# A Robust Energy Consumption Forecasting Model using ResNet-LSTM with Huber Loss

**Saleh Albelwi**

Department of Computer Science, Faculty of Computing and Information Technology
Industrial Innovation and Robotics Center
University of Tabuk, Tabuk 47731, Saudi Arabia

**Summary**
Energy consumption has grown alongside dramatic population increases. Statistics show that buildings in particular utilize a significant amount of energy, worldwide. Because of this, building energy prediction is crucial to best optimize utilities' energy plans and also create a predictive model for consumers. To improve energy prediction performance, this paper proposes a ResNet-LSTM model that combines residual networks (ResNets) and long short-term memory (LSTM) for energy consumption prediction. ResNets are utilized to extract complex and rich features, while LSTM has the ability to learn temporal correlation; the dense layer is used as a regression to forecast energy consumption. To make our model more robust, we employed Huber loss during the optimization process. Huber loss obtains high efficiency by handling minor errors quadratically. It also takes the absolute error for large errors to increase robustness. This makes our model less sensitive to outlier data. Our proposed system was trained on historical data to forecast energy consumption for different time series. To evaluate our proposed model, we compared our model's performance with several popular machine learning and deep learning methods such as linear regression, neural networks, decision tree, and convolutional neural networks, etc. The results show that our proposed model predicted energy consumption most accurately.
***Keywords:***
*Energy consumption predication, ResNet-LSTM, ResNet, LSTM, Huber loss, deep learning*

## 1. Introduction

Growing populations and economies require increased energy resources. Statistics show that buildings in particular utilize a significant amount of energy, worldwide. For instance, the buildings in China consumed around 28% of the country's total electricity usage, an amount that grew to 35% in 2020. Similarly, the buildings in the United States consume about 39% of their total energy [1]. Because buildings use significant energy resources, their energy utilization should be more efficiently managed. One strategy for achieving this is energy consumption prediction, which assists building managers in reducing energy consumption and thereby improving their energy utilization rate [2]. Additionally, accurate energy prediction models can help in balancing power consumption with its production, which can lower both resource waste and operating expenses.

In the past decades, a wide range of prediction approaches have been developed for building energy consumption prediction. In general, these energy predictive models can be categorized into two main classes: physical models and data-driven models. The former, also called white-box or forward models, are based on physical laws, as the name suggests. These models need information about a building's HVAC system, for example, in order to predict energy consumption accurately. Software such as EnergyPlus and TRNSYS make use of physical models. This first class of model is most effective when introduced at building design; they are less effective for existing buildings due to time and budget necessary to calibrate the model and locate appropriate parameters. [3].

Data-driven models, on the other hand, utilize historical data to train a predictive model to discover the hidden relationship between the outcome (in this case, the building's energy consumption) and the input features such as the day and time, building equipment information, weather and building information, tenant data, and operational schedule to understand energy behavior using statistical, machine learning, and deep learning methods. These methods construct a model that predicts future energy consumption based on previous data. Data-driven models are simple and flexible, and have thus achieved remarkable interest from researchers [4]. Machine learning algorithms and neural networks have been successfully implemented in energy usage and time series forecasting [5].

Forecasting energy consumption can be done in short, medium-, and long-term timeframes. Short-term models forecast a building's power consumption for the next few hours, up to one day. Mid-term models forecast power connumeration for a few days, up to several months. Long-term forecasting can predict energy usage between

Manuscript received July 5, 2022
Manuscript revised July 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.7.36

one and ten years [6]. Short-term energy forecasting is often the most valuable, because it is easier to identify practical energy-saving measures using hourly data. This makes the forecast more accurate, which is fundamental in operating smart buildings and lowering power consumption [7].

In this paper, we propose a ResNet-LSTM model that combines residual networks (ResNets) [8] and long short-term memory (LSTM) [9] for energy consumption prediction. Our proposed system was trained on historical data to forecast energy consumption with a different time series. ResNets are utilized to extract complex and rich features, while LSTM has the ability to learn temporal correlation; the dense layer is used as a regression to forecast energy consumption. To make our model more robust, we employed Huber loss during the optimization process. Huber loss obtains high efficiency by handling minor errors quadratically. It also takes the absolute error for large errors to increase robustness. This makes our model less sensitive to outlier data. The contributions of this paper can be summarized as follows:

- We proposed a ResNet-LSTM model for energy consumption prediction that combines ResNet blocks to extract complex and rich features, and an LSTM that has the capability of bridging long time lags between inputs over arbitrary time intervals. The use of this LSTM increases the model's effectiveness, because it can identify temporal patterns at varying frequencies. This is beneficial in the prediction of energy consumption over a given time.
- We used Huber loss to train our ResNet-LSTM to provide generalization. Huber loss is more robust against outliers and better at understanding diverse data.

The rest of our paper is structured as follows: the next section describes related works. Sec 3 explains our model design. Sec 4 presents and discusses the results. Finally, Sec 5 presents the conclusion and future works.

## 2. Related Works

Several forecasting techniques have been applied to energy consumption problems in the past twenty years. These include AI methods, neural networks, and machine learning, deep learning, and genetic algorithms. These are utilized most often because they build an intelligent system capable of identifying hidden cues in the data.

Machine learning methods are widely implemented for predicting energy consumption and control purposes. The basic premise is that these models use historical data to identify mathematical models and predict future energy consumption. Ahmad et al. [10] compared the performance of neural networks and random decision for predicting the hourly energy consumption of a hotel in Madrid. They found that neural networks performed better than random forest with small differences in term of mean squared error. Tso et al. [11] employed a decision tree algorithm to build an energy predictive model in Hong Kong. Wang et al. [12] estimated the hourly energy usage for two educational buildings using random forest. They found that different types of input variables, such as time factors, weather conditions, and building occupancy, produced reliable prediction. Paudel et al. [13] developed a model based on support vector machine for estimating the energy demand in a low-energy building. The authors used instance selection to create a small, representative dataset instead of using the full training set. The results showed that the model trained on the small dataset outperformed the model that trained on all the data.

Many studies have shown that a blend of linear and nonlinear models promotes higher accuacy than any single linear or nonlinear model. One of the most popular of these models are autoregressive integrated moving averyage (ARIMA) models. They have been applied across disciplines to create hybrid models with higher accruacy. [14]. As a result, Khashei & Bijari [15] exploited neural networks and ARIMA to build a hybrid model for forecasting energy consumption. In linear modeling, ARIMA models amplify the linear structures within the data to better train the neural networks for electricity consumption prediction. Li et al. [16] proposed a hybrid model called iPSO-ANN to predict the hourly electricity consumption using neural networks. They used a particle swarm pptimization algorithm to update the weights of the neural networks to minimize the error rarte. PCA was utilized to remove redundant information and reduce the dimensionility of the input features. Abedinia et al. [17] used genetic algorithms and intersection theory to create a hybrid model and select the most relevant, and least redundant, features.

Qiu et al. [18] proposed an ensemble deep learning model that combined a deep belief network and support vector regression. The outputs of several deep belief networks were fed into support vector regression for prediction for power forecasting. Chengdong et al. [19] utlized a deep stacked autoencoder to extract discrimatrive features. They then employed extreme machine leanring (ELM) as a predictor to achieve accurate prediction ouputs. Deep autoencoders are applied widely for converting a high-dimensional input into a lower deimensitonal input, and the research has shown that deep autoencoders are more effective than PCA to reduce dimensionality. Somu et al. [20] proposed a $k$CNN- LSTM framework for electricity consumption forecasting. K-means algorithm is used to discover patterns in energy consumption. Convolutional neural networks are then used for extracting complex

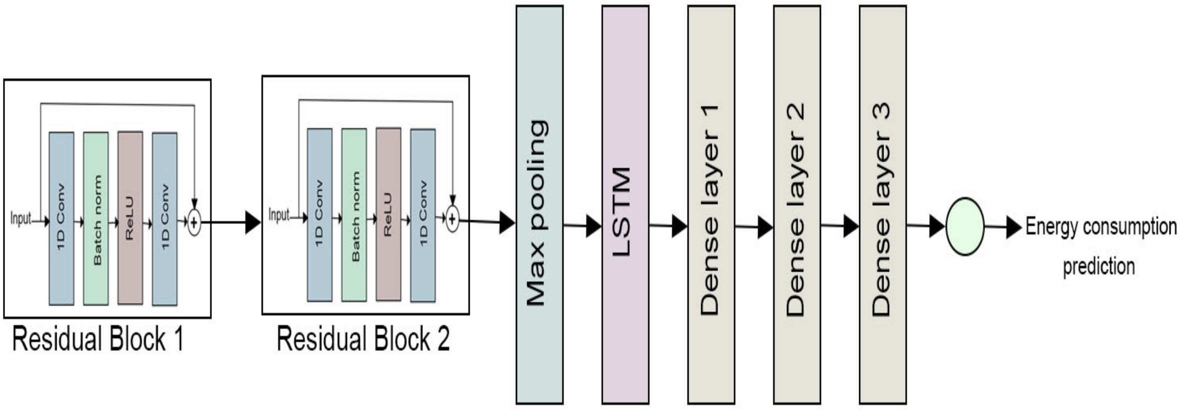


Fig. 1 Architecture of ResNet-LSTM.

features. LSTM neural networks manage long-term dependencies by modeling temporal data in the time series. Several techniques, such as transfer learning and data generation, are employed to build accurate energy forecasting and overcome problems such as insufficient dataset for training predictive models. For example, Tian et al. [21] exploited generative adversarial nets to generate parallel data training. This data was then combined with the original data to train different machine learning models such as BPNN and support vector regression. The results showed that the models trained on mixed data performed better than the models trained on the original data alone. Hooshmand et al. [22] built an energy predictive system based on convolutional neural networks (CNN). They applied transfer learning to address the challenge of insufficient data to train their CNN. The results showed that transfer learning improved the model's performance significantly as compared to existing forecasting models.

## 3. Model design

Our ResNet-LSTM model consists of three main parts: The first is a residual network, which is composed of convolutional layers, max pooling, and batch normalization layers, which extract useful features. The second part exploits the features generated by the LSTM and the fully connected layers. The last part is Huber loss, which optimizes our model during training to reduce the error rate. The ResNet architecture is illustrated in Fig. 1. In this section, we provide a description of the core components in our proposed model.

### 3.1 Residual networks

ResNets have shown superior results in various challenging tasks. They simplify deep network training by bypassing signals from one layer to the next via identity connections.

ResNets use residual blocks as illustrated in Fig. 2, including skip connections, which are a shortcut path for gradient flow. This reduces issues such as vanishing gradients, even if the network is too deep. The element-wise addition of gradients is carried out in residual blocks. ResNets are employed to extract complex and abstract features. This is because skip connections are beneficial in the collection of historical data and the reduction of lost features and information; this allows the model to learn and extract richer features. Each residual block consists of two 1D convolutional layers with a filter size of 1, followed by batch normalization and a ReLU activation function.

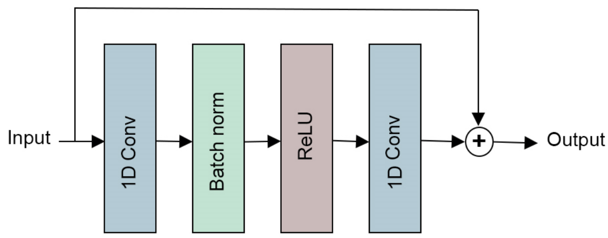


Fig. 1 The structure of Residual block in the ResNet-LSTM model. Each Residual block consists of two 1D convolutional layers, ReLU activation function, and batch normalization layer.

### 3.2 LSTM

LSTM [9] is a special type of recurrent neural network (RNN) designed to address vanishing gradients in RNNs. A gate control mechanism prevents long-term dependence, and is well suited for natural language generation as well as time series processing and prediction. Fig. 3 shows a basic LSTM neural network. It consists of a memory block, which contains a memory cell and three gates to control that cell: the input, output, and forget gates. The forget gate is designed to determine which information needs to be retained and which needs to be forgotten. The forget gate output is computed as follows:

$$f_t=a(W_f + U_f h_{t-1} + b_f) \quad (1)$$

where $W_f$ is a weight matrix, $h_{t-1}$ indicates the output of the previous cell, and $a$ is a sigmoid activation function to identify the information that must be preserved. If the value of the output is close to zero, the information should be forgotten. If it is closer to 1, that means it should be retained [23].

The input gate is built to select which information will be updated using sigmoid activation. It uses a tanh activation function to obtain the $\tilde{C}_t$ value, which is then updated from $C_{t-1}$ to $C_t$. The calculations are done using the following equations:

$$i_t = (W_i + x_t + U_i i h_{t-1} + b_i) \quad (2)$$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (3)$$

$$C_t = f_t . C_{t-1} + i_t . \tilde{C}_t \quad (4)$$

The output gate calculates the current information output. It is filtered using a sigmoid activation function, which obtains $o_t$. The tanh activation function is then used to acquire the desired information $h_t$:

$$o_t = \text{a}\ (W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t . \tanh(C_t) \quad (6)$$

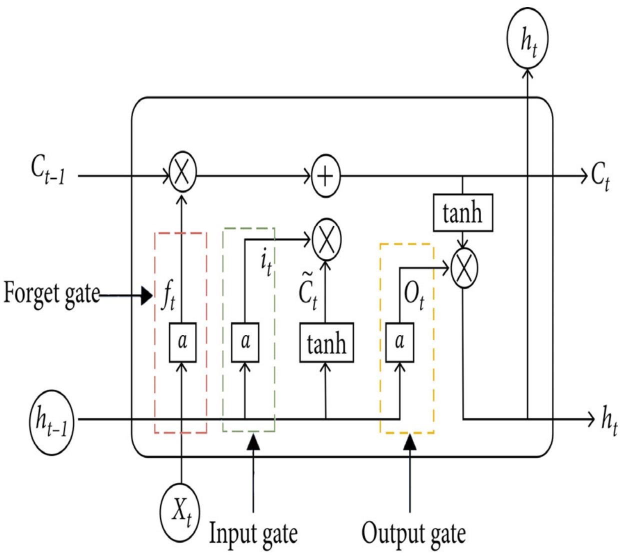


Figure 2. The structure of an LSTM unit.

### 3.3 Huber loss

In supervised learning, loss functions have a significant role in obtaining accurate results. Therefore, selecting the appropriate loss function, based on the noise within the training set, is critical in achieving a generalizable performance. To make our model more robust, we employed Huber loss instead of mean squared error, which is a combination of mean squared error and mean absolute error. Huber loss is a quadratic function (MSE) when the error between the target and predicted outputs is small, and a linear function when the errors are larger. Because of this, Huber loss handles outliers and diverse data better than other options. Huber loss is defined as follows:

$$L = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 & |y_i - \hat{y}_i| \leq \delta \\ \frac{1}{n}\sum_{i=1}^{n}(|y_i - \hat{y}_i| - \frac{1}{2}\delta) & |y_i - \hat{y}_i| > \delta \end{cases} \quad (7)$$

Where $y_i$ indicates the target output for example $i$ , $\hat{y}_i$ indicates the predicted output for example $i$, $n$ represents the number of training set, and $\delta$ is the hyperparameter that defines the threshold for the loss function to transfer from quadratic to linear.

The CNN-LSTM model consists of two residual blocks for feature extraction. Then, the extracted features are fed into an LSTM layer to learn the temporal correlation. The final four layers are dense layers. The last dense layer consists of one neuron with a linear activation function to predict energy consumption. Each residual block consists of two 1D convolutional layers, a batch normalization layer, and 1D max pooling.

## 4. Experimental results and discussion

### 4.1 Dataset

We used the PJM East region dataset (PJMER), which estimates the hourly energy consumption in megawatts for the eastern United States. The dataset consists of 145,366 records of hourly energy consumption data from 2002 to 2018. The dataset consists of two columns: The first is a datetime input variable, which contains information about both data and time. The second contains the energy consumption in megawatts. We have split the datetime into eight features as follows: hour, day of week, quarter, month, year, day of year, day of month, and week of year. Table 1 provides a descriptive summary about the dataset, such as the min, max, mean, standard deviation, skew, and kurtosis.

Table 1. Statistical information about PJMER dataset

| *Statistic value* | *Value* |
|---|---|
| Min | 14544.00 |
| Max | 62009.00 |
| Mean | 32080.22 |
| Standard deviation | 6464.01 |
| Skew | 0.739 |
| Kurtosis | 0.736 |
| Min | 14544.00 |
| Max | 62009.00 |

The dataset records from 2002 to 2015 were used to train the model, and the years 2016, 2017, and 2018 were used for testing, as shown in Fig. 4. The figure also shows the hourly energy consumption prediction for PJMER. The dataset is normalized between 0 and 1 using a min-max technique.

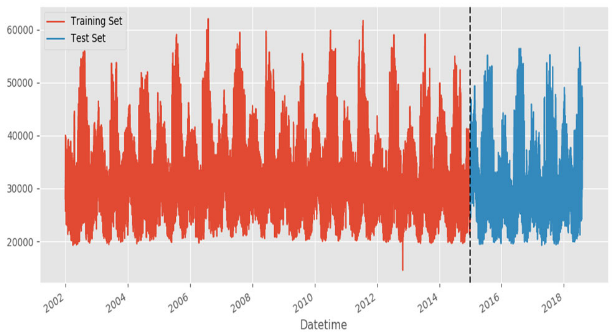


Fig. 4 Hourly energy consumption from the PJMER dataset, from 2002 to 2018. The data from 2002 to 2015 was used for training, and the data from 2016 to 2018 was used for testing.

## 4.2 Experimental setup and evaluation matrices

The whole model was built in Python. The ResNet-LSTM was developed using the TensorFlow framework. Several machine learning algorithms were applied via the sklearn library. Huber loss was minimized using an Adam optimizer during the optimization process with a learning rate of 0.002, a batch size set to 128, and the total number of epochs at 100. Dropout was applied with dense layers with a dropout rate of 0.3 to reduce overfitting. The value of $\delta$ set to 1.

In this work, we evaluated the performance of the models using mean squared error (MSE), mean absolute error (MAE), and root-mean-squared error (RMSE). They equations are defined, respectively, as follows:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{8}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{9}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{10}$$

where $y_i$ is the actual value (correct energy consumption) for $i$-example, and $\hat{y}_i$ is the predicted output from the model for $i$-example.

## 4.3 Results and Discussion

To validate the effectiveness of the ResNet-LSTM presented in this paper, we compared our ResNet-LSTM with the most popular predictive models such as linear regression, neural networks, support vector regression, convolutional neural networks, and LSTM. The results, as shown in Table 2 and Fig. 5-7, illustrate that our ResNet-LSTM obtained the best performance among all other prediction models. It also has the lowest error in terms of MSE, MAE, and RMSE. The RMSE of our ResNet-LSTM is 4246.72, the MSE is 18034620.37, and the MAE is 3256.85.

Table 2. The results of the MSE, MAE, RMSE comparisons among different prediction models.

| **Algorithm** | **MSE** | **MAE** | **RMSE** |
|---|---|---|---|
| Linear regression | 32471863.78 | 4586.08 | 5698.41 |
| Support vector regression | 33910528.53 | 4492.44 | 5823.27 |
| Neural networks | 32043511.08 | 4547.90 | 5660.69 |
| Decision Tree | 26260089.45 | 3732.70 | 5124.45 |
| CNN | 28964812.00 | 3876.43 | 5381.89 |
| ResNet-LSTM | 18034620.37 | 3256.85 | 4246.72 |

The ResNet-LSTM has shown impressive results as compared to other forecasting models in terms of error value. Our model accurately predicts energy consumption for a particular hour within a day, week, month, or year.

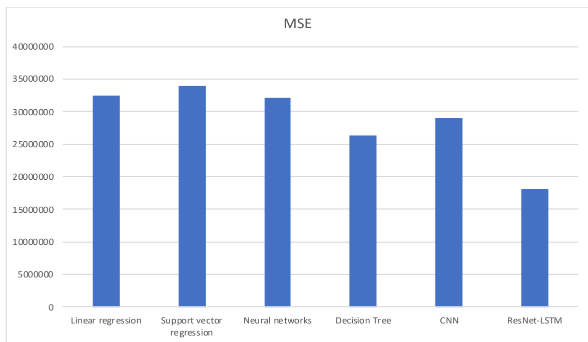


Fig. 5. The results of MSE among different predictive models.

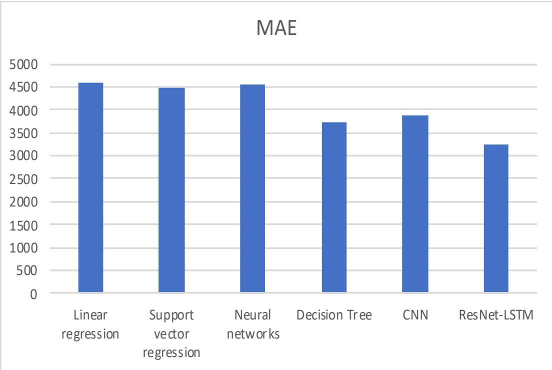


Fig. 6 The results of MAE among different predictive models.

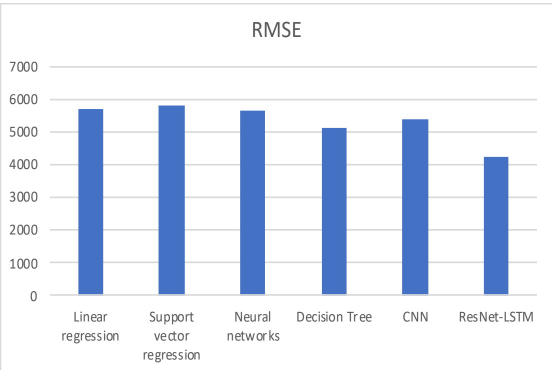


Fig 7. The results of RMSE among different predictive models.

## 5. Conclusion

Reliable energy consumption prediction models are of great significance for both energy planning and the enhancement of energy utilization. Deep learning algorithms have demonstrated powerful predictions and learning in time series applications. In this paper we employed the power of residual networks and LSTM to build a robust energy consumption prediction model. The proposed model is called ResNet-LSTM, and it is composed of residual blocks that function as feature extractors. LSTM can learn long-term dependencies in series, which is useful in time series prediction, and the fully connected layer is used as a regressor to predict energy consumption. The results showed that our ResNet-LSTM outperformed a variety of machine learning methods. Future work will focus on models that take into accounts other factors, such as the number of occupants, information about the building, information on holidays, etc.

## References


[1] G. Xu and W. Wang, "China's energy consumption in construction and building sectors: An outlook to 2100," *Energy,* vol. 195, p. 117045, 2020.

[2] Q. Fu, Z. Han, J. Chen, Y. Lu, H. Wu, and Y. Wang, "Applications of reinforcement learning for building energy efficiency control: A review," *Journal of Building Engineering,* vol. 50, p. 104165, 2022.

[3] J. Runge and R. Zmeureanu, "Forecasting energy use in buildings using artificial neural networks: A review," *Energies,* vol. 12, no. 17, p. 3254, 2019.

[4] Y. Chen, M. Guo, Z. Chen, Z. Chen, and Y. Ji, "Physical energy and data-driven models in building energy prediction: A review," *Energy Reports,* vol. 8, pp. 2656-2671, 2022.

[5] G. Tardioli, R. Kerrigan, M. Oates, O. D. James, and D. Finn, "Data driven approaches for prediction of building energy consumption at urban level," *Energy Procedia,* vol. 78, pp. 3378-3383, 2015.

[6] K. Karabulut, A. Alkan, and A. S. Yilmaz, "Long term energy consumption forecasting using genetic programming," *Mathematical and Computational Applications,* vol. 13, no. 2, pp. 71-80, 2008.

[7] D. Mariano-Hernández *et al.*, "A Data-Driven Forecasting Strategy to Predict Continuous Hourly Energy Demand in Smart Buildings," *Applied Sciences,* vol. 11, no. 17, p. 7886, 2021.

[8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[9] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation,* vol. 9, no. 8, pp. 1735-1780, 1997.

[10] M. W. Ahmad, M. Mourshed, and Y. Rezgui, "Trees vs Neurons: Comparison between random forest and ANN for high-resolution prediction of building energy consumption," *Energy and buildings,* vol. 147, pp. 77-89, 2017.

[11] G. K. Tso and K. K. Yau, "Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks," *Energy,* vol. 32, no. 9, pp. 1761-1768, 2007.

[12] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, "Random Forest based hourly building energy prediction," *Energy and Buildings,* vol. 171, pp. 11-25, 2018.

[13] S. Paudel *et al.*, "A relevant data selection method for energy consumption prediction of low energy building based on support vector machine," *Energy and Buildings,* vol. 138, pp. 240-256, 2017.

[14] C. N. Babu and B. E. Reddy, "A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data," *Applied Soft Computing,* vol. 23, pp. 27-38, 2014.

[15] M. Khashei and M. Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting," *Applied soft computing,* vol. 11, no. 2, pp. 2664-2675, 2011.

[16] K. Li, C. Hu, G. Liu, and W. Xue, "Building's electricity consumption prediction using optimized artificial neural networks and principal component analysis," *Energy and Buildings,* vol. 108, pp. 106-113, 2015.
[17] O. Abedinia, N. Amjady, and H. Zareipour, "A new feature selection technique for load and price forecast of electrical power systems," *IEEE Transactions on Power Systems,* vol. 32, no. 1, pp. 62-74, 2016.
[18] X. Qiu, L. Zhang, Y. Ren, P. N. Suganthan, and G. Amaratunga, "Ensemble deep learning for regression and time series forecasting," in *2014 IEEE symposium on computational intelligence in ensemble learning (CIEL)*, 2014: IEEE, pp. 1-6.
[19] C. Li, Z. Ding, D. Zhao, J. Yi, and G. Zhang, "Building energy consumption prediction: An extreme deep learning approach," *Energies,* vol. 10, no. 10, p. 1525, 2017.
[20] N. Somu, G. R. MR, and K. Ramamritham, "A deep learning framework for building energy consumption forecast," *Renewable and Sustainable Energy Reviews,* vol. 137, p. 110591, 2021.
[21] C. Tian, C. Li, G. Zhang, and Y. Lv, "Data driven parallel prediction of building energy consumption using generative adversarial nets," *Energy and Buildings,* vol. 186, pp. 230-243, 2019.
[22] A. Hooshmand and R. Sharma, "Energy predictive models with limited data using transfer learning," in *Proceedings of the tenth ACM international conference on future energy systems*, 2019, pp. 12-16.
[23] W. Lu, J. Li, Y. Li, A. Sun, and J. Wang, "A CNN-LSTM-based model to forecast stock prices," *Complexity,* vol. 2020, 2020.