.571	−2.571	−2.571	−2.591	−2.749	−2.576

and perioded of collection to test and evaluate the ability of the TLIA model:

- **The first dataset is Victoria Daily Electricity:** It is a collected daily electricity dataset for the second-largest state in Australia, Victoria, from the beginning of 2015 until 2020[1]. It has 13 columns beside the date column, and this study uses the following columns:
  - o **Demand:** It is the daily total electricity demand in megawatt-hours.
  - o **RRP:** a suggested retail price is AUD$ per megawatt-hour.
  - o **Solar exposure:** It is a daily total solar energy unit in MJ/m^2.
- **Oil Price Dataset:** It provides monthly information for the West Texas Intermediate (WTI) crude oil price[2] (USD/Bbl). WTI crude oil is regarded as the New York Mercantile Exchange (NYMEX) 's

and its equation is as follows:

$$MAE = 1 \Big/ n \sum_{i=1}^{n} |y_i - \hat{y}_i| \tag{18}$$

where y_i represents the real value, $\hat{y}_i$ represents the prediction value, and n indicates the number of samples.

RMSE expresses the variations between the forecast and actual values. It calculates the size of the residuals and estimates their distribution. RMSE is determined as below:

$$RMSE = \sqrt{1 \Big/ n \sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \tag{19}$$

RMSPE measures the results compared to the actual ones, giving the error percentage form. It is calculated as follows:

[1] https://www.kaggle.com/datasets/aramacus/electricity-demand-in-victoria-australia.

[2] https://www.kaggle.com/datasets/sc231997/crude-oil-price.

[3] https://opendata.elia.be.

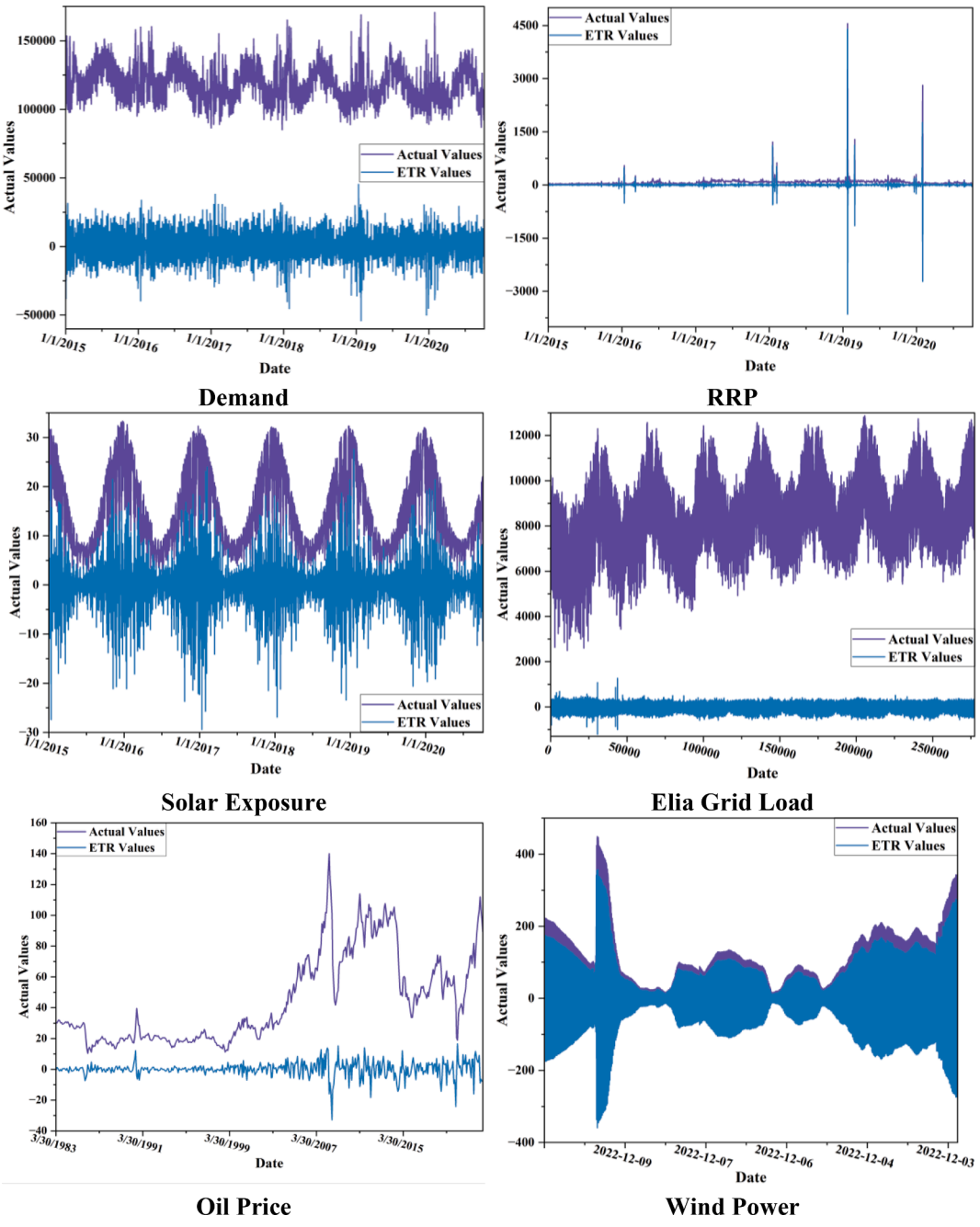


Fig. 4. The actual and ETR values for the six datasets.

Table 4
The RMSE results for the forecasting model with and without transferring.

MAE	LSTM-ANN	TLIA
Demand	0.036	**0.027**
RRP	0.019	**0.018**
Solar Exposure	0.015	**0.011**
Elia Grid Load	**0.068**	**0.068**
Oil Price	**0.066**	0.088
Wind Power	**0.008**	**0.008**

Table 5
The training and testing times for the LSTM-ANN and TLIA models are in seconds.

Dataset	LSTM-ANN (s)		TLIA (s)	
	Training T.	Testing T.	Training T.	Testing T.
Demand	126.207	**19.374**	**40.231**	19.568
RRP	126.094	**18.519**	**36.158**	21.489
Solar Exposure	124.720	**19.437**	**34.408**	19.484
Elia Grid Load	20806.077	**2893.82**	**11543.739**	3116.245
Oil Price	31.400	**4.174**	**12.362**	4.813
Wind Power	91.018	**13.220**	**31.140**	15.772

Table 6
The RMSE values for all datasets.

Dataset	CNN-LSTM	LSTM	GRU	RNN	2LSTM	2GRU	ANN	MM-TLIA	TLIA
Demand	5940.126	4162.758	3241.593	1711.042	23.260	1860.969	2597.872	50.982	**0.027**
RRP	50.227	0.810	8.380	7.915	2.858	10.190	0.250	0.247	**0.018**
Solar Exposure	4.793	3.504	3.829	1.676	0.026	0.026	1.004	0.022	**0.011**
Elia Grid Load	6.993	72.831	72.290	44.473	44.804	43.904	6.993	6.993	**0.068**
Oil Price	1.025	0.545	0.991	0.865	0.463	1.249	0.280	0.127	**0.088**
Wind Power	112.991	106.476	111.619	16.739	1.260	1.903	25.523	0.211	**0.008**

Table 7
The MAE results for the six datasets.

Dataset	CNN-LSTM	LSTM	GRU	RNN	2LSTM	2GRU	ANN	MM-TLIA	TLIA
Demand	4318.544	3286.544	2685.226	1290.880	16.826	1516.840	2019.962	50.982	**0.026**
RRP	22.652	0.452	1.705	7.698	0.875	2.959	0.250	0.247	**0.018**
Solar Exposure	3.222	2.574	2.858	1.272	0.026	0.025	0.704	0.022	**0.010**
Elia Grid Load	6.993	56.584	56.662	36.627	35.335	35.062	6.993	6.993	**0.068**
Oil Price	0.770	0.479	0.809	0.804	0.370	0.942	0.280	0.119	**0.088**
Wind Power	106.763	105.525	109.377	14.300	1.029	1.195	20.380	0.210	**0.008**

Table 8
The RMSPE (%) evaluation for the six used datasets.

Dataset	CNN-LSTM	LSTM	GRU	RNN	2LSTM	2GRU	ANN	MM-TLIA	TLIA
Demand	5.225	3.896	2.900	1.375	0.022	1.666	2.451	0.045	**0**
RRP	100.732	2.169	11.857	36.399	12.723	7.019	1.116	1.105	**0.082**
Solar Exposure	88.291	69.432	74.393	28.934	0.274	0.343	19.338	0.241	**0.116**
Elia Grid Load	0.082	0.830	0.827	0.509	0.514	0.500	0.082	0.082	**0.001**
Oil Price	3.080	1.435	2.515	1.473	1.168	3.149	0.562	0.226	**0.176**
Wind Power	186.923	199.238	198.475	30.696	2.340	2.060	58.469	0.409	**0.015**

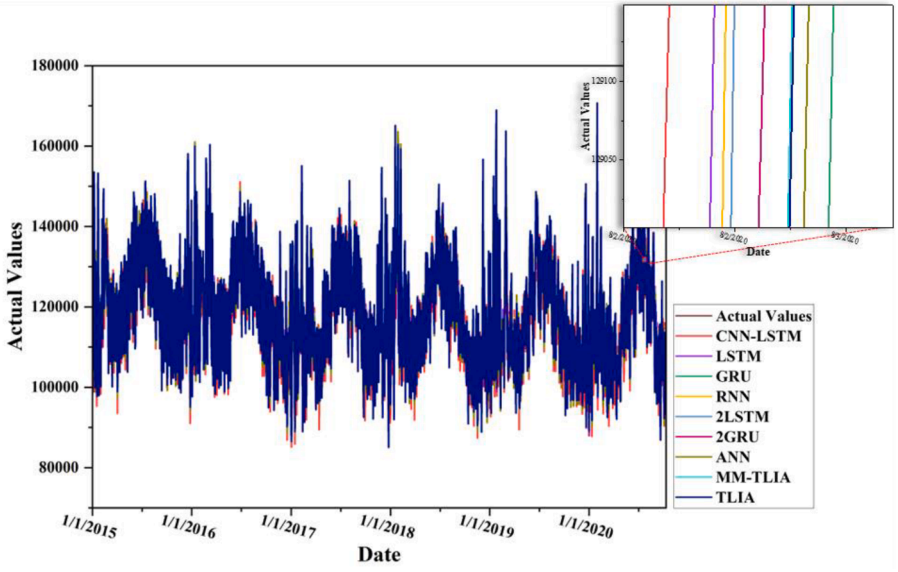


Fig. 5. The results of the nine models for the Demand dataset.

$$\text{RMSPE}\,(\%) = \left(\sqrt{1\Big/n\sum_{i=1}^{n}\left((y_i-\hat{y}_i)/y_i\right)^2}\right) \times 100 \tag{20}$$

3.3. Experiment results

All the used datasets were first cleaned of any null and non-numeric values. Then sixty-time steps (features) were created from each dataset. These sixty newly generated features were fed into the ETR method to enhance the stationarity of the data. The output of the ETR method was then split into training and testing sets. The train set was entered into the TLIA model. Finally, the production of the TLIA model was evaluated by the test set and the different evaluation measurements. Fig. 3 summarizes and depicts all these steps.

3.3.1. Stationariness test

An Augmented Dickey-Fuller Test (ADF) is a statistical test to determine whether the time series data is stationary [53]. Having a stationary dataset helps obtain better forecasting and more accurate predictions than a nonstationary one. First, the entire datasets were

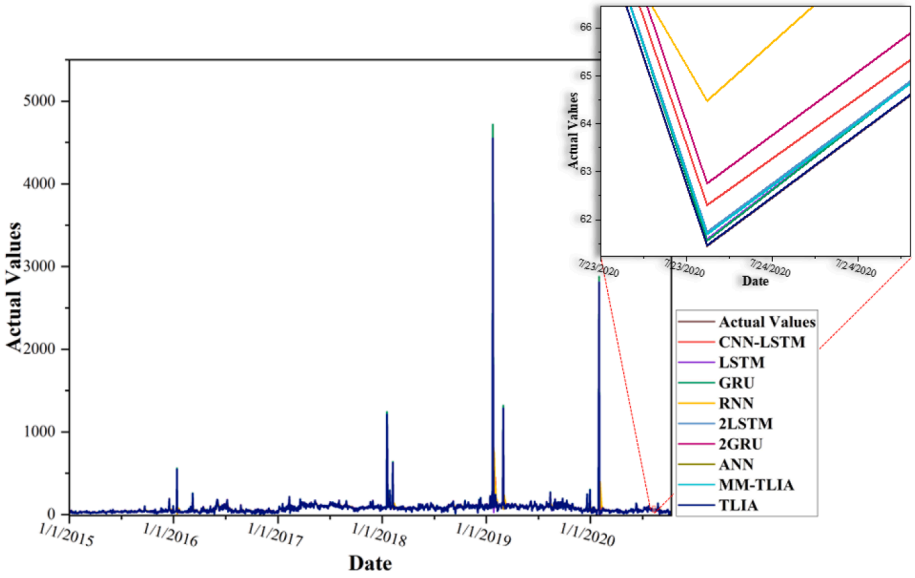


Fig. 6. The results of the nine models for the RRP dataset.

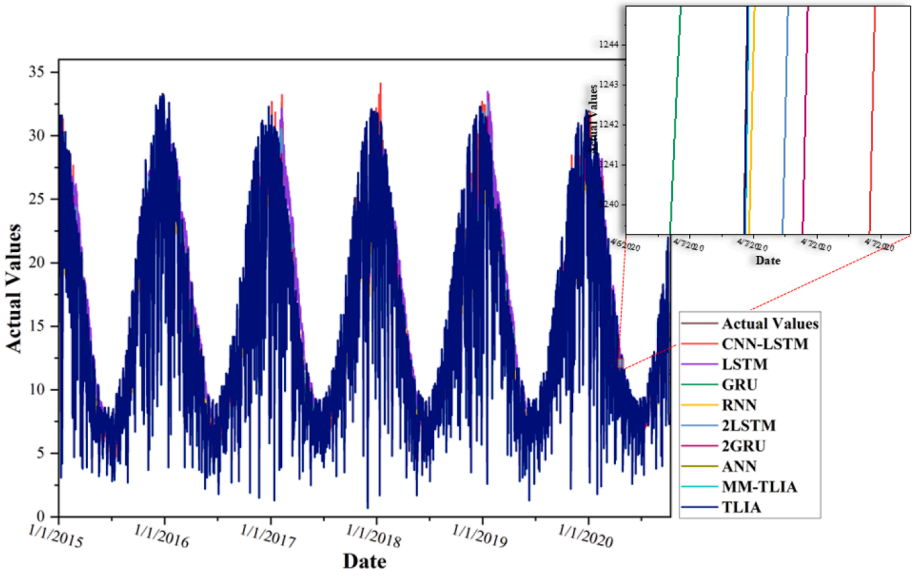


Fig. 7. The results of the nine models for the Solar Exposure dataset.

examined before and after employing the ETR, as shown in Table 2. Second, particular temporal values were taken and evaluated from each dataset before and after applying the ETR, Table 3. Some datasets (Demand, RRP, and Elia Grid Load) are shown to be stationary in Table 2 since the ADF statistical results are greater than the critical levels. In contrast, the output of the ETR has higher stationarities for all datasets; there is a significant difference between the values of the ADF for actual and ETR data.

Table 3 displays the results of the ADF for a certain period: From the first three datasets, the first year (365 days) was examined, one day (69 recordings) was evaluated from the Elia Grid Load dataset, one year (12 records) was examined from the Oil Price dataset, and one day (194 records) was tested from the Wind Power dataset. All these parts exhibit non-stationarity for their real values but were transformed to stationarity using the ETR method. The p-value of the ETR output is less than 0.05, and the ADF statistic for the six datasets is less than the critical levels. All experiments were conducted on Windows 10 with Python running on a GPU with 4 GB of RAM, and the Elia Grid Load dataset was run on a server with a GPU with 26 GB of RAM because it is a huge dataset that requires a large RAM.

Fig. 4 depicts the differences between the actual and ETR series values for all datasets. All the ETR series appear to be centred around zero, with no noticeable trend or seasonality. Importantly, the ETR approach minimized the presence of skewed and outlier data, and all datasets appeared to be stationary. The ETR reduced the distance between the values, which increased the correlation and removed the time dependence between them. It transformed the input data to be more stable and normal, with no time-dependent structure and consistent variation over time.

3.3.2. *Transferring LSTM into ANN model*

The TLIA model was fed the ETR output. The initial layer of the model is an LSTM layer, followed by a dropout layer to reduce overfitting and accelerate processing. The LSTM is distinguished by its long memory, which memorizes lengthy time series and contributes to its high accuracy in forecasting. The output of the initial dropout layer was

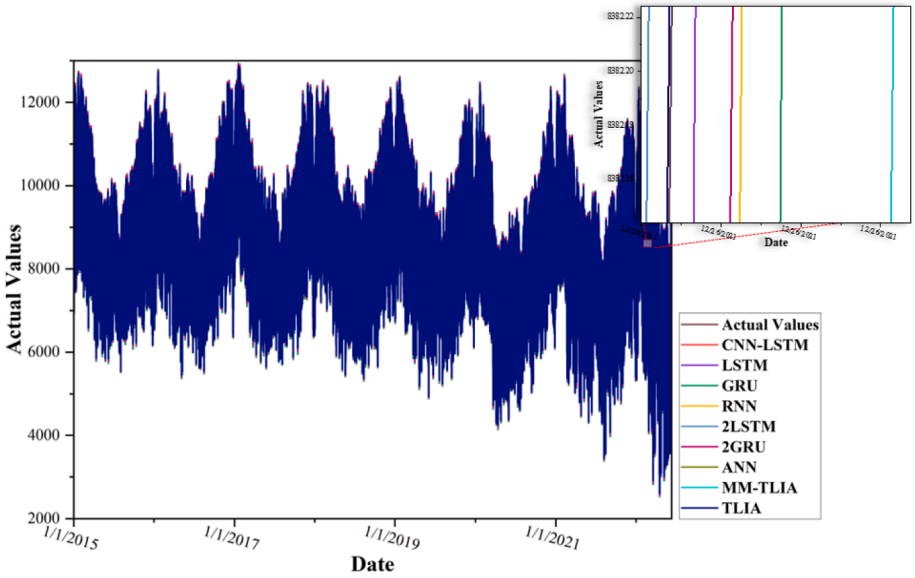


Fig. 8. The results of the nine models for the Elia Grid Load dataset.

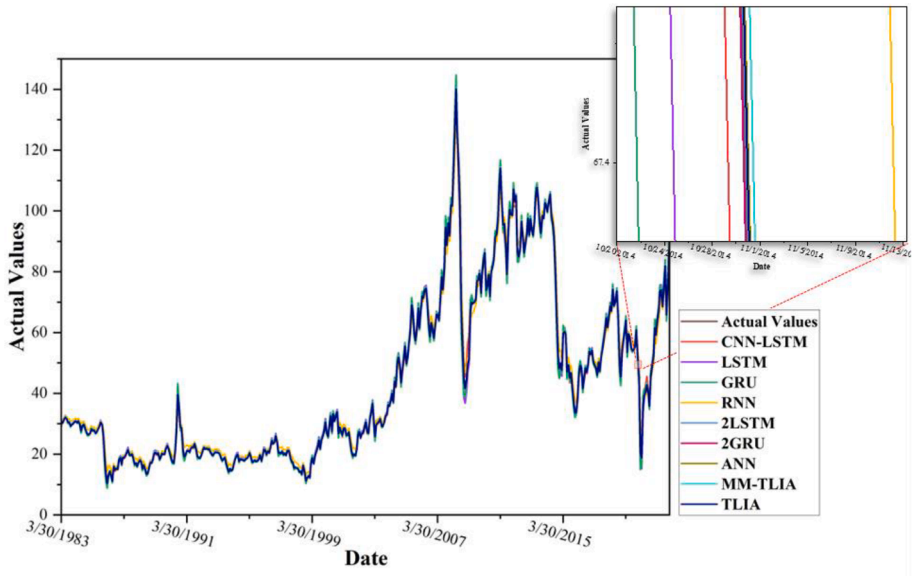


Fig. 9. The results of the nine models for the Oil Price dataset.

passed to the second LSTM layer and a subsequent dropout layer. The weights produced by these layers were transmitted to the first layer of the ANN using the ReLu activation function in order to filter the ANN outputs and send them to the subsequent layers. The layers of ANN strive to extract and enhance the extracted features.

The primary idea of the TLIA model is to stop updating the generated weights of the LSTM layers throughout the training time; the weights were only generated during the forward operation. ANN layers adjust the weights' generation operation in the model according to the model loss function (mean squared error). The LSTM-ANN model was developed to permit backwards updating, whereas the TLIA model does not permit backwards updating in the LSTM layers. The LSTM-ANN model was designed to illustrate the difference between permitting and prohibiting updating the backpropagation (backward) operation in the LSTM layers. Table 4 displays the differences in RMSE values between the LSTM-ANN and TLIA models for each dataset.

The results in Table 4 do not present significant differences between the two models. On the other hand, using the transfer weights from the LSTM to the ANN accelerated the processing by more than 100 s for many datasets and 9 k seconds for the Elia Grid Load dataset, compared to the processing of the LSTM-ANN model without the transfer approach, as shown in Table 5. The trainable parameters in TLIA were reduced from 30 to six. In addition, there was no difference between the test times of the two models. So, the TLIA model saved time and did not waste it updating generated weights in the forward and backward operations during the training time.

3.3.3. *Comparing forecasting models*

To evaluate the performance of the TLIA model, it was compared to seven other forecasting models. Convolutional neural networks with long short-term memory (CNN-LSTM) [54], long short-term memory (LSTM), gated recurrent units (GRU), recurrent neural networks (RNN), 2LSTM, 2GRU, and artificial neural network (ANN) models constitute the seven models.

The CNN-LSTM model comprises two CNN (1D-CNN) layers with 60 filters, a kernel size of 5, and the ReLu activation function. The third

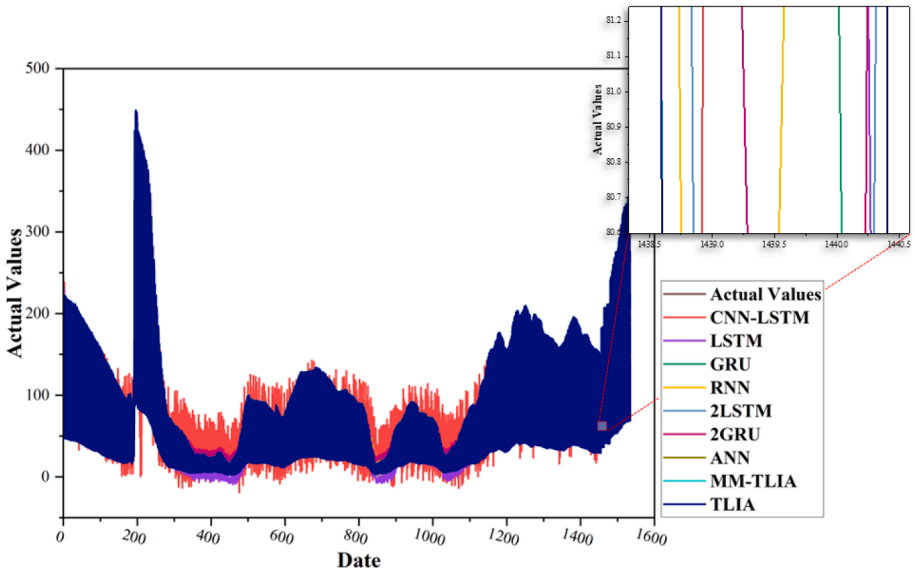


Fig. 10. The results of the nine models for the Wind Power dataset.

Table 9
The training time of the nine models for the six datasets.

Dataset	CNN-LSTM	LSTM	GRU	RNN	2LSTM	2GRU	ANN	MM-TLIA	TLIA
Demand	76.61	60.53	62.23	86.56	118.58	115.17	**8.91**	35.07	40.23
RRP	92.32	57.94	60.86	84.39	118.06	117.11	**8.88**	33.53	36.16
Solar Exposure	82.01	58.99	61.03	92.29	116.04	122.37	**8.91**	34.19	34.41
Elia Grid Load	16237.04	11697.51	11166.37	153716.64	19845.14	18398.57	**6070.65**	12712.91	11543.74
Oil Price	23.39	14.72	15.41	19.66	29.83	28.45	**2.62**	10.28	12.36
Wind Power	68.11	45.54	43.85	70.52	85.92	85.10	**6.64**	25.63	31.14

Table 10
The RMSE evaluation results for the three models (ETR-2LSTM, ETR-ANN, and TLIA) using the ETR method to transfer the input data of the six datasets.

Dataset	ETR-2LSTM	ETR-ANN	TLIA
Demand	**0.028**	**30.016**	**0.027**
RRP	**0.140**	**1.655**	**0.018**
Solar Exposure	**0.112**	**0.844**	**0.011**
Elia Grid Load	**3.814**	**22.995**	**0.068**
Oil Price	**0.089**	**0.199**	**0.088**
Wind Power	**0.233**	**10.465**	**0.008**

layer is composed of sixty LSTM units. Then, two dense layers (FC layers) consisting of 30 and 10 units are followed by the ReLu activation function. The last layer is dense, with a single unit for output and results. The GRU model comprises a GRU layer. Only one LSTM layer is present in the LSTM model. The dense layer for the final results follows the GRU and LSTM layers in the GRU and LSTM models.

The 2GRU model, in comparison, comprises two GRU layers, each followed by a 30% dropout layer. Similar to the LSTM model, the 2LSTM model consists of two layers of LSTM instead of one, along with a 30% dropout layer after each layer, to speed up processing and prevent overfitting. The RNN model consists of two RNN layers, each followed by a 30% dropout layer. Before feeding the data into these models, the MinMaxScaler (MM) method was used to scale the input data to a range between −1 and 1 to accelerate processing.

The three models with the poorest performance were CNN-LSTM, LSTM, and GRU. Perhaps the CNN layers did not supply the right features to the LSTM layers, resulting in a substantially less accurate outcome from the model. On the other hand, the LSTM and GRU models lack sufficient layers to enhance memory and obtain more features, which improve forecasting, particularly for large datasets such as Elia Grid Load, Tables 6-8.

It can be observed from Tables 6-8 that 2LSTM and ANN are the best models for most datasets compared to the other seven models. Their precision differs based on the input dataset. The 2LSTM model is more accurate than the 2GRU model since it has more gates and more sophisticated calculations. On the other hand, the operation of extracting the features from the ANN model is extremely complex. It depends on the number of nodes, activation function, optimizer, and a variety of other techniques that assist in guiding the direction of achieving the highest accuracy in a very short time.

Most of the time, the hybrid models produce better outcomes than the single-method models [6]. The CNN-LSTM, MM-TLIA, and TLIA models are hybrids, but as can be seen in Tables 6-8, the MM-TLIA outperformed the CNN-LSTM, and the TLIA model performed the best. Moreover, as seen in Tables 6-8, the CNN-LSTM model performed worse than non-hybrid models like LSTM and GRU.

The final comparison is between the models MM-TLIA and TLIA. In MM-TLIA, the input data were scaled using the MinMaxScaler (MM) method, but in the TLIA model, the input data were enhanced using the ETR method. The MM-TLIA model was created to evaluate the impact of the ETR on the TLIA model's output. Tables 6-8 show that the ETR aided the TLIA model in producing the most accurate forecasts, outperforming MM-TLIA and the other seven models. The TLIA is the optimal model for all available datasets. In the Demand dataset, the TLIA is more than 50 times superior to the MM-TLIA. Extremely substantial differences exist between TILA and other models' outcomes. Figs. 5-10 illustrate the contrast between the output of the nine models and the actual values of the six datasets. As can be seen, the lines that are close to the real value line are the MM-TLIA and TLIA lines; however, the TLIA model is the closest.

Table 9 indicates how long it took for each model to be trained on each dataset. It shows that the slowest model is the 2LSTM and the

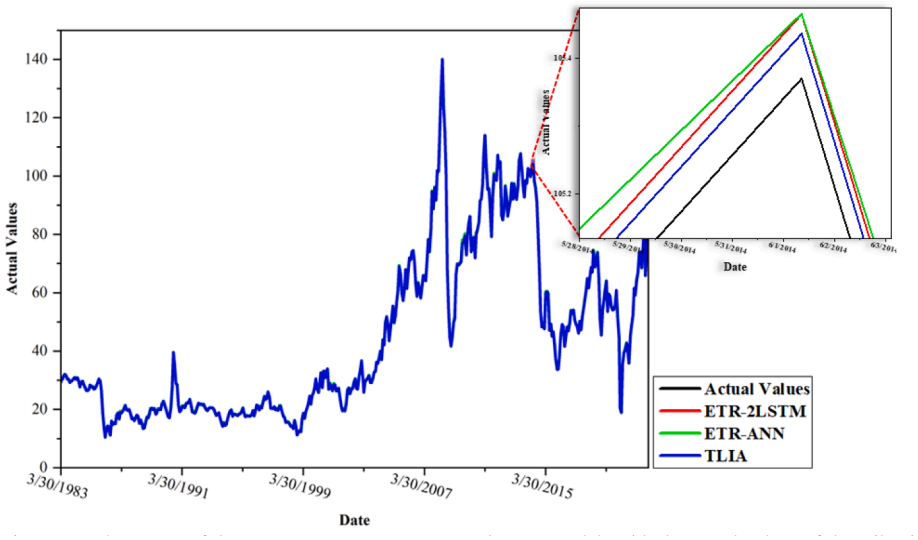


Fig. 11. The output of the ETR-2LSTM, ETR-ANN, and TLIA models with the actual values of the Oil Price dataset.

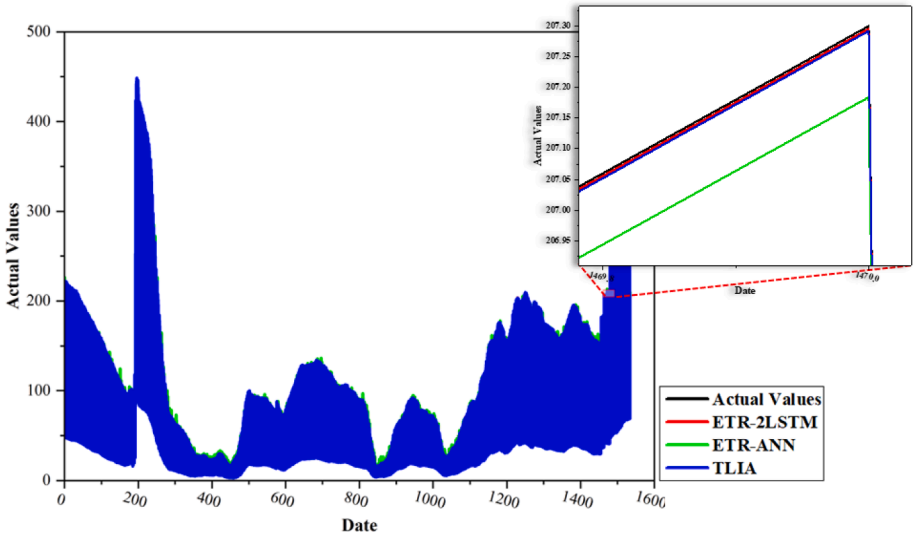


Fig. 12. The output of the ETR-2LSTM, ETR-ANN, and TLIA models with the actual values of the Wind Power dataset.

Table 11
The results of the RMSE for short-term forecasting of the eight models for all datasets and the size of each dataset.

Dataset	Data Set Period	CNN-LSTM	LSTM	GRU	RNN	2LSTM	2GRU	ANN	TLIA
Demand	365 days	8640.761	2312.791	778.680	1085.617	307.907	1452.692	2549.676	**0.017**
RRP	365 days	12.611	1.360	1.727	0.775	0.791	2.737	0.516	**0.014**
Solar Exposure	365 days	6.985	1.700	1.513	0.566	0.185	1.053	0.377	**0.003**
Elia Grid Load	A day (96 records)	67.846	45.225	33.026	101.839	73.563	63.527	24.305	**0.039**
Oil Price	6 years (72 records)	0.763	0.541	0.509	1.173	0.417	0.743	0.474	**0.022**
Wind Power	A day (192 records)	83.932	11.665	33.243	23.472	5.380	14.524	0.984	**0.006**

fastest is the ANN model. As has been explained, the structure of the TLIA models contains the 2LSTM and ANN. Combining the calculations of the LSTM and ANN, the TLIA model obtains the benefits of these two methods. Moreover, freezing the backpropagation procedure in the LSTM layers during training helps extract the input data's features more thoroughly and in less time. In addition to enabling the TLIA model to achieve the highest level of accuracy, the ETR technique accelerated the model by 1169 s compared to the MM-TLIA model using the Elia Grid Load dataset, and the TLIA is the third fastest model with the other datasets.

As seen in Table 10, the best performance was ordered as black, green, and blue are the first, second, and third. The ETR method made the extraction of the 2LSTM and ANN models much better than the results in Table 6. The 2LSTM model outperformed the ANN model on most datasets using the ETR technique. This means that the ANN layers played a significant role in boosting the final accuracy of the TLIA model depending on the transferred weights from the LSTM layers, input data, and the evaluation of the loss function. The TLIA model maintained the excellent performance of the LSTM layers by preventing backpropagation modification in the LSTM layers and limiting it to the ANN

layers.

Furthermore, the processing of ANN and 2LSTM models that used the ETR took longer than those that used the MM but provided higher accuracy (Tables 6 and 10). After increasing the stationarity of the input data, the hybrid model with the transfer learning technique achieved the greatest performance across all datasets. Figs. 11 and 12 show the output of the models in Table 10 for two datasets, and the closest model to the actual data is the TLIA.

3.3.4. Short-Term forecasting

As shown in Table 11, eight models were trained with a tiny portion of the dataset to determine how they would perform with short-term series data. Each dataset has a varied size, except the Demand, RRP, and Solar Exposure databases have the same size. Additionally, each dataset was allocated 80% for training and 20% for testing. As shown by the differences between Tables 6 and 11, the majority of the models performed worse than training the entire dataset. In contrast, the ETR model provided the most accurate results for all datasets, and its results with short-term data are more accurate than those in Table 6, which were for the whole data. The ETR method helped the TLIA model be the best with any size or type of dataset.

4. Conclusion

This study provided a suitable model that can deal with sudden fluctuations in the energy market and different kinds of energy datasets with varying distributions of data and complexities. This study proposed a new method to smooth and increase the data correlation to be more stationary using the ETR method. The ETR helped the forecasting model obtain better and more accurate predictions and accelerated its performance. The forecasting model is called " Time-Series Forecasting Model using Long Short-Term Memory (LSTM) integrated with Artificial Neural Networks (ANN)" (TLIA). This model was a hybrid that combined the features of LSTM and ANN and employed the transfer learning technique to stop the modification of the backward propagation of LSTM layers and modify their output. The TLIA model used the output of the ETR method. The TLIA model was trained using six different energy datasets, each with unique data distributions and complexity levels. The findings of the TLIA model with ETR were compared with seven different forecasting models. The results and the comparisons highlighted the following: (1) The ETR method strengthened data correlation, improved data distribution, decreased outliers, and eliminated seasonality and trends from the input dataset. (2) The ADF test was used to evaluate the stationarity of the ETR output and revealed that the ETR method increased the stationarity of all datasets, whether for the entire dataset or a subset. (3) Because the TLIA model used the transfer learning technique, it maintained the best performance of the LSTM and ANN. (4) In several cases, the TL accelerated the training of the TLIA model by more than 100 s, and in the case of the Elia Grid Load dataset, by more than 9000 s. (5) The ETR improved the accuracy of the TLIA model by 50 times more than MinMaxScaler. (6) For the six datasets, the TLIA model with ETR was highly superior to the other seven well-known forecasting models and the most flexible model for all datasets. (7) In addition, it was also superior when using short-term data from each dataset, but most of the other models introduced less accuracy than training the entire dataset. The summary is that the study provided a novel approach to dealing with energy market fluctuations and provided the best forecast by the TLIA model using the ETR output, which was the most accurate and flexible and outperformed others. Future work will be focused on improving the forecasting model. Although the hybrid model performs better than a non-hybrid model, it takes longer to process data than some non-hybrid models. Consequently, designing a fast model with very high accuracy will be the future work's focus.

CRediT authorship contribution statement

Dalal AL-Alimi: Conceptualization, Methodology, Software, Validation, Formal analysis, Formal analysis, Visualization, Writing – original draft. **Ayman Mutahar AlRassas:** Writing – review & editing, Visualization, Resources. **Mohammed A.A. Al-qaness:** Validation, Supervision, Writing – review & editing. **Zhihua Cai:** Formal analysis, Validation, Writing – review & editing, Supervision. **Ahmad O. Aseeri:** Formal analysis, Validation, Writing – review & editing. **Mohamed Abd Elaziz:** Formal analysis, Validation, Writing – review & editing. **Ahmed A. Ewees:** Formal analysis, Validation, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data are public as mentioned in the main text.

Acknowledgements

This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2023/R/1444)

References

[1] Azarpour A, Mohammadzadeh O, Rezaei N, Zendehboudi S. Current status and future prospects of renewable and sustainable energy in North America: Progress and challenges. Energy Convers Manag 2022;269:115945. https://doi.org/10.1016/j.enconman.2022.115945.

[2] Bhandari R, Arce BE, Sessa V, Adamou R. Sustainability Assessment of Electricity Generation in Niger Using a Weighted Multi-Criteria Decision Approach. Sustainability 2021;13:385. https://doi.org/10.3390/su13010385.

[3] AlRassas AM, Al-qaness MAA, Ewees AA, Ren S, Abd Elaziz M, Damaševičius R, et al. Optimized ANFIS Model Using Aquila Optimizer for Oil Production Forecasting. Processes 2021;9(7):1194.

[4] Medeiros SEL, Nilo PF, Silva LP, Santos CAC, Carvalho M, Abrahão R. Influence of climatic variability on the electricity generation potential by renewable sources in the Brazilian semi-arid region. J Arid Environ 2021;184:104331. https://doi.org/10.1016/j.jaridenv.2020.104331.

[5] Hou R, Li S, Wu M, Ren G, Gao W, Khayatnezhad M, et al. Assessing of impact climate parameters on the gap between hydropower supply and electricity demand by RCPs scenarios and optimized ANN by the improved Pathfinder (IPF) algorithm. Energy 2021;237:121621. https://doi.org/10.1016/j.energy.2021.121621.

[6] Heidarpanah M, Hooshyaripor F, Fazeli M. Daily electricity price forecasting using artificial intelligence models in the Iranian electricity market. Energy 2023;263:126011. https://doi.org/10.1016/j.energy.2022.126011.

[7] Afanasyev DO, Fedorova EA. On the impact of outlier filtering on the electricity price forecasting accuracy. Appl Energy 2019;236:196–210. https://doi.org/10.1016/j.apenergy.2018.11.076.

[8] Zhang Y, He M, Wen D, Wang Y. Forecasting crude oil price returns: Can nonlinearity help? Energy 2023;262:125589. https://doi.org/10.1016/j.energy.2022.125589.

[9] Lin Y, Chen K, Zhang X, Tan B, Lu Q. Forecasting crude oil futures prices using BiLSTM-Attention-CNN model with Wavelet transform. Appl Soft Comput 2022;130:109723. https://doi.org/10.1016/j.asoc.2022.109723.

[10] Al-qaness MAA, Elaziz MA, Ewees AA. Oil Consumption Forecasting Using Optimized Adaptive Neuro-Fuzzy Inference System Based on Sine Cosine Algorithm. IEEE Access 2018;6:68394–402. https://doi.org/10.1109/ACCESS.2018.2879965.

[11] Laib O, Khadir MT, Mihaylova L. Toward efficient energy systems based on natural gas consumption prediction with LSTM Recurrent Neural Networks. Energy 2019;177:530–42. https://doi.org/10.1016/j.energy.2019.04.075.

[12] Del Ser J, Casillas-Perez D, Cornejo-Bueno L, Prieto-Godino L, Sanz-Justo J, Casanova-Mateo C, et al. Randomization-based machine learning in renewable energy prediction problems: Critical literature review, new results and perspectives. Appl Soft Comput 2022;118:108526. https://doi.org/10.1016/j.asoc.2022.108526.

[13] Al-qaness MAA, Ewees AA, Abd Elaziz MA, Samak AH. Wind Power Forecasting Using Optimized Dendritic Neural Model Based on Seagull Optimization Algorithm and Aquila Optimizer. Energies 2022;15:9261. https://doi.org/10.3390/en15249261.

[14] Li C, Xiao Z, Xia X, Zou W, Zhang C. A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting. Appl Energy 2018; 215:131–44. https://doi.org/10.1016/j.apenergy.2018.01.094.

[15] Yang Z, Ce L, Lian L. Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods. Appl Energy 2017;190:291–305. https://doi.org/10.1016/j.apenergy.2016.12.130.

[16] Kottath R, Singh P. Influencer buddy optimization: Algorithm and its application to electricity load and price forecasting problem. Energy 2023;263:125641. https://doi.org/10.1016/j.energy.2022.125641.

[17] Xiao X, Mo H, Zhang Y, Shan G. Meta-ANN – A dynamic artificial neural network refined by meta-learning for Short-Term Load Forecasting. Energy 2022;246: 123418. https://doi.org/10.1016/j.energy.2022.123418.

[18] Lee J, Cho Y. National-scale electricity peak load forecasting: Traditional, machine learning, or hybrid model? Energy 2022;239:122366. https://doi.org/10.1016/j.energy.2021.122366.

[19] Haider SA, Sajid M, Sajid H, Uddin E, Ayaz Y. Deep learning and statistical methods for short- and long-term solar irradiance forecasting for Islamabad. Renew Energy 2022;198:51–60. https://doi.org/10.1016/j.renene.2022.07.136.

[20] Zini M, Carcasci C. Machine learning-based monitoring method for the electricity consumption of a healthcare facility in Italy. Energy 2023;262:125576. https://doi.org/10.1016/j.energy.2022.125576.

[21] Sun J, Zhao P, Sun S. A new secondary decomposition-reconstruction-ensemble approach for crude oil price forecasting. Resour Policy 2022;77:102762. https://doi.org/10.1016/j.resourpol.2022.102762.

[22] Zhu B, Ye S, Wang P, He K, Zhang T, Wei Y-M. A novel multiscale nonlinear ensemble leaning paradigm for carbon price forecasting. Energy Econ 2018;70: 143–57. https://doi.org/10.1016/j.eneco.2017.12.030.

[23] Xu H, Wang M, Jiang S, Yang W. Carbon price forecasting with complex network and extreme learning machine. Phys A Stat Mech Its Appl 2020;545:122830. https://doi.org/10.1016/j.physa.2019.122830.

[24] Rudnik K, Hnydiuk-Stefan A, Kucińska-Landwójtowicz A, Mach Ł. Forecasting Day-Ahead Carbon Price by Modelling Its Determinants Using the PCA-Based Approach. Energies 2022;15:8057. https://doi.org/10.3390/en15218057.

[25] AlRassas AM, Al-qaness MAA, Ewees AA, Ren S, Sun R, Pan L, et al. Advance artificial time series forecasting model for oil production using neuro fuzzy-based slime mould algorithm. J Pet Explor Prod Technol 2022;12(2):383–95. https://doi.org/10.1007/s13202-021-01405-w.

[26] Yang W, Sun S, Hao Y, Wang S. A novel machine learning-based electricity price forecasting model based on optimal model selection strategy. Energy 2022;238: 121989. https://doi.org/10.1016/j.energy.2021.121989.

[27] Lago J, De Ridder F, De Schutter B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Appl Energy 2018;221:386–405. https://doi.org/10.1016/j.apenergy.2018.02.069.

[28] M H, E.a. G, Menon VK, K.p. S. NSE Stock Market Prediction Using Deep-Learning Models. Procedia Comput Sci 2018;132:1351–62. https://doi.org/10.1016/j.procs.2018.05.050.

[29] Vijh M, Chandola D, Tikkiwal VA, Kumar A. Stock Closing Price Prediction using Machine Learning Techniques. Procedia Comput Sci 2020;167:599–606. https://doi.org/10.1016/j.procs.2020.03.326.

[30] AL-Alimi D, Al-qaness MAA, Cai Z, Alawamy EA. IDA: Improving distribution analysis for reducing data complexity and dimensionality in hyperspectral images. Pattern Recognit 2023;134:109096. https://doi.org/10.1016/j.patcog.2022.109096.

[31] Al-qaness MAA, Ewees AA, Thanh HV, AlRassas AM, Abd EM. An optimized neuro-fuzzy system using advance nature-inspired Aquila and Salp swarm algorithms for smart predictive residual and solubility carbon trapping efficiency in underground storage formations. J Energy Storage 2022;56:106150. https://doi.org/10.1016/j.est.2022.106150.

[32] Lauriola I, Lavelli A, Aiolli F. An introduction to Deep Learning in Natural Language Processing: Models, techniques, and tools. Neurocomputing 2022;470: 443–56. https://doi.org/10.1016/j.neucom.2021.05.103.

[33] Alalimi A, AlRassas AM, Vo Thanh H, Al-qaness MAA, Pan L, Ashraf U, et al. Developing the efficiency-modeling framework to explore the potential of CO2 storage capacity of S3 reservoir, Tahe oilfield. China Geomech Geophys Geo-Energy Geo-Resources 2022;8(4):128. https://doi.org/10.1007/s40948-022-00434-x.

[34] Chen H, Birkelund Y, Zhang Q. Data-augmented sequential deep learning for wind power forecasting. Energy Convers Manag 2021;248:114790. https://doi.org/10.1016/j.enconman.2021.114790.

[35] Yazici I, Beyca OF, Delen D. Deep-learning-based short-term electricity load forecasting: A real case application. Eng Appl Artif Intell 2022;109:104645. https://doi.org/10.1016/j.engappai.2021.104645.

[36] Bashir T, Haoyong C, Tahir MF, Liqiang Z. Short term electricity load forecasting using hybrid prophet-LSTM model optimized by BPNN. Energy Rep 2022;8: 1678–86. https://doi.org/10.1016/j.egyr.2021.12.067.

[37] Kuo P-H, Huang C-J. An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks. Sustainability 2018:10. https://doi.org/10.3390/su10041280.

[38] Zhang J, Tan Z, Wei Y. An adaptive hybrid model for short term electricity price forecasting. Appl Energy 2020;258:114087. https://doi.org/10.1016/j.apenergy.2019.114087.

[39] Qiao W, Yang Z. Forecast the electricity price of U.S. using a wavelet transform-based hybrid model. Energy 2020;193:116704. https://doi.org/10.1016/j.energy.2019.116704.

[40] Ma Z, Zhong H, Xie L, Xia Q, Kang C. Month ahead average daily electricity price profile forecasting based on a hybrid nonlinear regression and SVM model: an ERCOT case study. J Mod Power Syst Clean Energy 2018;6:281–91. https://doi.org/10.1007/s40565-018-0395-3.

[41] Cerjan M, Petričić A, Delimar M. HIRA Model for Short-Term Electricity Price Forecasting. Energies 2019;12:568. https://doi.org/10.3390/en12030568.

[42] Yang Z, Keung J, Kabir MA, Yu X, Tang Y, Zhang M, et al. AComNN: Attention enhanced Compound Neural Network for financial time-series forecasting with cross-regional features. Appl Soft Comput 2021;111:107649. https://doi.org/10.1016/j.asoc.2021.107649.

[43] AL-Alimi D, Al-qaness MAA, Cai Z, Dahou A, Shao Y, Issaka S. Meta-Learner Hybrid Models to Classify Hyperspectral Images. Remote Sens 2022;14(4):1038. https://doi.org/10.3390/rs14041038.

[44] AL-Alimi D, Cai Z, Al-qaness MAA, Ahmed Alawamy E, Alalimi A. ETR: Enhancing Transformation Reduction for Reducing Dimensionality and Classification Complexity in Hyperspectral Images. Expert Syst Appl 2023;213:118971.

[45] Skrobek D, Krzywanski J, Sosnowski M, Kulakowska A, Zylka A, Grabowska K, et al. Implementation of deep learning methods in prediction of adsorption processes. Adv Eng Softw 2022;173:103190. https://doi.org/10.1016/j.advengsoft.2022.103190.

[46] Skrobek D, Krzywanski J, Sosnowski M, Kulakowska A, Zylka A, Grabowska K, et al. Prediction of Sorption Processes Using the Deep Learning Methods (Long Short-Term Memory). Energies 2020;13(24):6601. https://doi.org/10.3390/en13246601.

[47] Sunjaya BA, Permai SD, Gunawan AAS. Forecasting of Covid-19 positive cases in Indonesia using long short-term memory (LSTM). Procedia Comput Sci 2023;216: 177–85. https://doi.org/10.1016/j.procs.2022.12.125.

[48] Zhou L, Zhao C, Liu N, Yao X, Cheng Z. Improved LSTM-based deep learning model for COVID-19 prediction using optimized approach. Eng Appl Artif Intell 2023;122: 106157. https://doi.org/10.1016/j.engappai.2023.106157.

[49] Bacanin N, Sarac M, Budimirovic N, Zivkovic M, AlZubi AA, Bashir AK. Smart wireless health care system using graph LSTM pollution prediction and dragonfly node localization. Sustain Comput Informatics Syst 2022;35:100711. https://doi.org/10.1016/j.suscom.2022.100711.

[50] Drewil GI, Al-Bahadili RJ. Air pollution prediction using LSTM deep learning and metaheuristics algorithms. Meas Sensors 2022;24:100546. https://doi.org/10.1016/j.measen.2022.100546.

[51] Mao W, Zhu H, Wu H, Lu Y, Wang H. Forecasting and trading credit default swap indices using a deep learning model integrating Merton and LSTMs. Expert Syst Appl 2023;213:119012. https://doi.org/10.1016/j.eswa.2022.119012.

[52] Lin Y, Liao Q, Lin Z, Tan B, Yu Y. A novel hybrid model integrating modified ensemble empirical mode decomposition and LSTM neural network for multi-step precious metal prices prediction. Resour Policy 2022;78:102884. https://doi.org/10.1016/j.resourpol.2022.102884.

[53] Pastor DJ, Ewing BT. Exploding DUCs? Identifying periods of mild explosivity in the time series behavior of drilled but uncompleted wells. Energy 2022;254: 124298. https://doi.org/10.1016/j.energy.2022.124298.

[54] Kim T-Y, Cho S-B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019;182:72–81. https://doi.org/10.1016/j.energy.2019.05.230.