applied sciences

Article

A Hybrid DNN Multilayered LSTM Model for Energy Consumption Prediction

Mona AL-Ghamdi [1,*], Abdullah AL-Malaise AL-Ghamdi [1,2] and Mahmoud Ragab [3,4]

1 Information Systems Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; aalmalaise@kau.edu.sa
2 Information Systems Department, School of Engineering, Computing and Design, Dar Al-Hekma University, Jeddah 22246, Saudi Arabia
3 Information Technology Department, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia; mragab@kau.edu.sa
4 Mathematics Department, Faculty of Science, Al-Azhar University, Naser City, Cairo 11884, Egypt
* Correspondence: monagh5454@gmail.com

Abstract: The ability to predict energy consumption in a world in which energy needs are ever-increasing is important for future growth and development. In recent years, deep learning models have made significant advancements in energy forecasting. In this study, a hybrid deep neural network (DNN) multilayered long short-term memory (LSTM) model was used to predict energy consumption in households. When evaluating the model, the individual household electric power consumption dataset was used to train, validate, and test the model. Preprocessing was applied to the data to minimize any prediction errors. Afterward, the DNN algorithm extracted the spatial features, and the multilayered LSTM model was used for sequential learning. The model showed a highly accurate predictive performance, as the actual consumption trends matched the predictive trends. The coefficient of determination, root-mean-square error, mean absolute error, and mean absolute percentage error were found to be 0.99911, 0.02410, 0.01565, and 0.01826, respectively. A DNN model and LSTM model were also trained to study how much improvement the proposed model would provide. The proposed model showed better performance than the DNN and LSTM models. Moreover, similar to other deep learning models, the proposed model's performance was superior and provided accurate and reliable energy consumption predictions.

Keywords: DNN; LSTM; deep learning; prediction

Citation: AL-Ghamdi, M.; AL-Ghamdi, A.A.-M.; Ragab, M. A Hybrid DNN Multilayered LSTM Model for Energy Consumption Prediction. *Appl. Sci.* **2023**, *13*, 11408. https://doi.org/10.3390/app132011408

Academic Editor: Douglas O'Shaughnessy

Received: 24 August 2023
Revised: 1 October 2023
Accepted: 15 October 2023
Published: 18 October 2023

1. Introduction

Electricity's importance in people's lives, institutions, and companies is undeniable, and social welfare and economic activities are directly related to consumption. Prediction is a suitable method for predicting consumption and use, based on parameters like months, hours, weather, cost, economic conditions, and location [1]. The energy industry's growth has led to an increase in electrical consumption prediction, promoting hourly trading and enhancing market interaction, thereby ensuring stability and reducing waste [2,3], which is especially important due to the development of smart grids in the future [4].

Global energy trends vary across nations due to income levels, altitudes, and energy efficiency measures. Asia and the Pacific region have the highest energy consumption, with China being the largest. Brazil, Russia, India, and China's strong economic performance and advocacy for electrification contribute to increased global energy consumption, reducing greenhouse gas emissions [5].

Simulating energy consumption and determining efficiencies are critical to achieving a low-carbon emission economy [6]. Predicting energy consumption accurately is crucial for energy efficiency design in industries and households, enabling the development of monitoring tools and peak energy regulations for large sectors [7]. In addition, some industries are important in securing public resources and supplies in certain areas. Therefore, predicting energy consumption can maintain household energy distribution in different industries.

To predict energy consumption, a high-performance model is needed, which requires accurate feature selection and a suitable modeling process. Advances in deep learning and artificial intelligence can aid in examining energy consumption phenomena.

Energy consumption prediction studies utilize traditional machine learning and deep learning models, often employing hybrid models to enhance performance. However, complex data, nonlinear relationships, and computational challenges restrict the scope of the study, prompting the need for further development.

In this paper, a deep neural network (DNN) was integrated with a multilayered long short-term memory (LSTM) network to predict energy consumption. The IHEPC data were utilized as a dataset that was cleaned and transferred to the DNN layer. It was then passed to the multilayered LSTM model, which consists of two layers. The second module of the model is the multilayer LSTM. The LSTM layers are used for information processing and time series prediction. Finally, the LSTM layers are connected to the dense layer, which modifies the output dimensionality of the LSTM layers. Therefore, to improve energy consumption prediction, this study makes the following contributions:

- It presents a new integrated deep neural network (DNN) with a multilayered long short-term memory (LSTM) network to predict energy consumption;
- It uses a real-life dataset called individual household electric power consumption (IHEPC) to develop the proposed model, as well as DNN and LSTM models for comparing and checking how much improvement the proposed model could give compared with the DNN and LSTM.

The rest of this study is organized as follows. Section 2 introduces the related studies. Section 3 elaborates on the proposed model, datasets, and experimental design. Section 4 demonstrates model evaluation. The experimental results and discussion are exhibited in Section 5. In Section 6, the conclusion and future work are shown.

2. Related Studies

Load prediction is a promising field, and research has shown its potential in the field through publications. These publications are divided into different classes based on the learning algorithms: persistence, physical, statistical, and artificial intelligence.

Persistence models are applied in predicting future time series trends in electricity consumption or predictions. However, these models failed after several hours and were unable to produce accurate predictive results [8]. Persistence predictive models are not decisive and cannot be recommended for energy consumption prediction. Physical models use mathematical expressions that take into account historical and meteorological data.

Artificial intelligence algorithms can be learned from nonlinear and complex datasets. They can be classified into traditional and deep learning techniques.

2.1. Traditional Models

Traditional algorithms, including artificial neural networks [9], support vector machines [10], extreme learning machines [11], energy-aware clustering schemes [12], and random forest algorithms, did not perform optimally in predictions. Nevertheless, they showed promising performance in feature mining. Therefore, improving the models' predictive performance requires adding more selection and feature extraction algorithms, which remains challenging [13,14]. Artificial learning techniques are further classified into deep learning and machine learning techniques. Several studies have been conducted on energy consumption prediction for machine learning algorithms.

Mohan et al. [15] developed a dynamical–empirical model to predict energy consumption using physical models. The developed models were found to be unreliable in energy prediction based on the high computational space and memory used and needed. Relative to physical learning models, statistical learning models are not as expensive computationally as they apply autoregressive models, such as ARIMA [16], GARCH [17], and linear regression techniques. Learning models were developed using linear datasets. Khodayar et al. [8] used the GARCH model to determine predictive uncertainty, but its ability was limited to capturing nonlinear and nonstationary elements of energy consumption data. Studies involving linear regression have also been conducted. Fumo et al. [18] developed a multiple and linear regression model for energy consumption prediction, while Vu et al. [19] and Braun [20] developed multiregression energy consumption prediction models. However, statistical energy consumption models are not capable of accurately capturing uncertainty trends directly; instead, they employ other methods, as shown in Shi et al.'s [13] study, to minimize prediction errors.

Chen et al. [21] used support vector regression (SVR) to predict energy consumption using temperature and consumption datasets. Cao and Wu [22] applied a hybrid SVR model to forecast seasonal energy consumption. In their work, the SVR algorithm was combined with the fruit fly algorithm to improve the model's performance. Zhong et al. developed an SVR model to predict energy consumption and change the nonlinearity of the features found in the dataset into linear to deal with the nonlinearity problem of the model and the dataset used [23]. Li et al. used a random forest regression algorithm for energy consumption prediction. The researchers extracted the required features based on frequency using fast Fourier transforms and then applied the random forest model in their simulation to make predictions [24].

2.2. Deep Learning Models

Several deep learning models have been developed for energy consumption prediction. Kong et al. used an LSTM model to predict energy consumption [25]. Wang et al. developed a bidirectional LSTM model linked to a rolling update and attention algorithm to make energy consumption forecasts [26]. Shi et al. developed a pooling deep RNN to forecast household energy consumption. In this study, overfitting was an issue in the algorithm [13]. Feature extraction techniques involve various regular patterns, noise, and spectral analysis. However, the algorithms decrease energy consumption accuracy, while the methods decrease accuracy because of the nonlinearity and irregular trends of the electrical energy consumption datasets. The LSTM [25] and CNN [27] algorithms excel in pattern and sequence learning. CNN models often struggle with learning temporal features in energy consumption data. Han et al. proposed a GRU model to predict future energy consumption for short intervals of time, decreasing error rates [28]. To deal with the shortcomings of single algorithms, hybrid models have been developed to effectively predict energy consumption. Ullah et al. [29] used a hybrid bidirectional LSTM and CNN. Kim et al. [30] used a hybrid CNN and LSTM to predict energy consumption. Chi et al. proposed a WT multiple LSTM model to improve power consumption prediction performance [31]. Khan et al. proposed the CNN-BiGRU model to reduce errors in individual household electricity consumption prediction [32]. Moreover, Khan et al. proposed the CNN-ESN model, which generated high prediction accuracy for power consumption forecasting [33]. Alsharekh et al. developed a deep R-CNN with ML-LSTM model for short-term electricity load forecasting and conducted an experiment for daily and hourly power consumption forecasts [34].

Therefore, based on previous studies, a model that could perform better with complex data and excel in sequence learning together, which initiates the idea of building a hybrid DNN multilayered LSTM model, should be developed. Such a model could improve predictive performance, which could be used in energy management. Table 1 presents a summary of the related studies.

Table 1. Summary of the related studies.

Research Objective	Methodology Used	Result	Limitation	Resource
To develop an innovative framework for short-term electricity load forecasting	Deep R-CNN with ML-LSTM	The proposed model resulted in the lowest error rates for the MAE, MSE, RMSE, and MAPE, and the highest R2 for two datasets.	Experiments were carried out for daily and hourly power consumption forecasts. The error rates were significantly reduced. Nevertheless, the proposed model should be tested across medium- and long-term electrical load predictions in future work. Environmental sensor data should be integrated to assist in estimating future electricity consumption.	Alsharekh et al. (2022) [34]
To introduce an efficient and effective hybrid model for power generation and consumption forecasting	CNN-ESN	The proposed model achieved higher prediction accuracy and demanded significantly reduced running time compared with state-of-the-art prediction models.	The model achieved notable results in terms of error rate. However, using other measures, such as the Pearson correlation coefficient or R-squared, could help determine statistical relationships and explain how dependent variables can be explained by predictors or independent variables.	Khan et al. (2022) [33]
To establish a two-step methodology for residential building load prediction	CNN-BiGRU	The proposed model reduced errors in the individual household electricity consumption prediction.	The model was utilized for short-term power consumption prediction, which means predictions of up to several days. Medium- and long-term horizons might need to be tested or generalized using the model.	Khan et al. (2020) [32]
To improve power consumption prediction performance using a wavelet transform-based model	WT multiple LSTM	The prediction performance of the proposed model was better than the traditional LSTM and bidirectional LSTM. Wavelet denoising could further improve the prediction performance of the model.	The model showed remarkable results in predicting daily energy consumption. The proposed model could be generalized to hourly, weekly, or monthly power consumption in future studies.	(Chi et al., 2022) [31]
To propose a CNN-LSTM neural network that can extract spatial and temporal features to effectively predict housing energy consumption	CNN-LSTM	The proposed method achieved an almost perfect prediction performance for electric energy consumption. It generated the smallest value of the root-mean-square error compared with the conventional forecasting methods for the dataset on individual household power consumption.	A potential limitation of the CNN-LSTM model is the relatively substantial efforts by trial and error to establish the best hyperparameters. To overcome this challenge, the search for optimal hyperparameters should be automated. The authors are now working on a genetic algorithm that can automatically search for the CNN-LSTM hyperparameter space.	Kim and Cho (2019) [30]
To present an intelligent hybrid technique that combines a convolutional neural network (CNN) with a multilayer bidirectional long short-term memory (M-BDLSTM)	CNN-M-BDLSTM	The proposed method outperformed the existing methods (Seq2Seq, FCBRM, BPTT, and CNN-LSTM), achieving the smallest value for the MSE.	The proposed method needs to be improved by taking into account different time scales and horizons, such as months, years, and decades. These horizons can be examined for energy management in many industries and smart grids to enable efficient electrical energy utilization.	Ullah et al. (2020) [29]
To propose a deep learning-based framework for intelligent energy management to predict future energy consumption for short time intervals	GRU	The lowest error rate was obtained by the proposed model on two datasets. Evidence showed that it could be used for household and industrial sectors in real-world scenarios.	The authors suggested integrating fuzzy logic with sequential learning, which could help to improve real-time energy forecasting and generalize power consumption prediction on different scales.	Han et al. (2021) [28]

According to related studies, deep learning has been widely used and has presented remarkable results compared to traditional machine learning methods. The CNN has been utilized in various studies to extract and capture informative features, thereby improving prediction results. LSTM has been preferred over the RNN for energy consumption prediction, especially for time series problems. The transformer is one of the newest and most commonly employed techniques for sequence prediction and, therefore, requires further investigation in power consumption prediction studies. Furthermore, most of the introduced models have been used for short-term power consumption prediction, which typically covers predictions up to several days. However, medium- and long-term horizons should be considered for developing generalized models.

2.3. Shortcomings of the Existing Models

As the related works were limited to the challenges of complex data, selecting the most suitable hyperparameters limited the time ranges in the experiments. Moreover, owing to the need to develop new hybrid models due to their better performance over single algorithms, we developed a suitable model that could perform effectively with feature selection and sequential learning.

3. Proposed Model

The methodology of this paper has the following steps:

- Step 1: Introduce the IHEPC dataset, the theoretical background of the proposed model, and the model evaluation measurements.

In this step, a glimpse of the dataset is given, and an integrated DNN with a multi-layered LSTM model is introduced. This obtains trends and time-based evidence from the dataset used. The model evaluation measurements used in the experiment are listed below.

- Step 2: Practical application

This step is divided into three parts:

a. Preprocess the IHEPC dataset, such as resolving the issues of missing values and outliers, variable normalization, and data resampling.

The IHEPC dataset contains several missing values and outliers that can be attributed to metering errors, weather conditions, and even uncharacteristic customer consumption trends. The atypical trends and redundancies in the data result in poor predictive performance and ambiguous predictions. To resolve this issue, the data are preprocessed.

b. Train the desired model, a DNN model or an LSTM model, with the dataset and draw graphs to summarize the results.
c. Compare the results of the proposed model with those of the developed DNN and LSTM models to check for improvements. The proposed model is also compared with other models in related research, such as the CNN multilayered bidirectional LSTM, CNN-BiGRU, GRU, CNN-ESN, R-CNN with ML-LSTM, linear regression models, LSTM, bidirectional LSTM, multiple LSTM, and WT multiple LSTM models.

3.1. Dataset

The dataset has 2,075,259 measurements collected from households in Sceaux, France, over a 47-month time span (December 2006 to November 2010) [35]. The dataset contained several missing measurement values, approximately 1.25% of the row data. All date timestamps were present in the data. However, some values were missing. A dash between the dataset attribute separators represents the missing values. The variables found in the data included minutes, hours, days, months, years, global active power, global reactive power, global intensity, voltage, and sub-metering 1, 2, and 3 [35], as shown in Figure 1.

#	Variable	Description
1	Day	A value from 1 to 31
2	Month	A value from 1 to 12
3	Year	A value from 2006 to 2010
4	Hour	A value from 0 to 23
5	Minute	A value from 1 to 60
6	Global active power	The household global minute-averaged active power (in kilowatt)
7	Global reactive power	The household global minute-averaged reactive power (in kilowatt)
8	Voltage	The minute-averaged voltage (in volt)
9	Global intensity	The household global minute-averaged current intensity (in ampere)
10	Sub metering 1	an oven and a microwave, hot plates being not electric but gas powered (in watt-hour of active energy)
11	Sub metering 2	This variable corresponds to the laundry room, containing a washing machine, a tumble-drier, a refrigerator and a light. (in watt-hour of active energy)
12	Sub metering 3	This variable corresponds to an electric water heater and an air conditioner (in watt-hour of active energy)

Figure 1. IHEPC dataset features.

3.2. The DNN Multilayered LSTM Model

This two-stage model includes data preprocessing and the hybrid DNN multilayered LSTM model. The preprocessing stage involves filling the missing values with the mean values, eliminating any outliers in the dataset, and data normalization for effective training. The data processing stage receives the IHEPC data, which are cleaned and passed onto the DNN layer, as in the first module. They are then transferred to the multilayered LSTM model, which is made up of two layers. The multilayer LSTM is the second module of the model. The purpose of the LSTM layers is information analysis and the prediction of the time series. The LSTM layers are then connected to the dense layer, which changes the output dimensionality of the LSTM layers.

3.2.1. The DNN Model

The DNN layer comprises two hidden layers that are similar to traditional multilayered perceptrons. The layers are made up of a network structure, with each layer containing several nodes. The number of nodes and activation functions found in the output layers are tailored to the energy consumption classification problem. In addition to the output and input layers, there is a hidden layer that extracts any required complex features [36].

The module addresses the problem of matrix factorization, as it can incorporate items and query features based on input layer flexibility. This allows the model to capture the desired output and improve the model's performance. As the DNN classifier is made up of neurons that work using the rectified linear unit (ReLU), the mathematical form of the DNN is expressed as follows: $f(x) = x^+ = \max(0, x)$ [37], where x is the neuron input and RAMP function. The ReLU function is a smooth approximation function of the rectifier that relies on the analytic function, which is referred to as the softplus function:

$$f(x) = \ln[1 + e^x]$$

After prediction, a raw descriptor is generated from the hidden layers, which is expressed as follows:

$$X_{l+1} = H(W_l X_l + B_l), \quad l = 1, 2 \ldots l$$

where H is the activation function, W_l is the weight matrix, B_l is the bias in the hidden layers, and the ReLU function selects the parameters [37]. The output of the DNN module is then fed to the multilayered LSTM model. The output is composed of the weighted inputs from a multilinear transformation.

3.2.2. Multilayered LSTM Model

LSTM was specifically designed for modeling temporal sequences. LSTM is a modified RNN, but it stands out from other RNNs because of its long-range dependencies, which increase its accuracy. Using typical RNN algorithms eliminates the main problem of being unable to determine the significant connection between steps in data over 10. LSTM has special units, which are referred to as memory blocks, found in the recurrent hidden layer. The memory blocks have memory cells that store the temporal state of the network. These memory cells allow LSTM to have long-term memory capabilities that enable the updating of the previous hidden state. Memory cells provide feedback in every neuron. The input data to a memory cell usually contain the cell state, the previous step output, and the input data of the current step. This makes the output of the current cell dependent on the inputs and weights of the current step as well as the inputs of the previous step. This functionality is responsible for the LSTM's ability to understand temporal relationships with long-term sequences. It can create relationships even when the data contain 1000 steps. Each memory block has three gates that manage the cell state information for the LSTM. The forget gate is the first gate, and it is responsible for selecting and removing a part of the cell's memory. The input gate is the second gate, and it is responsible for controlling the input data flow into the cell's memory. The output gate is the last gate, and it is responsible for the output data flow from the cell's memory. The LSTM process is computed through a sequence of Equations (1)–(6) [38].

$$f_t = \sigma\left(x_t W_{xf} + h_{(t-1)} W_{hf} + b_f\right) \tag{1}$$

$$i_t = \sigma\left(x_t W_{xi} + h_{(t-1)} W_{hi} + b_i\right) \tag{2}$$

$$o_t = \sigma\left(x_t W_{xo} + h_{(t-1)} W_{ho} + b_o\right) \tag{3}$$

$$\widetilde{c_t} = tanh\left(x_t W_{xc} + h_{(t-1)} W_{hc} + b_c\right) \tag{4}$$

$$c_t = f_t \odot C_{t-1} + i_t \odot \widetilde{c_t} \tag{5}$$

$$h_t = o_t \odot tanh(c_t) \tag{6}$$

The current LSTM cell is fed by the output of the previous nodes (h_{t-1}), input x_t, and the previous node's cell state c_{t-1}. The cell states contain hidden long-term data. Information is passed through three different LSTM gates: input (i_t), forget (f_t), and output (o_t). Using the gates, the LSTM node can forget or keep cell states and compute the next outputs. In the mathematical equations, σ is the sigmoid function and $\odot$ is element-wise multiplication. The bias and weight values are represented by W and b, respectively. The values for the LSTM cells are shown in Figure 2, which has three cells. The middle cell is the current cell.

- Equation (1), $f_t = \sigma\left(x_t W_{xf} + h_{(t-1)} W_{hf} + b_f\right)$, determines the cell's information state. The forget gate is responsible for this action. It observes the previous nodes (h_{t-1}), input x_t, and the previous nodes' cell state c_{t-1}. It also produces binary outputs 1 and 0, with 1 directing the complete retention of information and 0 directing the information to be discarded.
- Equation (2), $i_t = \sigma\left(x_t W_{xi} + h_{(t-1)} W_{hi} + b_i\right)$, determines which information needs updating. The input gate is responsible for this action.
- Equation (3), $o_t = \sigma\left(x_t W_{xo} + h_{(t-1)} W_{ho} + b_o\right)$, is responsible for the output of the current node. The output gate is responsible for this action. The state of the output

gate is expressed by the values 0 and 1, with 1 indicating that the gate is open and 0 indicating that the gate is closed.

- Equation (4), $\widetilde{c_t} = tanh\left(x_t W_{xc} + h_{(t-1)} W_{hc} + b_c\right)$, calculates the memory gate, which is responsible for selecting the information that needs to be memorized from the temporary cell state.
- Equation (5), $c_t = f_t \odot C_{t-1} + i_t \odot \widetilde{c_t}$, determines the current memory cell's value. The value includes the weight between the memory unit and the input data and the weight between the memory unit and the hidden layer.
- Equation (6), $h_t = o_t \odot tanh(c_t)$, is the final equation that determines the final output at any given time t for the LSTM unit. The model's final output is related to Equations (3) and (5), which are the output value and the current memory value, respectively.

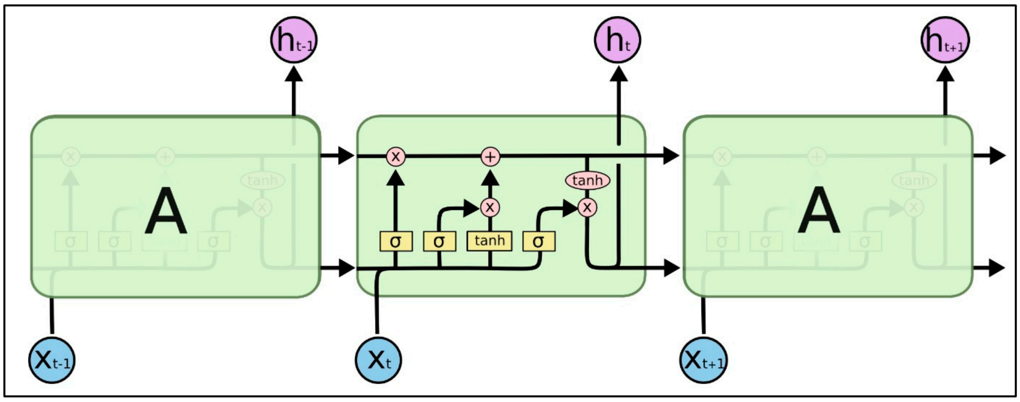

Figure 2. LSTM cells: previous cell, current cell state, and next cell.

Combining the DNN module, which is composed of one layer, with the multilayered LSTM model, which is made up of three LSTM layers, improves prediction accuracy. The dataset goes through cleaning first, in which the missing values are replaced with null and mean values, and then the outliers are removed. Noise is also filtered from the data, and the output from this module is passed on to the DNN layer. The weighted inputs produced as the outputs of the DNN layer are then transferred to the LSTM multilayers. The first layer computes the data and passes them on to the second layer, which refines the output, thus obtaining more accurate predictions. The output from the multilayered LSTM model is then passed on to the fully connected layer, in which the energy consumption values are generated.

3.2.3. Dense Layer

The dense layer is a deeply connected layer connected to the multilayered LSTM layer, in which the layer neurons are linked to every neuron in the multilayered LSTM layer. The main attributes of the dense layer are the weight matrix, the activation function, and the bias vector, which are all used to obtain the desired output. The dense layer is expressed as follows: Output $=$ activation(.(input, kernel) $+$ bias).

The architecture of the proposed model is shown below in Figure 3.

The hybrid forecasting model combines the DNN and LSTM algorithms. The reason for developing the hybrid model is to add a layer that is capable of learning and another layer that is capable of making relationships based on nonlinear data. The proposed model's inputs are divided into sequential time-based information (e.g., past energy consumption) and regular data (e.g., days of the week). Dividing the inputs is important, as the LSTM algorithm depends on time series value sequences as its input. However, considering every possible regressor in energy consumption prediction, some datasets do not have such properties. Generally, energy consumption and differentiation between the data types can be achieved by accounting for the interval data exemplified by the dataset. When the dataset shows an assortment of historical energy consumption measurements, it can usually be demonstrated as sequential time data points and uses the LSTM as regressors. In comparison, when energy data characterize an explicit element related to future data, they cannot be developed as a time series in the sequence [39].

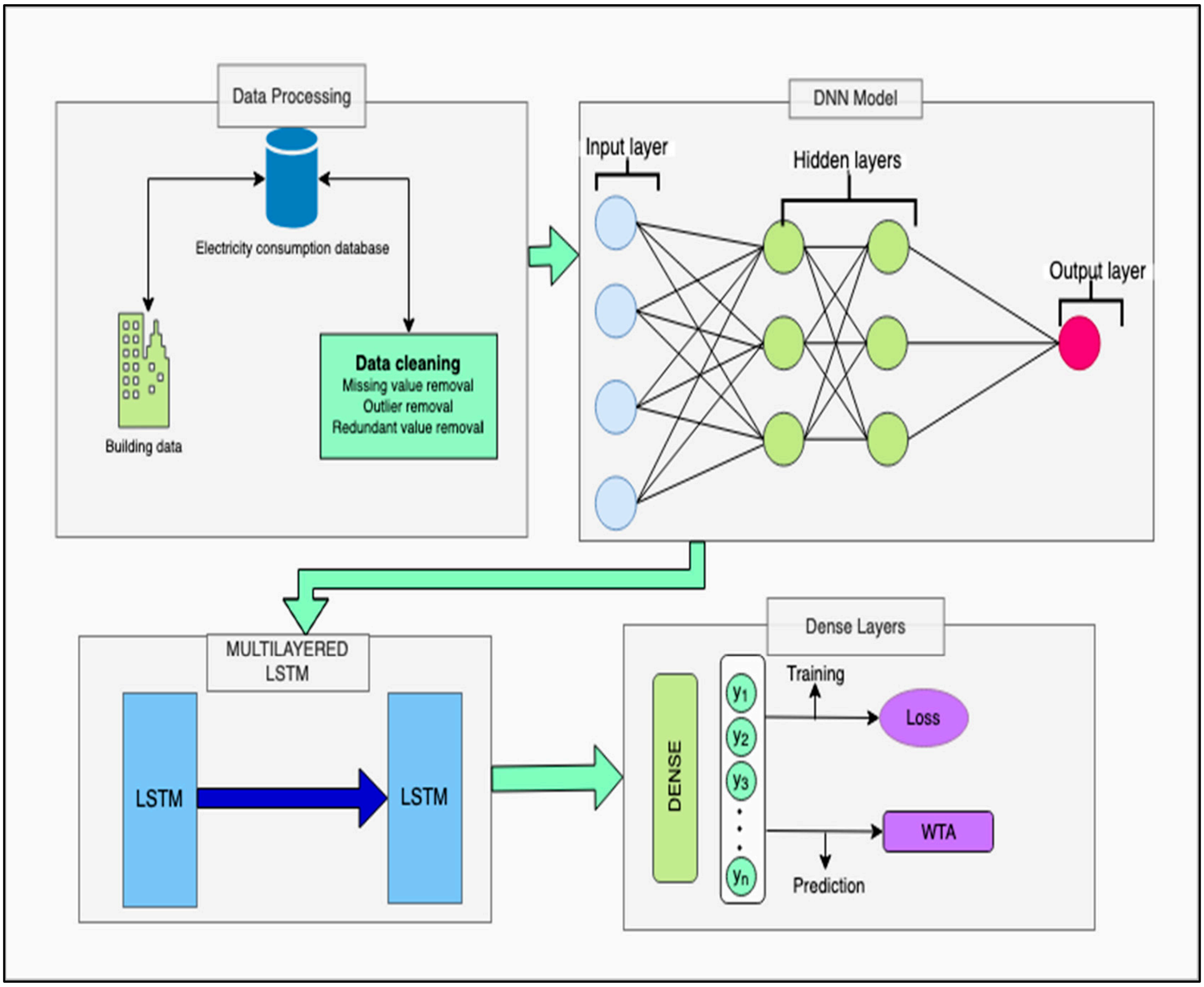

Figure 3. The proposed model.

3.3. Model Architecture

This section introduces RNN-LSTM architecture as a pseudocode.

Prepossess the dataset

- Fill missing values by the mean of values;
- Normalize values using MinMaxScaler;
- Resample the dataset into short-term electrical load predictions: hourly power consumption (34,589 samples);
- Transform the dataset into a time series supervised learning;
- Split the dataset into 70%:30% samples, x_train, x_test, y_train, and y_test.

Construct a DNN-LSTM model

- Three LSTM layers with one hundred neurons; three DNN layer (2048,512) neurons;
- One output layer with one neuron.

The Proposed Model

```
Input: x_train, x_test, y_train, and y_test
Output: DNN-LSTM trained_model
Train DNN-LSTM
Set num_epochs = 50
Set batch_size= 50
Set DNN_LSTM. loss_fn = 'mean_squared_error'
Set DNN_LSTM. Optimizer = 'adam'
For i = 0: num_epochs
    For j = 0: batch_size
        x_batch = x_train[ len(x_train)/batch_size]
        y_batch = y_train[ len(y_train)/batch_size]
        trained_model = DNN_LSTM(x_batch, y_batch)
        //updates weights and biases
        trained_model = optimize(trained_model)
    End
End
Evaluate and Test trained_model (x_test, y_test)
```

4. Model Evaluation

Energy consumption prediction is a time series problem; therefore, error loss metrics were used to measure the differences between the predicted value and the actual value. The metrics used for the evaluation are mean absolute error (MAE), root-mean-square error (RMSE), mean absolute percentage error (MAPE), and the coefficient of determination (R^2). The MAE, RSME, and MAPE are metrics that measure the amount of error in the prediction generated by the model, while R^2 was used to determine how much of a power consumption variability may be explained by its relationship to other factors. These metrics were selected based on the type of the problem and their usage in related studies, such as [28,32–34].

The MAE measures the absolute deviation between the true values and the predicted values; the lower the MAE, the better the model in fitting the data. The RMSE describes the degree of deviation between the true values and the predicted values. Lower RMSE values indicate a more stable model. The MAPE evaluates the model's error through the ratio reflection of the absolute deviation between the true values and the predicted values to the actual values. The model is the most accurate when the index is closest to 0. R^2 represents the proportion of variation in the dependent variable explained by the model. The values of its output range from 0 to 1, and the closer the value is to 1, the better the fitting effect of the model [31]. The formulation of the metrics is shown below. In all cases, N represents the sample size.

4.1. MAE

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N}$$

4.2. RMSE

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

4.3. MAPE

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \hat{y}_i|}{|y_i|}$$

4.4. R^2

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - y_i)^2}$$

where y_i is the true value, $\hat{y}_i$ is the predicted value, and y_i is the mean of the sample.

5. Experimental Results

The experimental setup, evaluation metrics, and performance assessment of the model are discussed in this section.

5.1. Experimental Setup

To evaluate the efficiency of the developed model, the IHEPC was used in the model's training, testing, and validation. The hybrid DNN multilayered LSTM model was trained on a MacBook Pro with an Intel Core i7 processor and 64 GB RAM running on macOS Ventura 13.2. The model was developed using Python coupled with the TensorFlow and Keras frameworks. Among the datasets, 70% were used for training, and 30% were used for testing and model validation. The model settings were a batch size of 50 and an epoch number of 50.

5.2. Data Preparation

5.2.1. Cleaning

The original dataset contained 2,075,259 collected measurements. First, the missing values were cleaned during importation by replacing them with the mean values and removing outliers. Second, the original dataset had more than 2 million records in which energy consumption was collected every minute. Thus, a new dataset was created from the original dataset by resampling hourly, which we used to train the models, and it contained 34,589 rows. This decreased the model's training time without affecting the results that we would have had if we had used the original dataset.

5.2.2. Transformation

While testing for stationarity, the Dicky–Fuller test was used to improve the model accuracy. According to the test, the p-value should be less than 0.05 for the null hypothesis to be true; if greater, the alternative hypothesis should be true. In this study, the p-value was 2.816, indicating that the time series data were nonstationary.

5.2.3. Normalization and Feature Selection

Normalization was applied to the dataset using the MinMax scalar function, and the data were normalized such that the feature values were between 0 and 1. The features from the dataset were selected using the Pearson correlation coefficient, the values of which ranged from −1 to 1. To determine correlation, covariance was applied between the two terms to show how strongly the features correlated. The voltage feature showed a negative correlation relative to the other features, as shown in Figure 4.

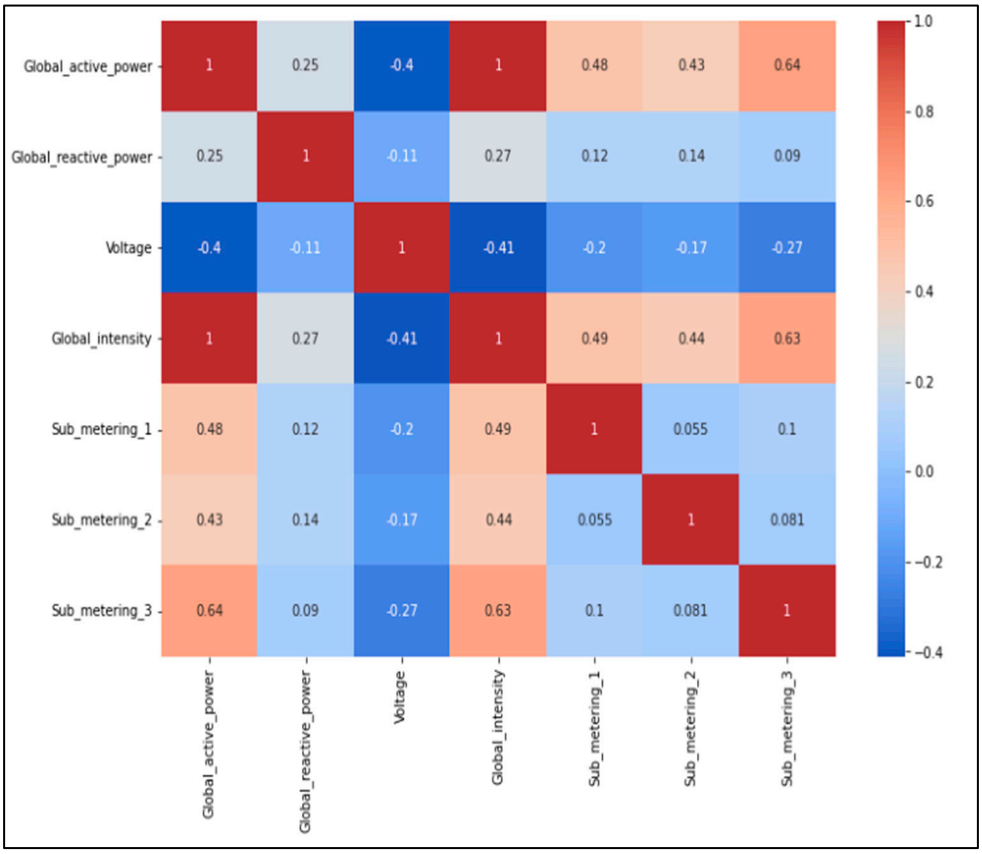

Figure 4. Pearson correlations for feature selection.

5.3. Results and Evaluation Metrics

The MAE, R^2, RMSE, and MAPE metrics were used to evaluate the models.

5.4. Model Training and Loss

The hybrid DNN multilayered LSTM model, DNN model, and LSTM model were implemented. The ReLU activation function was used in the dense layer of the model. Model training involves determining the best values in the datasets for every bias and weight. Model losses often result in poor predictions and indicate insufficient data preprocessing practices. Thus, a perfect predictive model should have zero losses.

The model losses for the proposed model, DNN model, and LSTM model are shown in Figures 5–7, respectively. To check the prediction performance, actual versus predicted values are graphed for the proposed model in Figure 8, the DNN model in Figure 9, and the LSTM model in Figure 10. The resolution was adapted to a 60 min resolution time step for the first 750 h.

Figures 5–7 show the loss of the model for training and testing. There was overfitting observed in the DNN model, which is indicated by a large difference between the training and testing results. This suggests that the DNN model alone cannot effectively capture the relationship between predictors and the response variable. However, when LSTM was integrated with the DNN, the model exhibited consistent results, and the issue of overfitting disappeared.

In addition, there were differences between the predicted values of the DNN model and the actual power consumption, as shown in Figure 9. However, as illustrated in Figures 8 and 10, LSTM improves the effectiveness of the model both as a standalone model and when integrated with the DNN. Therefore, LSTM has the most significant impact on reducing overfitting and improving prediction results.

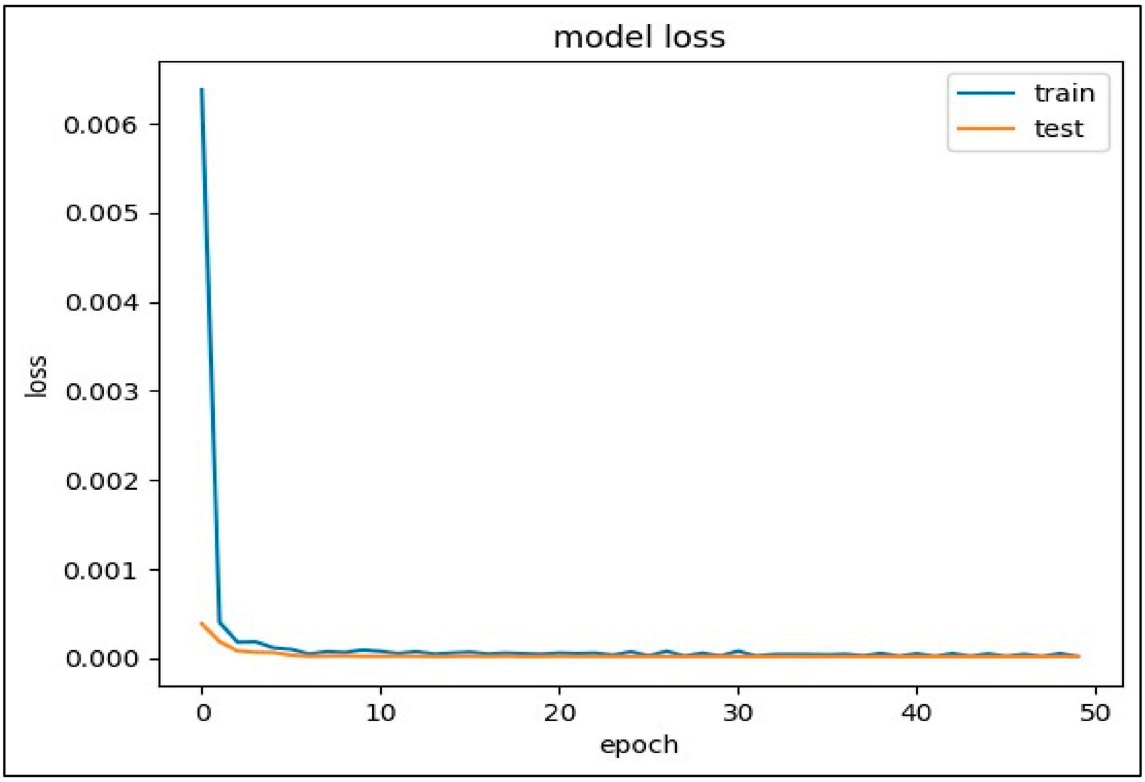

Figure 5. Hybrid DNN multilayered LSTM model loss.

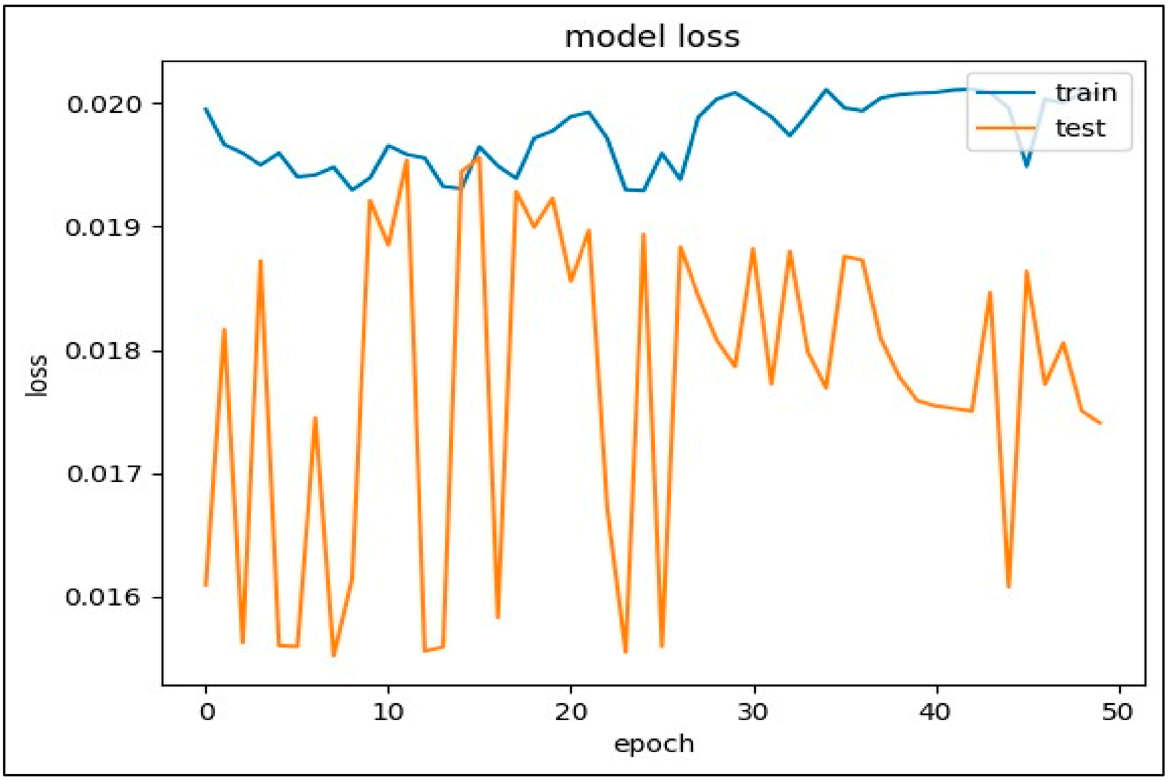

Figure 6. DNN model loss.

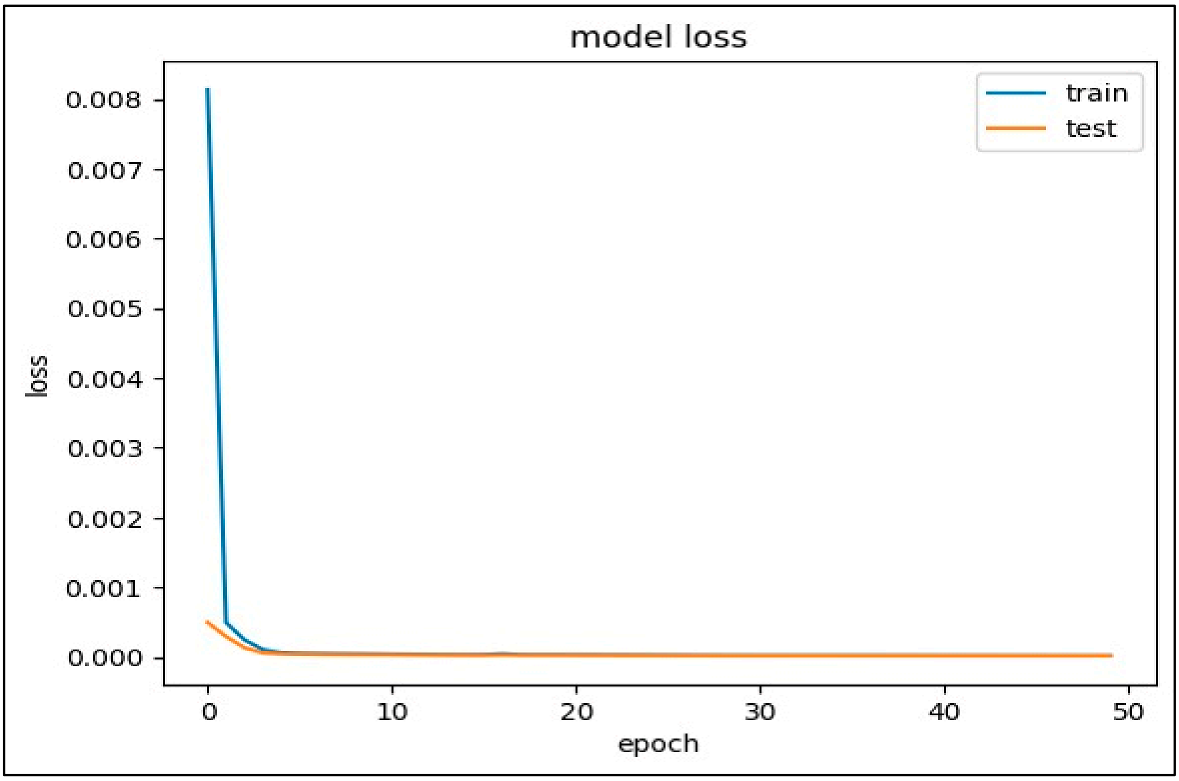

Figure 7. LSTM model loss.

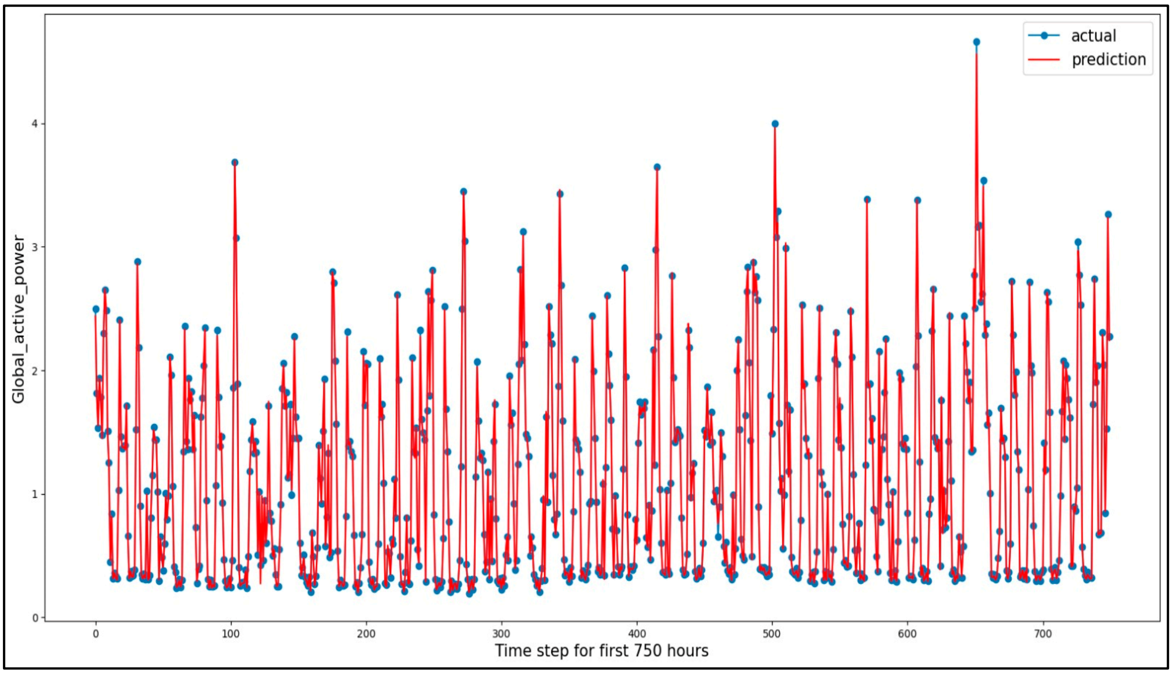

Figure 8. Hybrid DNN multilayered LSTM model, actual vs. prediction.

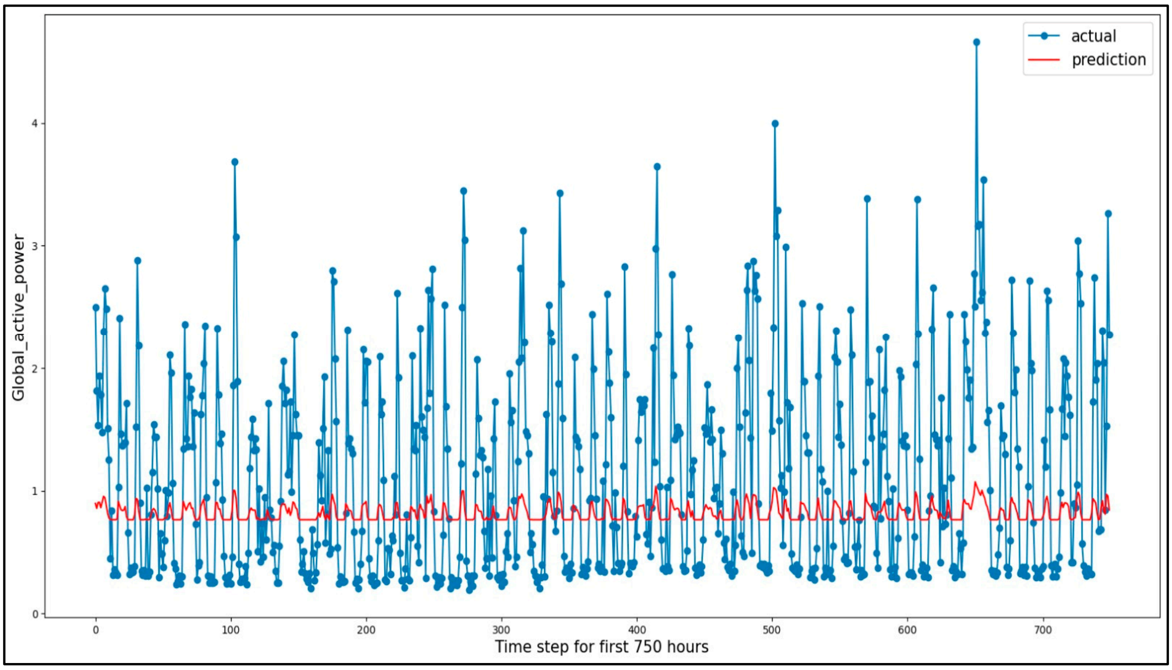

Figure 9. DNN model, actual vs. prediction.

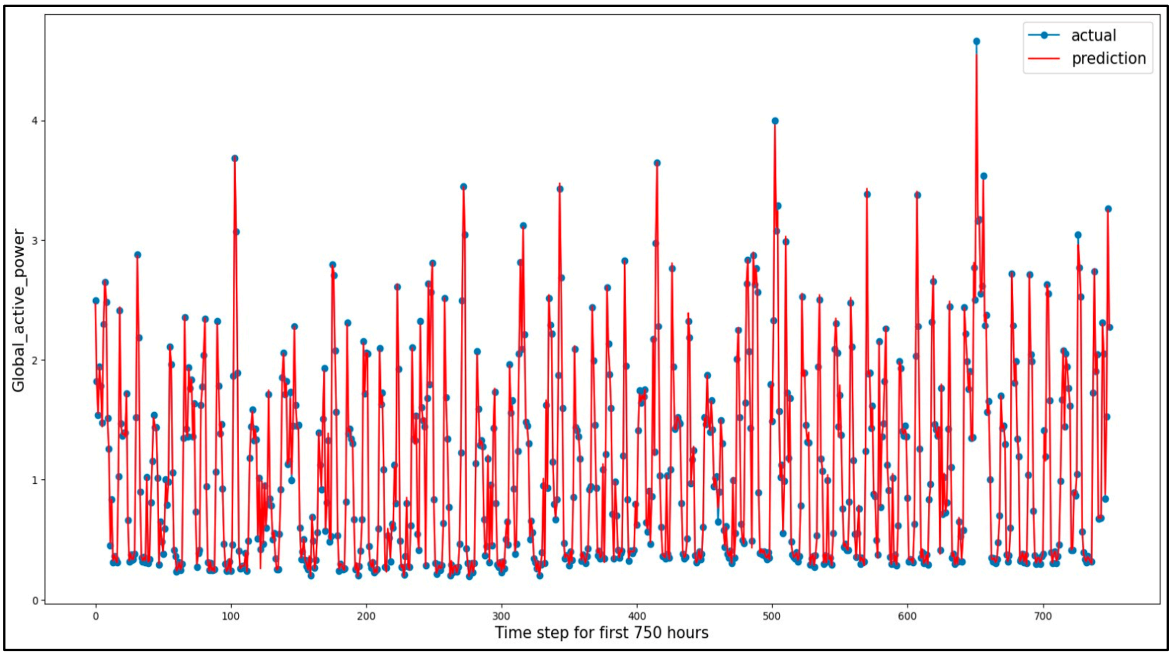

Figure 10. LSTM model, actual vs. prediction.

5.5. Evaluation of the Proposed Model, DNN Model, and LSTM Model

As shown in Figures 5–10, the hybrid DNN multilayered LSTM model performed slightly better than the LSTM model and much better than the DNN model.

The train and test error lines in the model loss graphs showed that the LSTM model and the proposed model performed almost equally and that both were much better than the DNN model. Moreover, the graphs of the actual versus predicted values showed that the DNN model prediction performed poorly, while the prediction performance of the proposed model and the LSTM model was excellent.

Table 2 shows the evaluation metrics for the three models to compare their performances.

Table 2. Model comparison of the proposed model, DNN model, and LSTM model.

Model	RMSE	MAE	MAPE	R^2
DNN	0.81695	0.61333	0.68997	-
LSTM	0.02497	0.01656	0.01957	0.99904
Proposed model: DNN multilayered LSTM	0.02410	0.01565	0.01826	0.99911

The proposed model showed better performance than the DNN model, with a percentage of improvement of 97.05% in RMSE, 97.49% in MAE, and 97.35% in MAPE. This shows the superiority of the proposed model over the DNN model. Conversely, the proposed model has a slightly better improvement than the LSTM model, with a percentage of improvement of 3.5% in RMSE, 5.5% in MAE, 6.7% in MAPE, and 0.007% in R^2. This shows that the proposed model performed better than using the DNN or LSTM alone.

The hybrid DNN multilayered LSTM model was also compared with other deep learning models in previous studies that used the IHEPC dataset to determine the model's performance, as shown in Table 3. The model's performance was evaluated against the CNN multilayered bidirectional LSTM [29], linear regression algorithm [30], CNN-BiGRU [32], GRU [28], CNN-ESN [33], and R-CNN with ML-LSTM [34]. The proposed model was found to be better than the evaluated models, including the GRU [28] and the CNN multilayered bidirectional LSTM [29]. Khan, Hussain, and Baik attested to the adequate potential of hybrid models in the extraction of spatiotemporal features from historical electricity consumption data and renewable power generation [33]. In all the models evaluated, the ability to make accurate pattern predictions stands out, despite the various improvements to increase their reliability in power predictions. Compared with the other models shown in Table 3, the proposed model had the lowest error rates and the highest R^2 value, indicating that the predictions made by the model had minimal errors and were identical to the observed values.

Table 3. Model comparison with various models that used the IHEPC dataset.

Model	RMSE	MAE	MAPE	R^2
CNN multilayered bidirectional LSTM [29]	0.565	0.346	0.2910	-
Linear regression [30]	0.6517	0.5022	83.74	-
CNN-BiGRU [32]	0.42	0.29	-	-
GRU [28]	0.22	0.19	60.00	-
CNN-ESN [33]	0.0472	0.0266	-	-
R-CNN with ML-LSTM [34]	0.0325	0.0144	1.024	0.9841
Proposed model: DNN multilayered LSTM	0.02410	0.01565	0.01826	0.99911

The comparison in Table 3 showed a variety of models whose structure and parameters are different from those of the proposed model, except for the CNN multilayered bidirectional LSTM and R-CNN with ML-LSTM. The two models incorporated LSTM, which has the same structure and parameters as the proposed model. However, all the models had lower prediction accuracy than the proposed model, as indicated by the MAPE values. Only two models used R^2 as an indicator and between the two, the proposed model stood out as the better model. The degree of deviation represented by the RMSE was also the lowest for the proposed model, an indication of a more stable model.

Table 4 summarizes the percentage of improvement when comparing the proposed model with each of the above-mentioned models. R-CNN with ML-LSTM was the only model that showed a better MAE metric than the proposed model. However, considering the other metrics, we can assure that the proposed model is better.

Table 4. The proposed model's percentage of improvement compared with various models.

Model	% of RMSE	% of MAE	% of MAPE	% of R^2
CNN multilayered bidirectional LSTM [29]	95.73%	95.48%	93.73%	
Linear regression [30]	96.30%	96.88%	99.98%	
CNN-BiGRU [32]	94.26%	94.60%		
GRU [28]	89.05%	91.76%	99.97%	
CNN-ESN [33]	48.94%	41.17%		
R-CNN with ML-LSTM [34]	25.85%	−8.68%	98.22%	1.53%

The hybrid model was also compared with other LSTM methods to determine its performance. It was put against LSTM, bidirectional LSTM, multiple LSTM, and WT multiple LSTM. The purpose of the comparison was to establish how the changes in LSTM affected the model. The comparison also fit well because all the models used the same parameters and structure. If the model had LSTM alone or if it was a bidirectional LSTM alone, then multiple LSTM alone, multiple LSTM with WT, or our model, multiple LSTM with a DNN, was used. The indicators used for the comparison were the RMSE, MAPE, and R^2. The results are shown in Table 5.

Table 5. Model comparison with LSTM models.

Model	RMSE	MAPE	R^2
LSTM [31]	0.041	0.00946	0.988
Bidirectional LSTM [31]	0.053	0.01193	0.98
Multiple LSTM [31]	0.023	0.0052	0.996
WT multiple LSTM [31]	0.019	0.00404	0.997
Proposed model: DNN multilayered LSTM	0.02410	0.01826	0.99911

Table 5 shows that the values of all the three indicators improved as the LSTM model improved. The RSME values moved from 0.041 for a lone LSTM model but decreased with the multiple LSTM models. This is good because lower values indicate a more stable model. The MAPE values also decreased while moving from the lone LSTM to the multiple LSTM. The proposed model's value was low but was not the closest to 0 in the comparison. However, the results showed that multiple LSTM had better accuracy than the other models. The model performed best with the coefficient of determination. The value was close to 1, which makes the model have the best-fitting effect. Based on the comparison results, multiple LSTM or multilayered LSTM models provide an improved fitting effect and prediction accuracy.

6. Conclusions

The DNN multilayered LSTM model was proposed in this study to predict energy consumption in households. The IHEPC dataset was used, and the proposed model was discussed. Preprocessing methods were applied to clean, transform, and conduct feature selection for effective training. The proposed model was then trained. The DNN of the proposed model was applied to learn the patterns in the energy consumption data, and the outputs were used as inputs in the multilayered LSTM model. The DNN and LSTM models were also trained on the same dataset.

The three models were compared, and the results showed that the proposed model performed better than the DNN and LSTM models. Moreover, the results of the proposed model were compared with those of various models from related studies. The proposed model also had a better performance than these models.

As the proposed model produced satisfactory results, it can be tested over real-time energy forecasting and integrated into sensor data, which can be used to make energy consumption predictions. In addition, it can be tested in different domains, such as renewable energy consumption, production prediction, and fault detection prediction in industries. In the future, the focus will be on developing a hybrid DNN with other LSTM networks, such as bidirectional LSTM.

Author Contributions: Conceptualization, M.A.-G. and A.A.-M.A.-G.; methodology, M.A.-G. and M.R.; software, M.A.-G.; validation, M.A.-G. and A.A.-M.A.-G.; formal analysis, M.A.-G.; investigation, A.A.-M.A.-G., M.A.-G. and M.R.; resources, M.A.-G.; data curation, M.A.-G., A.A.-M.A.-G. and M.R.; writing—original draft preparation, M.A.-G.; writing—review and editing, M.A.-G.; visualization; M.A.-G.; supervision, A.A.-M.A.-G. and M.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available upon request.

Acknowledgments: The authors gratefully acknowledge the support provided by the Faculty of Computing and Information Technology (FCIT), King Abdulaziz University (KAU), Jeddah, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Yar, H.; Imran, A.S.; Khan, Z.A.; Sajjad, M.; Kastrati, Z. Towards Smart Home Automation Using IoT-Enabled Edge-Computing Paradigm. *Sensors* **2021**, *21*, 4932. [CrossRef] [PubMed]
2. Huang, N.; Lu, G.; Xu, D. A Permutation Importance-Based Feature Selection Method for Short-Term Electricity Load Forecasting Using Random Forest. *Energies* **2016**, *9*, 767. [CrossRef]
3. Vrablecová, P.; Bou Ezzeddine, A.; Rozinajová, V.; Šárik, S.; Sangaiah, A.K. Smart Grid Load Forecasting Using Online Support Vector Regression. *Comput. Electr. Eng.* **2018**, *65*, 102–117. [CrossRef]
4. Dang-Ha, T.-H.; Bianchi, F.M.; Olsson, R. Local Short Term Electricity Load Forecasting: Automatic Approaches. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 4267–4274. [CrossRef]
5. Sönnichsen, N. Global Primary Energy Consumption 2010–2022. 2023. Available online: https://www.statista.com/statistics/265598/consumption-of-primary-energy-worldwide/ (accessed on 29 August 2023).
6. Allouhi, A.; El Fouih, Y.; Kousksou, T.; Jamil, A.; Zeraouli, Y.; Mourad, Y. Energy Consumption and Efficiency in Buildings: Current Status and Future Trends. *J. Clean. Prod.* **2015**, *109*, 118–130. [CrossRef]
7. Zhao, L.; Zhang, J.; Liang, R. Development of an Energy Monitoring System for Large Public Buildings. *Energy Build.* **2013**, *66*, 41–48. [CrossRef]
8. Khodayar, M.; Kaynak, O.; Khodayar, M.E. Rough Deep Neural Architecture for Short-Term Wind Speed Forecasting. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2770–2779. [CrossRef]
9. Tsekouras, G.J.; Hatziargyriou, N.D.; Dialynas, E.N. An Optimized Adaptive Neural Network for Annual Midterm Energy Forecasting. *IEEE Trans. Power Syst.* **2006**, *21*, 385–391. [CrossRef]
10. Wang, Y.; Xia, Q.; Kang, C. Secondary Forecasting Based on Deviation Analysis for Short-Term Load Forecasting. *IEEE Trans. Power Syst.* **2011**, *26*, 500–507. [CrossRef]
11. Li, S.; Wang, P.; Goel, L. A Novel Wavelet-Based Ensemble Method for Short-Term Load Forecasting with Hybrid Neural Networks and Feature Selection. *IEEE Trans. Power Syst.* **2016**, *31*, 1788–1798. [CrossRef]
12. Binyamin, S.S.; Ragab, M. Modified Dwarf Mongoose Optimization Enabled Energy Aware Clustering Scheme for Cognitive Radio Wireless Sensor Networks. *Comput. Syst. Sci. Eng.* **2023**, *47*, 105–119. [CrossRef]
13. Shi, H.; Xu, M.; Li, R. Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. *IEEE Trans. Smart Grid* **2018**, *9*, 5271–5280. [CrossRef]
14. Kong, W.; Dong, Z.Y.; Hill, D.J.; Luo, F.; Xu, Y. Short-Term Residential Load Forecasting Based on Resident Behaviour Learning. *IEEE Trans. Power Syst.* **2018**, *33*, 1087–1088. [CrossRef]
15. Mohan, N.; Soman, K.P.; Sachin Kumar, S. A Data-Driven Strategy for Short-Term Electric Load Forecasting Using Dynamic Mode Decomposition Model. *Appl. Energy* **2018**, *232*, 229–244. [CrossRef]
16. Boroojeni, K.G.; Amini, M.H.; Bahrami, S.; Iyengar, S.S.; Sarwat, A.I.; Karabasoglu, O. A Novel Multi-Time-Scale Modeling for Electric Power Demand Forecasting: From Short-Term to Medium-Term Horizon. *Electr. Power Syst. Res.* **2017**, *142*, 58–73. [CrossRef]

17. Bikcora, C.; Verheijen, L.; Weiland, S. Density Forecasting of Daily Electricity Demand with ARMA-GARCH, CAViaR, and CARE Econometric Models. *Sustain. Energy Grids Netw.* **2018**, *13*, 148–156. [CrossRef]
18. Fumo, N.; Rafe Biswas, M.A. Regression Analysis for Prediction of Residential Energy Consumption. *Renew. Sustain. Energy Rev.* **2015**, *47*, 332–343. [CrossRef]
19. Vu, D.H.; Muttaqi, K.M.; Agalgaonkar, A.P. A Variance Inflation Factor and Backward Elimination Based Robust Regression Model for Forecasting Monthly Electricity Demand Using Climatic Variables. *Appl. Energy* **2015**, *140*, 385–394. [CrossRef]
20. Braun, M.R.; Altan, H.; Beck, S.B.M. Using Regression Analysis to Predict the Future Energy Consumption of a Supermarket in the UK. *Appl. Energy* **2014**, *130*, 305–313. [CrossRef]
21. Chen, Y.; Xu, P.; Chu, Y.; Li, W.; Wu, Y.; Ni, L.; Bao, Y.; Wang, K. Short-Term Electrical Load Forecasting Using the Support Vector Regression (SVR) Model to Calculate the Demand Response Baseline for Office Buildings. *Appl. Energy* **2017**, *195*, 659–670. [CrossRef]
22. Cao, G.; Wu, L. Support Vector Regression with Fruit Fly Optimization Algorithm for Seasonal Electricity Consumption Forecasting. *Energy* **2016**, *115*, 734–745. [CrossRef]
23. Zhong, H.; Wang, J.; Jia, H.; Mu, Y.; Lv, S. Vector Field-Based Support Vector Regression for Building Energy Consumption Prediction. *Appl. Energy* **2019**, *242*, 403–414. [CrossRef]
24. Li, C.; Tao, Y.; Ao, W.; Yang, S.; Bai, Y. Improving Forecasting Accuracy of Daily Enterprise Electricity Consumption Using a Random Forest Based on Ensemble Empirical Mode Decomposition. *Energy* **2018**, *165*, 1220–1227. [CrossRef]
25. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Trans. Smart Grid* **2019**, *10*, 841–851. [CrossRef]
26. Wang, S.; Wang, X.; Wang, S.; Wang, D. Bi-Directional Long Short-Term Memory Method Based on Attention Mechanism and Rolling Update for Short-Term Load Forecasting. *Int. J. Electr. Power Energy Syst.* **2019**, *109*, 470–479. [CrossRef]
27. Raza, M.Q.; Mithulananthan, N.; Li, J.; Lee, K.Y. Multivariate Ensemble Forecast Framework for Demand Prediction of Anomalous Days. *IEEE Trans. Sustain. Energy* **2020**, *11*, 27–36. [CrossRef]
28. Han, T.; Muhammad, K.; Hussain, T.; Lloret, J.; Baik, S.W. An Efficient Deep Learning Framework for Intelligent Energy Management in IoT Networks. *IEEE Internet Things J.* **2021**, *8*, 3170–3179. [CrossRef]
29. Ullah, F.U.M.; Ullah, A.; Haq, I.U.; Rho, S.; Baik, S.W. Short-Term Prediction of Residential Power Energy Consumption via CNN and Multi-Layer Bi-Directional LSTM Networks. *IEEE Access* **2020**, *8*, 123369–123380. [CrossRef]
30. Kim, T.-Y.; Cho, S.-B. Predicting Residential Energy Consumption Using CNN-LSTM Neural Networks. *Energy* **2019**, *182*, 72–81. [CrossRef]
31. Chi, D. Research on Electricity Consumption Forecasting Model Based on Wavelet Transform and Multi-Layer LSTM Model. *Energy Rep.* **2022**, *8*, 220–228. [CrossRef]
32. Khan, Z.A.; Ullah, A.; Ullah, W.; Rho, S.; Lee, M.; Baik, S.W. Electrical Energy Prediction in Residential Buildings for Short-Term Horizons Using Hybrid Deep Learning Strategy. *Appl. Sci.* **2020**, *10*, 8634. [CrossRef]
33. Khan, Z.A.; Hussain, T.; Baik, S.W. Boosting Energy Harvesting via Deep Learning-Based Renewable Power Generation Prediction. *J. King Saud Univ.-Sci.* **2022**, *34*, 101815. [CrossRef]
34. Alsharekh, M.F.; Habib, S.; Dewi, D.A.; Albattah, W.; Islam, M.; Albahli, S. Improving the Efficiency of Multistep Short-Term Electricity Load Forecasting via R-CNN with ML-LSTM. *Sensors* **2022**, *22*, 6913. [CrossRef] [PubMed]
35. Hebrail, G.; Berard, A. Individual Household Electric Power Consumption Data Set, UC Irvine Machine Learning Repository. Available online: https://archive.ics.uci.edu/dataset/235/individual+household+electric+power+consumption (accessed on 29 August 2023).
36. Dargan, S.; Kumar, M.; Ayyagari, M.R.; Kumar, G. A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning. *Arch. Comput. Methods Eng.* **2020**, *27*, 1071–1092. [CrossRef]
37. Yuvaraj, N.; Raja, R.A.; Kousik, N.V.; Johri, P.; Diván, M.J. 15—Analysis on the Prediction of Central Line-Associated Bloodstream Infections (CLABSI) Using Deep Neural Network Classification. In *Computational Intelligence and Its Applications in Healthcare*; Verma, J.K., Paul, S., Johri, P., Eds.; Academic Press: Cambridge, MA, USA, 2020; pp. 229–244. [CrossRef]
38. Tovar, M.; Robles, M.; Rashid, F. PV Power Prediction, Using CNN-LSTM Hybrid Neural Network Model. Case of Study: Temixco-Morelos, México. *Energies* **2020**, *13*, 6512. [CrossRef]
39. Balraj, E.; Harini, R.M.; Kavi Priya, N.; Sneha Prabha, S.B.; Janani, S. A DNN Based LSTM Model for Predicting Future Energy Consumption. In Proceedings of the 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 9–11 May 2022; pp. 1667–1671. [CrossRef]