Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 1 (2025)

Nonlinear Ensemble Deep Learning Model for Energy Consumption Prediction with Bayesian Optimization

Ejigu Tefera[1], Kula Kekeba[1], Ravindra Babu.B[2], and M. Mart´ınez-Ballesteros[3]
[1]Big Data and HPC Center of Excellence, Department of Software Engineering, Addis Ababa Science & Technology University, Addis Ababa P.O. Box 16417, Ethiopia, ejigu.tefera@aastustudent.edu.et, kuulla@gmail.com
[2]Distributed Systems Research Group (SIG), Adama Science and Technology University, P.O. Box 1888, Adama, Ethiopia, ravindrababu4u@yahoo.com
[3]Department of Computer Science, University of Seville, ES-41012 Seville, Spain, mariamartinez@us.es

Article History:

Received: *10-07-2024*

Revised: *23-08-2024*

Accepted: *06-09-2024*

Abstract:
Accurate prediction of electric energy consumption is crucial for efficient load dispatching, energy utilization, and grid operation. Traditional statistical and classical machine learning methods struggle with the nonlinear nature of energy consumption data, often leading to higher prediction errors. Additionally, deep learning models using a single approach face challenges such as convergence to local minima and poor generalization. This paper proposes a nonlinear ensemble deep learning model for residential energy consumption prediction, incorporating Bayesian optimization for hyperparameter tuning. The model combines Long Short-Term Memory (LSTM), Bidirectional LSTM (BiLSTM), and 1D Convolutional Neural Networks (1D-CNN), leveraging their powerful nonlinear feature learning capabilities. A k-means clustering approach is used to preprocess and reduce variability in the data, enhancing the ensemble model's performance. The ensemble model was tested on real energy consumption data from two districts in Addis Ababa, showing significant improvements in prediction accuracy with lower MAE, RMSE, and MAPE values compared to single models and un-clustered data. The integration of clustering and Bayesian optimization further enhanced model generalizability and minimized overfitting, demonstrating the effectiveness of a nonlinear approach in capturing complex energy consumption patterns.

Keywords: Bayesian Optimization, Deep Learning, Ensemble Learning, Hyperparameter tuning, k-means Clustering.

1. Introduction

Nowadays, people are highly dependent on the supply of sufficient and stable eclectic energy to live comfortably [31]. Consequently, electricity consumption demand has been rising due to the growth of urbanization along with the rapid growth of the human population throughout the world [32].

More importantly, energy demand in Africa keeps growing annually at an average rate of 4%, the highest in the world [13]. Similar to Africa, energy consumption demand steadily growing in Ethiopia. From the Ethiopian context, the residential consumption demand accounts for 39% which is the largest followed by the industrial (34%) and commercial (27%) sectors. Specifically, household energy demand is expected to exceed population growth because economic improvements have driven households to have appliances and become owners of energy dependent technological devices that use energy continuously. However, energy supply and distribution are characterized by frequent power interruption, inefficient utilization, and substantial waste [6]. Apart from the high demand frequent power outages problems, and the high level of energy demand especially in Ethiopia, once it is generated, storing and preserving the produced electricity energy sufficiently using the current energy storage technology [1] is difficult. In other words, energy waste will occur if the electricity is adequately distributed and consumed as soon as it is produced following the consumption demand in each district.

To this end, given the large contribution of the residential sector to total energy demand, it is feasible to study consumption trends and develop accurate models using state-of-the-art data-driven algorithms for effective planning and demand-supply management. Moreover, for reliable and efficient grid systems, effective load dispatching, and efficient energy utilization, accurate electricity consumption forecasting has become indispensable for energy companies. In this regard, machine learning and deep learning models have been widely used for electric load forecasting, power system monitoring, and anomalous energy usage detection [21]. However, the existing machine learning methods are incapable of capturing nonlinear energy consumption data and cannot yield accurate prediction results [5, 15]. Moreover, deep learning methods with a single model have been plagued by a poor capacity for generalization and a tendency to converge to local minima [11, 29]. In existing methods, little attention is given to the fine graining of the input data, which accounts for model complexity and larger prediction errors [35]. In general, despite several studies have been conducted for electric load forecasting based on deep learning and ensemble methods, enhancement is required to get optimal prediction performance by ensembling multiple deep learning algorithms with clustering and fine graining of the input data to learn nonlinear and complex energy data effectively [29, 25].

In this regard, the electricity consumption prediction method is imperative to ensure efficient load dispatching, scheduling, and efficient energy utilization [19]. This paper aims to investigate the effectiveness of an ensemble deep learning model for energy consumption prediction with fin-graining of input data including identifying optimal clusters of residential energy consumption profiles. The contributions of this paper can be summarized as follows:

1. K-means clustering was applied for energy consumption profile characterization to acquire a more thorough understanding of how power consumption patterns of users behave. Moreover, optimal clusters were identified that will lead the subsequent ensemble model to learn the detail features and intrinsic behaviors of energy consumption data.
2. Optimal hyperparameter combination is searched using a Bayesian optimization algorithm to get an improved prediction model.
3. The robust ensemble model has been developed based on optimal cluster-generated energy consumption data.

4. The ensemble deep learning model’s effectiveness in predicting the monthly aggregate residential energy

consumption is evaluated and verified against the base models using MAE, RMSE, and MAPE.

2 Related Works

Accurate electric energy consumption forecasting at both long-term and short-term horizons is necessary to establish a more stable supply-and-demand equilibrium [30]. To this end, several studies have been conducted on energy consumption forecasting problems. Wen et al. [35] proposed a deep-learning model to forecast the load demand for residential buildings with a one-hour resolution. Hyperopt hyperparameter tuning was employed to find the optimal hyperparameter combination. Moreover, accurate forecasting of electricity consumption is a very challenging task due to the high volatility of energy consumption. A. Salam and A. El Hibaoui in [24] introduced an improved intelligent energy prediction model based on deep feedforward neural networks and Long Short-Term Memory. M. Cai et al. [2] proposed deep neural network models, namely recurrent neural networks (RNN) and convolutional neural networks (CNNs). The proposed model is compared with the Seasonal ARIMAX model's accuracy, computational efficiency, generalizability, and robustness. Among all the investigated deep learning techniques, the gated 24- h CNN model achieved the best performance, improving the forecasting accuracy by 22.6% compared to the seasonal ARIMAX.

N. Somu, et al. in [27] proposed a hybrid model for building energy consumption forecasting using long short-term memory networks. In this work, a novel *Haar* wavelet-based mutation operator was introduced to improve the divergence nature of the sine cosine optimization algorithm while dealing with hyperparameter tuning using the sine cosine optimization algorithm. On the other hand, a hybrid of wavelet transform and machine learning model is proposed in [26] to estimate electrical load consumption using the historical time-series information of energy usage. To investigate the effectiveness of combining different deep learning algorithms for estimating residential household energy consumption, Authors in [14] employed a hybrid ensemble model consisting of CNN, multilayer LSTM, and BiLSTM algorithms, by which the CNN framework can extract spatial and non-linear patterns of the energy data and multilayer LSTM used to learn temporal dependencies.

An ensemble method [36] is developed to forecast the residential short-term energy consumption. Vector auto-regression, Gaussian process regression, and the long short-term memory neural network model were trained as base learners. Another ensemble method is proposed in [23] by combining the deep LSTM and Auto-regressive Integrated Moving Average (ARIMA) models. In this work, the ARIMA was used to capture the stationary pattern of load data, and the nonlinearity of the complex energy consumption data was tackled using LSTM architecture. The performance of the proposed model surpasses existing short-term load forecasting models with less computation complexity.

Moreover, ensemble learning methods provide a powerful tool for improving accuracy and stability in power load forecasting by leveraging the strengths of multiple predictor techniques [30, 11, 29]. Hadjout et al. [10] introduced an ensemble model for monthly industrial energy consumption forecasting. The proposed model combines LSTM, GRU, and TCN based on weighted averages. Similarly, W. Khan et al. [16] developed an effective ensemble model but at this time the authors employed a stacking-based ensemble approach using simple neural networks (ANN) and LSTM as base learners for solar energy forecasting. XGBoost algorithm was used as a meta-learner to combine the base models and the proposed model exhibited better consistency and stability in different cases.

In general, most of the related works [16, 23, 24] have employed grid search for optimization tasks to improve deep learning and ensemble model performance. However, this optimization approach is highly criticized for high computation time requirements and is ineffective when the number of hyperparameter spaces and the type of hyperparameters have been increased. In addition, the presence of outliers in a dataset degrades model prediction performance and reduces model generalization abilities. This problem has been observed in the above-mentioned related works that little attention is given to the fine-graining of the input data, which accounts for model complexity and larger prediction errors [35].

3 Methodology

The proposed method comprises three main phases: (1) Data preprocessing and clustering for energy consumption pattern identification; (2) Deep learning model hyperparameter tuning using a Bayesian optimization algorithm and (3) Individual base model training and model fusion to develop ensemble model and performance evaluation. In general, figure 11 shows the details of the proposed model.

3.1 Description of Data and Data Preprocessing

The data for this study was collected from the Ethiopian electric utility. The collected data is about two districts of Addis Ababa city, South and West Addis Ababa (hereafter South AA and West AA) district's monthly residential energy consumption data ranging from May 2019 to January 2021. Since each month's consumption data was obtained from a different monthly bill report in separate Excel files for each month, it is necessary to combine the individual monthly file into a single, sequentially arranged dataset for each district as a Figure 1 shows.

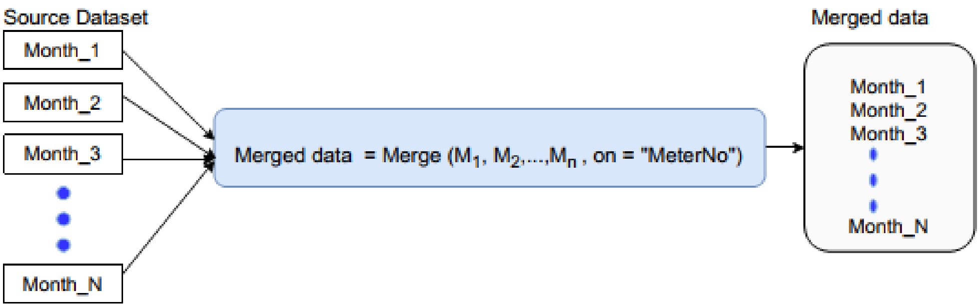


Figure 1: Data Aggregation

There were many missing values in each month's data because some customers may not have paid their consumption charge within the specified billing period. Consequently, users who have zero electric consumption for at least one month out of the 20 months or users who are absent for at least one month out of 20 months have been removed using filtering techniques because they are not good representatives of the samples. After filtering out the missing values, Table 1 shows the size of the input observation. Therefore, given the number of customers, C in each month for each selected district, and the number of months, M, the input observations or dataset D for each case study data is:

$$Dataset, D = C * M \tag{1}$$

where C is the number of customers in each month and M is the number of months considered in each district. Furthermore, Table 1 summarizes the descriptive statistics of each district dataset.

Table 1: Descriptive statistics of the load consumption dataset

Dataset	Count	Mean	Max.	Min.	Std.	Skewness	Kurtosis
South AA	309303	304.00	919	0.100	182.74	0.849	0.29
West AA	313491	279.72	9120	0.020	215.87	11.24	7.93

Moreover, Figure 2 shows the average monthly electricity consumption over 20 months for the South AA district on the left and West AA district on the right.

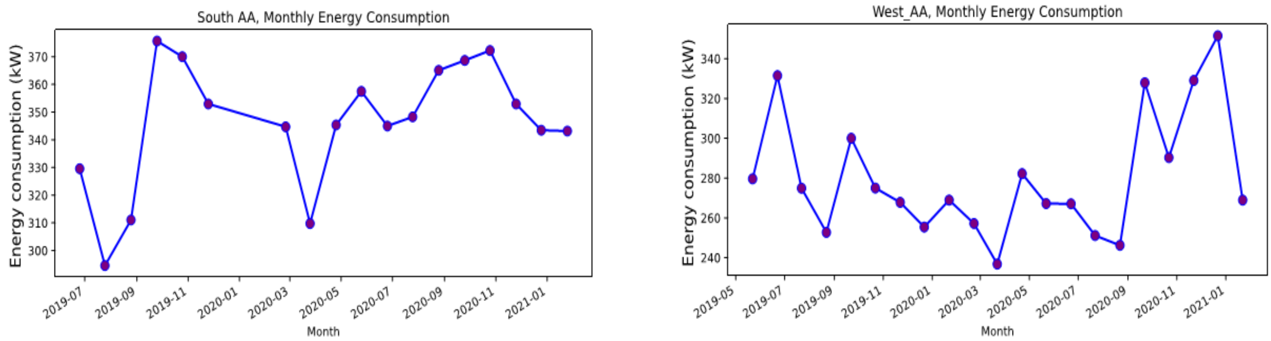


Figure 2: Aggregate Monthly Energy Consumption of west and West AA

The data indicates that there is significant variation in electricity consumption habits from month to month in both districts. Furthermore, the pattern of electricity usage in each district is non-linear and irregular making difficult accurate predictions of this data using classical machine learning models and traditional statistical techniques [7]. In other words, this kind of data requires an effective preprocessing method such as a k-means clustering algorithm to discover the optimal clusters comprising more stable and similar consumption profiles. Moreover, integrating advanced data preprocessing techniques such as k-means clustering will enable the subsequent ensemble deep learning model [5, 17, 22] to learn the nonlinear complex association between energy consumption features and make accurate predictions.

3.2 Experiment Setup

This section tried to discuss experimentation phases of our study which include data processing and ensemble deep learning model development based on the cluster-generated data. In this phase, the Keras framework on top of TensorFlow was selected to utilize a deep learning framework, hyperparameters, and a Bayesian optimization algorithm based on the BayesSearchCV interface.

3.2.1 Energy Consumption Clustering and Analysis

In this study, K-means++ clustering was employed to discover the optimal clusters because kmeans++ is developed as an enhanced version of k-means clustering in initial cluster center identification and can give faster computation advantages [33].

In this study, K-means++ clustering was employed to discover the optimal clusters because k -means++ is developed as an enhanced version of k -means clustering in initial cluster center identification and can give faster computation advantages [5].

The silhouette coefficient is used to evaluate how effective a clustering method is. It has a value between -1 and 1 . From this $a(i)$ represents the average distance from an item i in the cluster A to all other objects in A, and $d(i, C)$ represents the average distance from an object i to all objects in the cluster $C \neq A$. After computing $d(i, C)$ and $C \neq A$ for each cluster, the smallest cluster is chosen as described below.

$$b(i) = \min_{C \neq A} d(i, C) \text{ with } i \in A \tag{2}$$

The value $b(i)$ denotes to what extent a data point i is dissimilar to its nearest neighbor cluster. Thus, the silhouette values, silh(i) are given in Equation (3):

$$\text{silh }(i) = \frac{a(i) - b(i)}{\max\{a(i), b(i)\}} \tag{3}$$

Another cluster validation metric is the DBI which is used to determine the goodness of clusters. The DBI for K clusters C_i with $i = 1, \ldots, K$ is defined according to Equation (4):

$$DB_K = \frac{1}{K}\sum_{i=1}^{K} \max_{j \neq i} f_{i,j} \tag{4}$$

where:

$$f_{i,j} = \frac{\text{diam }(C_i) + \text{diam }(C_j)}{d(C_i, C_j)} \tag{5}$$

and, in this case, the diameter of a cluster is defined as:

$$\text{diam }(C_i) = \left(\frac{1}{n_i}\sum_{x \in C_i} \|x - z_i\|^2\right)^{\frac{1}{2}} \tag{6}$$

with n_i the number of data points and z_i the centroid of cluster C_i. The DBI will achieve very small values, which guarantees the presence of high-quality clusters. Therefore, the ideal number of clusters is discovered when this index is minimized depending on the input dataset.

Algorithm 1 Pseudocode for the K-means++ clustering algorithm.

Require: Dataset D, Number of clusters k

Randomly initialize the first cluster centroid c_1 from D
for $i = 2$ to k **do**
  Select c_i from D with probability proportional to the distance squared to the nearest existing centroid
end for
repeat
  for each datapoint x_i in D **do**
    Assign x_i to the closest centroid:

$$\arg \min_{c_j \in \{c_1, c_2, \ldots, c_k\}} \|x_i - c_j\|^2$$

  end for
  for $i = 1$ to k **do**
    Update c_i by calculating the average of all points assigned to cluster i
  end for
until No cluster assignments change

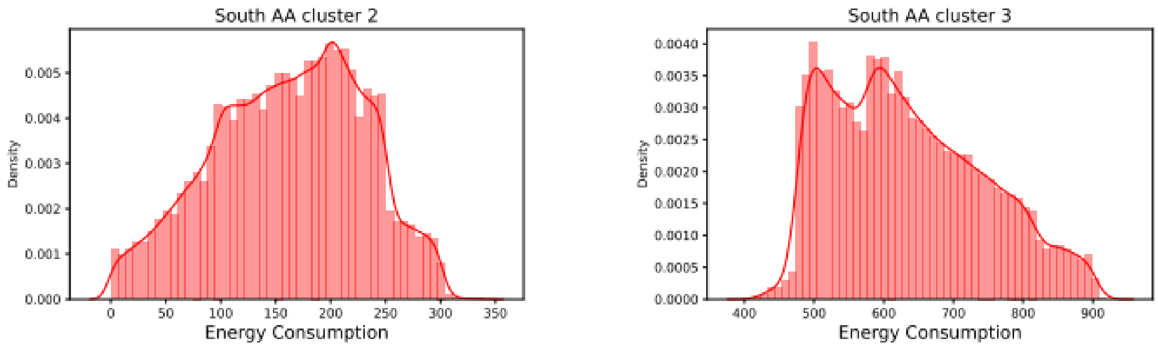
As indicated in Table 2, each district data is grouped into the best similar cluster, and each cluster has a different observation size. For example, in the south AA district (Cluter 1 = 111503, Cluster 2 = 152620, and Cluster 3 = 45183 observations). In the case of West AA data (Cluster 1 = 169800, Cluster 2 = 109227, and Cluster 3 = 34464 observations). Moreover, from Table 2, it has been indicated that the silhouette score for each cluster is > 0.5, which is higher, and the data points are correctly grouped in their proper cluster. Furthermore, the similarity (cohesion) of data points in a cluster is also very high as the larger silhouette score reveals the closeness of data points in a cluster. In general, the cluster validation results in Table 2 show that k-means clustering is a viable solution to characterize the energy consumption profiles and generate optimal clusters that will improve the prediction accuracy of the subsequent ensemble models.

Table 2: K-means clustering validation results

District	Dataset	#Cluster	Silhouette_Score	DBI_Score
South AA	309303	3	0.5478	0.5785
West AA	313492	3	0.550	0.593

A dataset with normal distribution has skewness and kurtosis values of 0 and 3, respectively. However, as Table 1 and Figure 3, Figure 4, and Figure 5 show, our dataset is positively skewed. This kind of asymmetrical data distribution and complicated energy consumption patterns [7] requires efficient data clustering and an ensemble deep learning model that can handle much better than the classical machine learning models and statistical techniques.

Moreover, k-means clustering results in Figure 7 illustrate that South AA data is grouped into 3 clusters of energy consumption profiles. Accordingly, Cluster 1 contains the medium size energy consumption profiles and the user's monthly energy usage is between 275kW and 575kW. Next to Cluster 1, Cluster 2 is indicated in the brown box and contains lower energy users; their monthly energy consumption is between 0.1kW and 275KW, but the largest observation or energy consumption profiles are grouped in this category. The last cluster consists of the group of consumption profiles that comprises the highest monthly energy consumption profiles whose monthly consumption revolves between 575 kW and 925 kW, but this group accommodates the smallest number of observations.



(a) Distribution of South AA Data in cluster 2.

(b) Distribution of South AA data in cluster 3.

Figure 3: Skewness test for South AA after clustering.

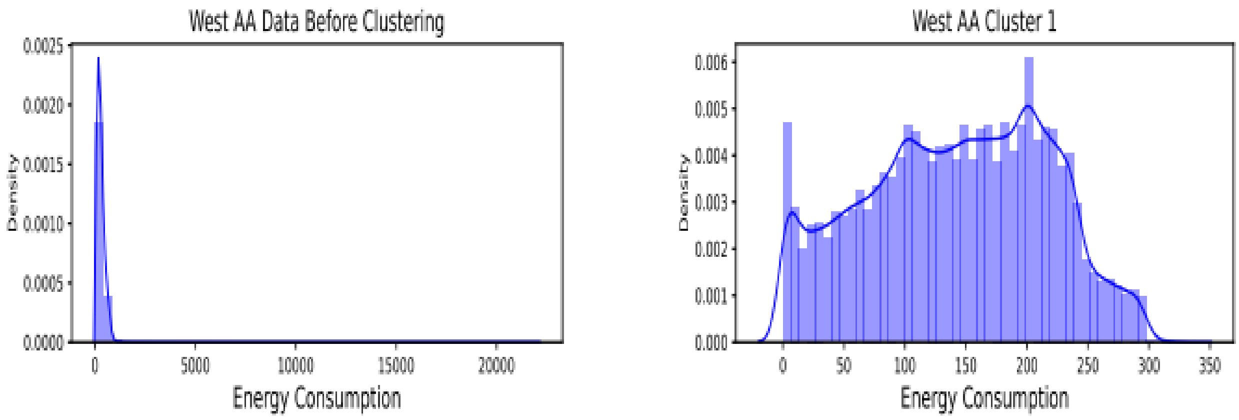


Figure 4: Skewness of West AA Data before Clustering and After Clustering

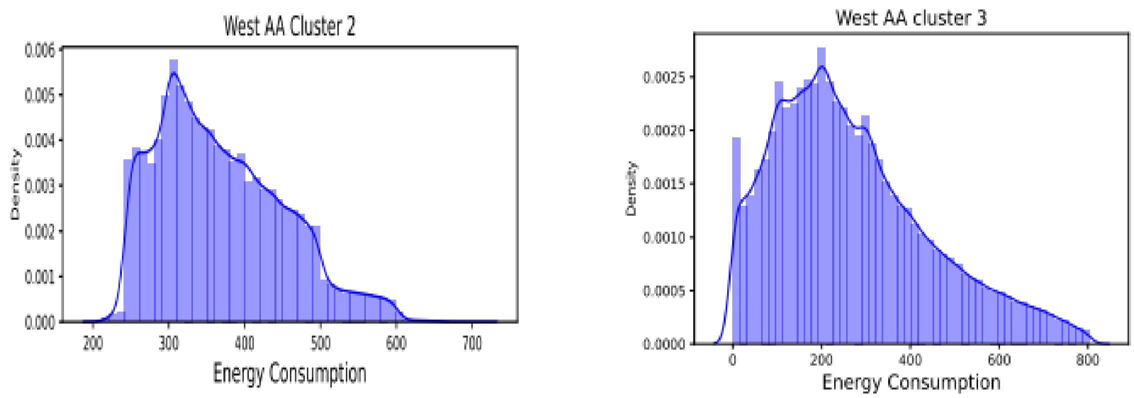


Figure 5: Skewness test for West AA data after clustering.

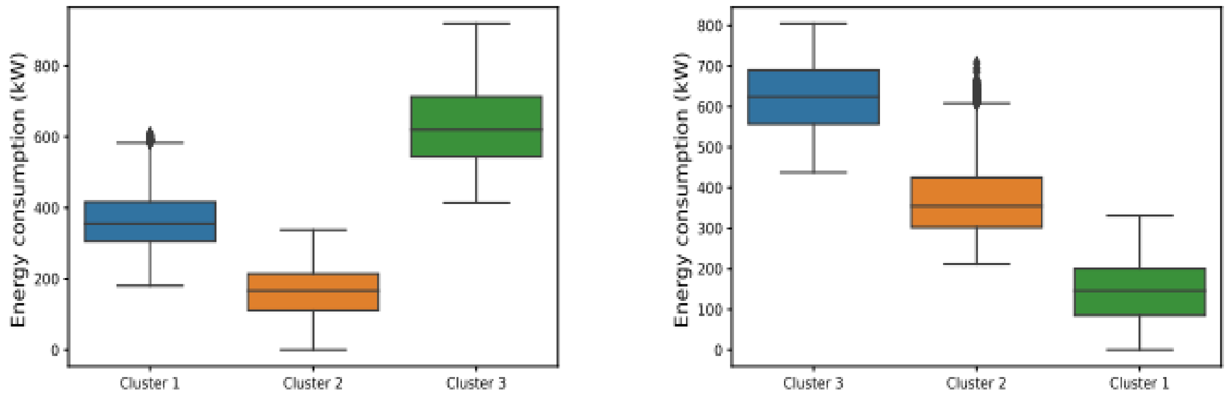


Figure 6: K-means clustering Results.

3.3 Deep Learning Model

Deep learning methods have gained greater attention because of their remarkable performance in image classification, natural language processing, and nonlinear electric load consumption prediction [9, 36]. Convolutional neural networks (CNN) [18], Long Short-Term Memory networks (LSTM) [8], and Gated Recurrent networks (GRU) [17] are the most widely used deep learning algorithms.

3.3.1 Long-Short Term Memory Neural Networks (LSTM)

Long short-term memory network (LSTM) is an improved version of a recurrent neural network frequently used in time forecasting and natural language processing. With its special memory cell, LSTM is capable of storing information involving long-range temporal dependencies [31]. LSTM network can establish long-term temporal correlation information and overcome vanishing gradient problems of RNNs networks as Figure 8 shows the sequential learning capabilities of LSTM. The self-connection of the LSTM memory block, referred to as the cell state, preserves (remembers) longer-range temporal dependencies of the data. Moreover, LSTM architecture is equipped with multiplicative gate modules which include an input gate, forget gate, and output gate [25]. The gate units are responsible for regulating the flow of information while sequential data processing is dealt with LSTM model.

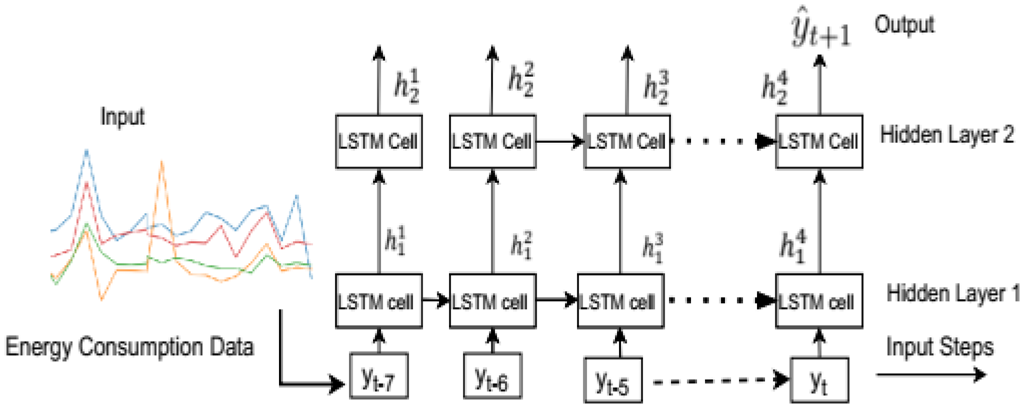


Figure 7: LSTM Sequential Learning Process

$$\hat{y}_{t+1} = f \sum (W_i x_i * b_i) \tag{7}$$

where the W_i and x_i are the updated weight vector and the input data respectively. Furthermore, f is the activation function.

3.3.2 Bidirectional LSTM (BiLSTM)

Bidirectional Long Short-Term Memory (BiLSTM) is the defamation of the LSTM algorithm which is capable of learning sequential data in both forward and backward directions as Figure 9 shows. It is more effective in learning the past and the future context information in two ways forward and backward directions allowing it to capture the context from both past and future information. This makes Bidirectional LSTM well-suited for tasks involving sequential data such as natural language processing and time series forecasting. The forward and backward operation of the BiLSTM can be expressed using Equations 8-10.

$$\vec{h}_t = \mathrm{LSTM}(x_t, \vec{h}_t) \tag{8}$$

$$\overleftarrow{h}_t = \mathrm{LSTM}(x_t, \overleftarrow{h}_t) \tag{9}$$

$$\hat{y}_{t+1} = w \rightarrow_y^h \vec{h}_t + W \leftarrow_{hy}, \overleftarrow{h}_t + b_y \tag{10}$$

BiLSTM trains two LSTM networks, where the first LSTM network processes the input sequence in the forward direction, and the second LSTM network processes in the reverse direction with a reversed copy of the input [15].

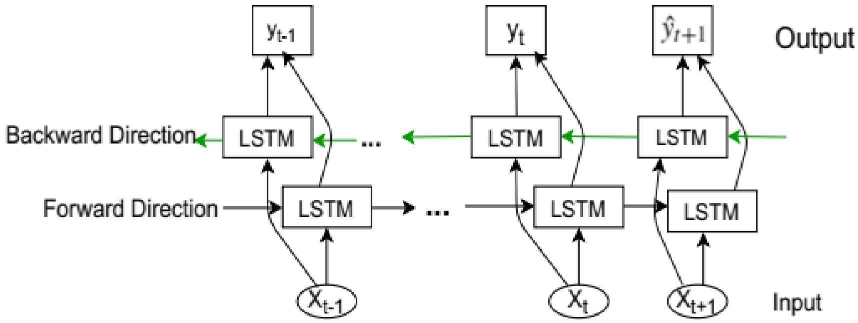


Figure 8: BiLSTM Structure

3.3.3 1D-Dimensional Convolution Neural Network (1D-CNN)

Model	Parameter name	Types / Range values	Optimal value selected
1D-CNN	Number filters	[32,64,128,256]	64
	Kernel size	[2,3,4,5]	3
	Pool type	[*MaxPooling*1*D*,*AveragePooling*1*D*]	MaxPooling1D
	Activation function	[*relu*,*tanh*,*Linear*]	relu
	Epoch	[40,80,120,160]	80
	Optimizer	[*RMSProp*,*Adam*,*Adadalta*]	Adam
	Batch size	[32,64,120,180]	32
	Learning rate	[0.0001,0.001,0.01,0.1]	0.001
BiLSTM	Activation function	[*relu*,*tanh*,*Linear*]	relu
	Dropout rate	[0.1,0.2,0.4,0.5]	0.2
	Optimizer	[*RMSProp*,*Adam*,*Adadalta*]	Adam
	Epoch	[40,80,120,200]	120
	Batch size	[32,64,128,256]	64
	Learning rate	[0.0001,0.001,0.01,0.1]	0.01
LSTM	Activation function	[*relu*,*tanh*,*Linear*]	tanh
	Dropout rate	[0.1,0.2,0.4,0.5]	0.2
	Epoch	[40,80,120,160]	160
	Batch size	[32,64,120,180]	64
	Optimizer	[*RMSProp*,*Adam*,*Adadalta*]	RMSProp
	Learning rate	[0.0001,0.001,0.01,0.1]	0.01

Convolutional neural networks (CNNS) are the most popular deep learning algorithms with similar human biological perception processing systems [18]. 1D-CNN is a special type of CNN network with powerful feature extraction and time series forecasting capabilities. 1DCNN is composed of three basic components

which include the convolution layer for feature extraction, the pooling layer for dimension reduction, and the neuron in fully connected layers use a weights matrix to apply a linear transformation to the input vector [12] as shown in Figure 10.

3.3.4 Ensemble Deep Learning Model

The ensemble model integrates multiple learning algorithms to obtain models that perform better than the single constituent base models. This means that the optimal and strong generalizable

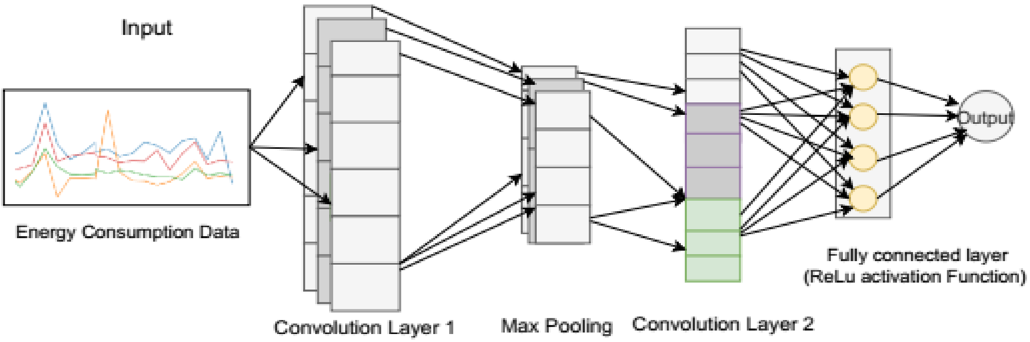


Figure 9: 1D-CNN Architecture

ensemble model can be obtained by combining multiple base models or algorithms. In this paper to build an ensemble deep learning model, first of all, each base models are trained with their optimal hyperparameter configuration as Table 3 shows the set of optimal hyperparameters of each base model. The optimal hyperparameter combination of each base model was determined using the Bayesian optimization (BO) algorithm. Bayesian optimization algorithm is a metaheuristic hyperparameter tuning approach based on the probability model of the global function which aims to intelligently identify the optimal combination of hyperparameters with reasonable computation time.

Table 3: Hyperparameters tuning for selected deep learning models.

After the optimal clusters have been generated, the observations in each cluster are divided into training and testing sets to train the proposed model with each cluster's data as Figure 10 illustrates. Despite the varying dataset size, 70% of the instances comprising the first 16 months of load consumption data of each cluster was used for model training by keeping the sequential order of the monthly energy consumption is relevant. Then, the remaining 30% approximately the final four months of the dataset was used to test the model's performance. This is because we aimed to predict the next four months' aggregate monthly load consumption using the previous 16 months' load consumption data. The pseudocode detailed in algorithm 2 shows the proposed model execution process, including data partitioning into training and test datasets.

Thus, given LSTM, BiLSTM, and CNN-GRU are chosen algorithms to build the base models, then each model prediction output can be represented by ˆ $m1$, ˆ $m2$ and ˆ $m3$, respectively. Similarly, their respective weights can be generated as $w1$, $w2$ and $w3$, for ˆ $m1$, ˆ $m2$ and ˆ $m3$ models, respectively, with different weight values based on the proportion of the performance of each base model that will yield an optimal ensemble model. The weight value for each base model should be a small fraction number ranging from 0 to 1. This means that the sum of the weight values of all base models should be ≤ 1. Then, the final weighted average ensemble model can be, represented as WAE, and be found by merging base models as illustrated in Equation 12.

$$WAE = \big((w_1 * \hat{m}_1) + (w_2 * \hat{m}_2) + (w_3 * \hat{m}_3)\big) \tag{11}$$

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 1 (2025)

Where w_1, w_2 and w_3 are the weights for $\hat{m}_1, \hat{m}_2$ and $\hat{m}_3$ models respectively.

Algorithm 2 Pseudocode for weighted ensemble model training

```
Input: Training dataset D, hyperparameters for individual Models
Output: prediction
{Split and Reshape Data}
n ← length(D)
Train_set ← D[0 : n × 0.7], Test_set ← D[n × 0.7 : n]
Xtrain ← reshape(Train_set), Xtest ← reshape(Test_set)
{Train and Load Models}
models ← list()
for model_type in [LSTM, BiLSTM, 1D-CNN] do
    model ← Sequential()
    model.add(model_type(...))
    for each epoch in 1 to n_epochs do
        model.fit(Xtrain, Ytrain, epochs=n_epochs, batch_size=n_batch)
    end for
    model.save(”my_model_” + model_type)
    loaded_model ← load_model(”my_model_” + model_type)
    models.append(loaded_model)
end for
{Evaluate Individual Models and Combine Predictions}
Predictions ← list()
for each model in models do
    Prediction ← model.predict(X_test)
    Predictions.append(Prediction)
end for
{Weighted Average Ensemble (WAE)}
Weights ← [w1, w2, w3]
WAE ← tensordot(Predictions, Weights, axes=((0),(0)))
Evaluate WAE using MSE, MAE, MAPE, RMSE
```

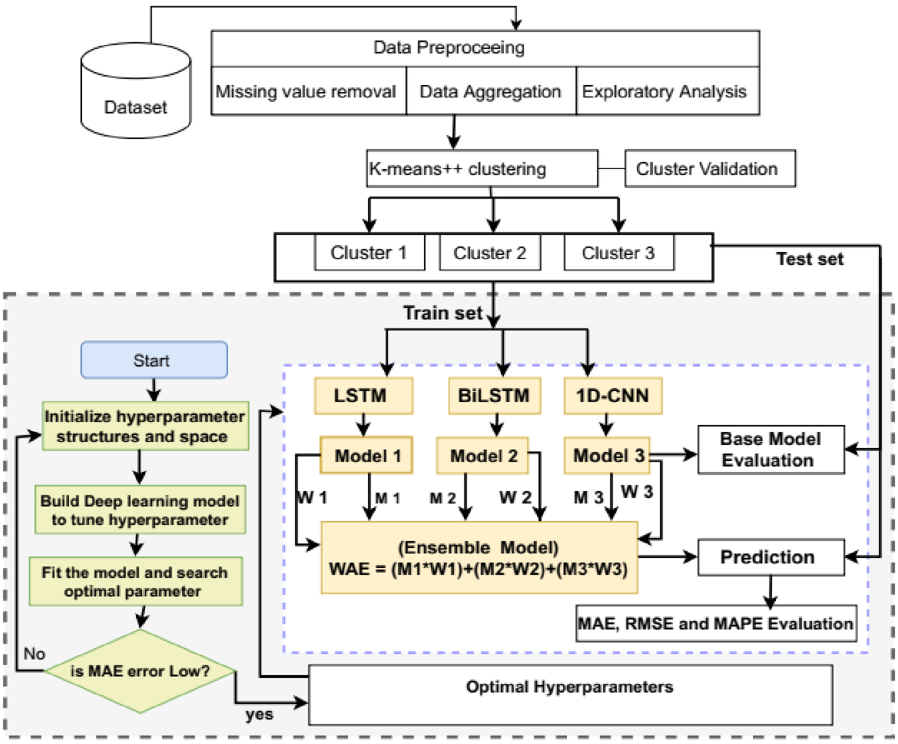


Figure 10: The Proposed Ensemble Model

3.3.5 Performance Evaluation

Whether machine learning models are developed for classification or regression tasks, several metrics can be used to quantitatively evaluate their efficacy and performance.

1. Mean Absolute Error (MAE): MAE is a statistical metric that can be used to quantify the discrepancies between the expected and target values [4].
2. Root Mean Square Error (RMSE): RMSE is one of the popular statistical metrics employed to calculate the difference between the actual and expected predicted values in regression model evaluation as expressed in Equation 13.
3. Mean Absolute Percentage Error (MAPE): The MAPE is used to calculate the percentage difference between the actual and expected values of load consumption data. [3]. The mathematical formula for MAPE is given in Equation 14.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y - \hat{y}| \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y - \hat{y})^2} \tag{13}$$

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \frac{|y - \hat{y}|}{y} * 100 \tag{14}$$

where y and $\hat{y}$ represent the actual and predicted values respectively. Furthermore, n denotes the total number of data that were utilized in the model evaluation process.

3.4 Results Analysis

This section provides the experimental results discussion and comparative analysis of the proposed ensemble deep learning model and the base models using different case study data, which includes South AA and West AA.

Table 4: Comparison of the proposed and baseline deep learning models on South AA data.

Data	Algorithm	MAE	RMSE	MAPE(%)
Cluster 1	1D-CNN	56.097	65.869	0.195
	GRU	63.499	74.679	0.210
	LSTM	56.658	66.646	0.179
	BiLSTM	57.904	67.777	0.186
	Ensemble	54.472	65.509	0.160
Cluster 2	1D-CNN	60.916	73.170	3.154
	GRU	51.953	63.904	3.180
	LSTM	51.965	64.054	3.195
	BiLSTM	59.711	69.964	0.195
	Ensemble	51.389	62.385	3.033
Cluster 3	1D-CNN	99.237	116.411	0.164
	GRU	103.903	120.980	0.175
	LSTM	99.014	116.439	0.163
	BiLSTM	97.567	117.137	0.146
	Ensemble	98.082	121.360	0.151

Table 4 shows the proposed model performance in comparison with base models (1DCNN, LSTM, BiLSTM) models using MAE, RMSE, and MAPE metrics. Accordingly, the proposed ensemble model shows superior performance than the base models on cluster 1 data with lower prediction error values 54.472, 65.509, and 0.16o for MAE, RMSE, and MAPE respectively. Similarly, the ensemble model outperforms the base models on cluster 2 data based on MAE(51.389) and RMSE(62.385) regardless of the larger MAPE values. But, in the case of cluster 3 data, BiLSTM has achieved the best performance with lower 97.567 and 0.146 error values for MAE and MAPE respectively.

Table 5: Model Performance comparison on Training and Test set with clustering and without clustering, South AA case study data

Data	Algorithm	Training		Test	
		MAE	RMSE	MAE	RMSE
Clustered	1D-CNN	58.358	70.057	60.916	73.170
	LSTM	58.579	70.123	51.965	64.054
	BiLSTM	61.927	74.520	59.711	69.964
	Ensemble	58.287	69.975	51.389	62.385
Un-Clustered	1D-CNN	178.995	256.540	182.874	266.111
	LSTM	162.288	250.252	167.094	261.151
	BiLSTM	174.277	255.662	179.108	261.151

	Ensemble	163.369	250.762	168.630	260.684

Table 5 summarizes the performance of training and test data accuracy on post-clustering and without-clustering energy consumption data. The results have demonstrated that the proposed ensemble model and the base models have achieved lower MAE and RMSE prediction errors on post-cluster data about the training and test accuracy. Moreover, the MAE and RMSE errors on test data are much lower compared with the training errors. From these results, we can conclude that integrating k-means++ clustering with deep learning methods enables the model to learn the new data very well and make better generalizations [20, 33]. Moreover, the Post-clustering based ensemble model demonstrates superior performance with a significant prediction error decrease of MAE (64.321%) and RMSE (72.095%) on training data, and MAE(69.525%) and RMSE (76.068%) on test data, as compared to the performance obtained without clustering.

Overall, the results show that the best prediction accuracy and the lowest MAE, RMSE, and MAPE errors are obtained when multiple deep learning techniques are combined and utilized in the form of an ensemble model [28, 37], coupled with hyperparameter optimization on post-clustering data.

Table 6: Comparison of the proposed model and baseline deep learning models on West AA data.

Data	Algorithm	MAE	RMSE	MAPE(%)
Cluster 1	1D-CNN	58.071	69.092	12.097
	GRU	58.538	69.884	12.418
	LSTM	58.039	69.079	12.086
	BiLSTM	58.487	69.758	12.312
	Ensemble	57.584	68.136	11.601
Cluster 2	1D-CNN	63.511	74.157	0.197
	GRU	67.750	78.317	0.218
	LSTM	62.458	76.379	0.180
	BiLSTM	65.867	76.337	0.210
	Ensemble	62.335	73.765	0.181
Cluster 3	1D-CNN	69.394	87.537	0.113
	GRU	69.007	87.457	0.115
	LSTM	69.138	88.323	0.110
	BiLSTM	70.374	90.007	0.171
	Ensemble	68.581	85.730	0.112

Moreover, Table 6 summarizes the performance of the proposed model and the base models on the West AA case study data. From this result, the ensemble model outperforms the base models in the case of cluster 1 data showing a significant error decrease of MAE(1.543%), and RMSE(2.325%) compared to the BiLSTM which is the worst-performing base model.

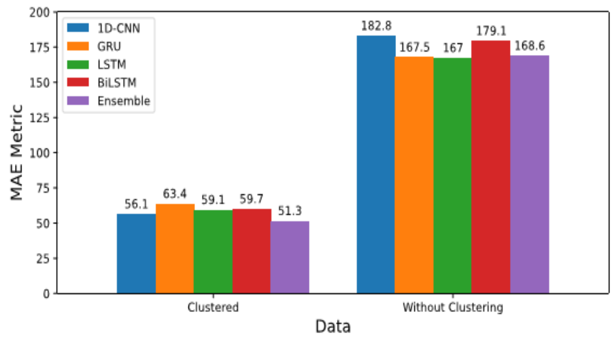


Figure 11: MAE Values for Clustered and Un-clustered, South AA Data

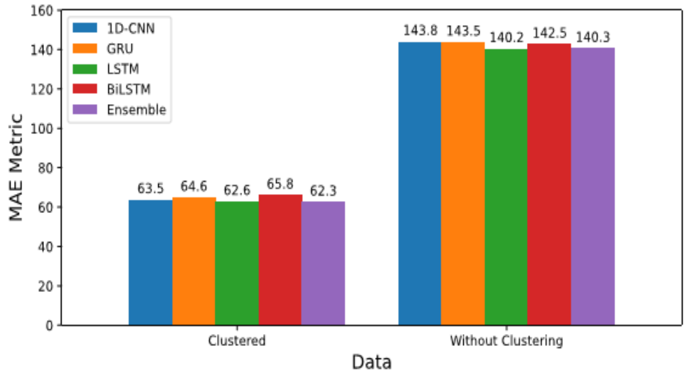


Figure 12: MAE Values for Clustered Vs Un-clustered, West AA Data

Figure 11 and Figure 12, shows that the performance of the ensemble model significantly improved in the case of cluster generated data with MAE(69.696%) error decrease south AA case study data and MAE(55.595%) error decrease on the west AA data. From these results, we can conclude that the potential of k-means clustering to find the optimal clusters of the energy data significantly contributes to the ensemble deep model to learn complex energy data effectively and exhibits lower prediction errors in the new dataset compared to the model performance on un-cluster data concerning MAE performance metric. To generalize the training of deep learning and their ensemble models with post-clustering data affirms the better performance. This is because in addition to outlier treatment, clustering of highly variable energy consumption data [17, 34] into more similar consumption patterns enables the proposed model to learn the detailed features of the input data.

4 Conclusion

In this study, the effectiveness of deep learning models (1D-CNN, LSTM, BiLSTM and GRU) and ensemble model is investigated for aggregate energy consumption prediction focusing on the residential users category. The model's performance was assessed on both the un-cluster and post-clustered energy datasets. The integration of k-means clustering with an ensemble model to find the optimal cluster that minimizes the high variability and complexity of energy consumption data has been investigated. The viability of the clustering technique to group energy consumption data into a more similar consumption profile was validated and promising results were found which has enabled the ensemble deep model to learn the complete and intrinsic nature of the energy consumption data. Hence, the integration of clustering approach with deep learning and ensemble techniques significantly improves the prediction performance of the proposed model with very low prediction errors when compared to the performance obtained without clustering. Furthermore, while properly combining the capabilities of multiple deep learning algorithms, results indicate that the proposed ensemble model has outperformed the optimal base model performance in all case study data sets used in this study. In addition, enhanced by the metaheuristic Bayesian based hyperparameter tuning method, the proposed ensemble deep learning model has demonstrated the best performance and better generalization abilities without facing the problem of model overfitting while the trained model is exposed to test data.

Overall, the ensemble model proposed in this study has demonstrated better capabilities for learning the complex energy consumption data and provides a significant MAE, RMSE and MAPE error decrease in both case study data as compared to base algorithms, i.e., LSTM, BiLSTM and 1D-CNN performance. In the future, the income level and family size information about the customers should be incorporated as exogenous variables to enhance the prediction accuracy of the energy consumption demand. Additionally, optimal time steps should be determined using automatic optimization methods.

References

[1] Musaed Alhussein, Khursheed Aurangzeb, and Syed Irtaza Haider. Hybrid cnn-lstm model for short term individual household load forecasting. *Ieee Access*, 8:180544–180557, 2020.

[2] Mengmeng Cai, Manisa Pipattanasomporn, and Saifur Rahman. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. *Applied Energy*, 236:1078–1088, 2019.

[3] Yaogang Chen, Guoyin Fu, and Xuefeng Liu. Air-conditioning load forecasting for prosumer based on meta ensemble learning. *IEEE Access*, 8:123673–123682, 2020.

[4] Jui-Sheng Chou, Dinh-Nhat Truong, and Ching-Chiun Kuo. Imaging time-series with features to enable visual recognition of regional energy consumption by bio-inspired optimization of deep learning. *Energy*, 224:120100, 2021.

[5] Behnam Farsi, Manar Amayri, Nizar Bouguila, and Ursula Eicker. On short-term load forecasting using machine learning techniques and a novel parallel deep lstm-cnn approach. *IEEE Access*, 9:31191– 31212, 2021.

[6] Dawit Habtu Gebremeskel, Erik O Ahlgren, and Getachew Bekele Beyene. Long-term evolution of energy and electricity demand forecasting: The case of ethiopia. *Energy Strategy Reviews*, 36:100671, 2021.

[7] Haibo Guo, Lingling Tang, and Yuexing Peng. Ensemble deep learning method for shortterm load forecasting. In *2018 14th International Conference on Mobile Ad-Hoc and Sensor Networks (MSN)*, pages 86–90. IEEE, 2018.

[8] Ejigu Tefera Habtemariam, Kula Kekeba, Mar´ıa Mart´ınez-Ballesteros, and Francisco Mart´ınez-Alvarez. A bayesian optimization-based lstm model for wind power forecast- ´ ing in the adama district, ethiopia. *Energies*, 16(5):2317, 2023.

[9] Ejigu T Habtermariam, Kula Kekeba, Alicia Troncoso, and Francisco Mart´ınez-Alvarez. ´ A cluster-based deep learning model for energy consumption forecasting in ethiopia. In *International Workshop on Soft Computing Models in Industrial and Environmental Applications*, pages 423–432. Springer, 2022.

[10] D. Hadjout, J. F. Torres, A. Troncoso, A. Sebaa, and F. Mart´ınez-Alvarez. Electricity ´ consumption forecasting based on ensemble deep learning with application to the algerian market. *Energy*, 243:123060, 2022.

[11] Ghulam Hafeez, Khurram Saleem Alimgeer, and Imran Khan. Electric load forecasting based on deep learning and optimized by heuristic algorithm in smart grid. *Applied Energy*, 269:114915, 2020.

[12] Ying-Yi Hong, Jonathan V. Taylar, and Arnel C. Fajardo. Locational marginal price forecasting in a day-ahead power market using spatiotemporal deep learning network. *Sustainable Energy, Grids and Networks*, 24:100406, 2020.

[13] Idowu David Ibrahim, Y Hamam, Yasser Alayli, Tamba Jamiru, Emmanuel Rotimi Sadiku, Williams Kehinde Kupolati, Julius Musyoka Ndambuki, and Azunna Agwo Eze. A review on africa energy supply through renewable energy production: Nigeria, cameroon, ghana and south africa as a case study. *Energy Strategy Reviews*, 38:100740, 2021.

[14] Muhammad Ishaq, Soonil Kwon, et al. Short-term energy forecasting framework using an ensemble deep learning approach. *IEEE Access*, 9:94262–94271, 2021.

[15] K. U. Jaseena and B. C. Kovoor. Decomposition-based hybrid wind speed forecasting model using deep bidirectional lstm networks. *Energy Conversion and Management*, 234:113944, 2021.16

[16] Waqas Khan, Shalika Walker, and Wim Zeiler. Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy*, 240:122812, 2022.

[17] Pratima Kumari and Durga Toshniwal. Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production*, 318:128566, 2021.

[18] Liying Liu and Yain-Whar Si. 1d convolutional neural networks for chart pattern classification in financial time series. *The Journal of Supercomputing*, 78(12):14191–14214, 2022.

[19] Haris Mansoor, Huzaifa Rauf, Muhammad Mubashar, Muhammad Khalid, and Naveed Arshad. Past vector similarity for short term electrical load forecasting at the individual household level. *IEEE Access*, 9:42771–42785, 2021.

[20] F. Mateo, J. J. Carrasco, A. Sellami, M. Millan-Giraldo, M. Dom ´ ´ınguez, and E. SoriaOlivas. Machine learning methods to forecast temperature in buildings. *Expert Systems with Applications*, 40(4):1061–1068, 2013.

[21] Manohar Mishra, Janmenjoy Nayak, Bighnaraj Naik, and Ajith Abraham. Deep learning in electrical utility industry: A comprehensive review of a decade of research. *Engineering Applications of Artificial Intelligence*, 96:104000, 2020.

[22] Tiago Pinto, Isabel Prac¸a, Zita Vale, and Jose Silva. Ensemble learning for electricity consumption forecasting in office buildings. *Neurocomputing*, 423:747–755, 2021.

[23] Zahra Qavidelfardi, Mohammad Tahsildoost, and Zahra Sadat Zomorodian. Using an ensemble learning framework to predict residential energy consumption in the hot and humid climate of iran. *Energy Reports*, 8:12327–12347, 2022.

[24] Abdulwahed Salam and Abdelaaziz El Hibaoui. Energy consumption prediction model with deep inception residual network inspiration and lstm. *Mathematics and Computers in Simulation*, 190:97–109, 2021.

[25] MyungJae Shin, David Mohaisen, and Joongheon Kim. Bitcoin price forecasting via ensemble-based LSTM deep learning networks. In *2021 International conference on information networking (ICOIN)*, pages 603–608. IEEE, 2021.

[26] SN Singh, Abheejeet Mohapatra, et al. Data driven day-ahead electrical load forecasting through repeated wavelet transform assisted svm model. *Applied Soft Computing*, 111:107730, 2021.

[27] Nivethitha Somu, Gauthama Raman MR, and Krithi Ramamritham. A hybrid model for building energy consumption forecasting using long short term memory networks. *Applied Energy*, 261:114131, 2020.

[28] Hui Song, Alex Kai Qin, and Flora D Salim. Evolutionary multi-objective ensemble learning for multivariate electricity consumption prediction. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[29] Zhiwei Song, Zhaojing Cao, Can Wan, and Shenglan Xu. An ensemble wavelet deep learning approach for short-term load forecasting. In *2019 IEEE Innovative Smart Grid Technologies-Asia (ISGT Asia)*, pages 1205–1210. IEEE, 2019.17

[30] Mao Tan, Siping Yuan, Shuaihu Li, Yongxin Su, Hui Li, and Feng He. Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning. *IEEE Transactions on Power Systems*, 35(4):2937–2948, 2019.

[31] J. F. Torres, F. Mart´ınez-Alvarez, and A. Troncoso. A deep LSTM network for the Spanish ´ electricity consumption forecasting. *Neural Computing and Applications*, 34(13):10533– 10545, 2022.

[32] Amin Ullah, Kilichbek Haydarov, Ijaz Ul Haq, Khan Muhammad, Seungmin Rho, Miyoung Lee, and Sung Wook Baik. Deep learning assisted buildings energy consumption profiling using smart meter data. *Sensors*, 20(3):873, 2020.

[33] Lingxiao Wang, Shiwen Mao, and Bogdan Wilamowski. Short-term load forecasting with lstm based ensemble learning. In *2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData)*, pages 793–800. IEEE, 2019.

[34] Yi Wang, Dahua Gan, Mingyang Sun, Ning Zhang, Zongxiang Lu, and Chongqing Kang. Probabilistic individual load forecasting using pinball loss guided lstm. *Applied Energy*, 235:10–20, 2019.

[35] Lulu Wen, Kaile Zhou, and Shanlin Yang. Load demand forecasting of residential buildings using a deep learning model. *Electric Power Systems Research*, 179:106073, 2020.

[36] Yu Yang, Fan Jinfu, Wang Zhongjie, Zhu Zheng, and Xu Yukun. A dynamic ensemble method for residential short-term load forecasting. *Alexandria Engineering Journal*, 63:75–88, 2023.

[37] Shuai Zhang, Yong Chen, Wenyu Zhang, and Ruijun Feng. A novel ensemble deep learning model with dynamic error correction and multi-objective ensemble pruning for time series forecasting. *Information Sciences*, 544:427–445, 2021.