

Electricity consumption forecasting using DFT decomposition based hybrid ARIMA-DLSTM model

Osman Yakubu¹, Narendra Babu C.²

¹Department of Computer Science, Garden City University College, Kumasi, Ghana

^{1,2}Department of Computer Science and Engineering, M S Ramaiah University of Applied Sciences, Bangalore, India

Article Info

Article history:

Received Jun 16, 2021

Revised Sep 8, 2021

Accepted Sep 16, 2021

Keywords:

ANN

ARIMA

DLSTM

Forecasting

Time series

ABSTRACT

Forecasting electricity consumption is vital, it guides policy makers and electricity distribution companies in formulating policies to manage production and curb pilfering. Accurately forecasting electricity consumption is a challenging task. Relying on a single model to forecast electricity consumption data which comprises both linear and nonlinear components produces inaccurate results. In this paper, a hybrid model using autoregressive integrated moving average (ARIMA) and deep long short-term memory (DLSTM) model based on discrete fourier transform (DFT) decomposition is presented. Aided by its superior decomposition capability, filtering using DFT can efficiently decompose the data into linear and nonlinear components. ARIMA is employed to model the linear component, while DLSTM is applied on the nonlinear component; the two predictions are then combined to obtain the final predicted consumption. The proposed techniques are applied on the household electricity consumption data of France to obtain forecasts for one day, one week and ten days ahead consumption. The results reveal that the proposed model outperforms other benchmark models considered in this investigation as it attained lower error values. The proposed model could accurately decompose time series data without exhibiting a performance degradation, thereby enhancing prediction accuracy.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Osman Yakubu

Department of Computer Science

Garden City University College

P.O. Box KS 12775, Kumasi, Ghana

Email: Osman.yakubu@gcuc.edu.gh

1. INTRODUCTION

Modern life is dependent on electricity; businesses and household consumers are heavily reliant on electricity for their daily energy requirements. Accurately forecasting electricity consumption, which falls under the domain of time series predictions, continues to engage the attention of researchers who have employed various prediction techniques. In both developed and developing economies, governments and their agencies emphasize the importance of predicting electricity consumption precisely for policy in [1]. Erroneously predicting electricity consumption has a negative financial impact [1]. Accuracy is the most significant factor considered in choosing a suitable forecasting model [2], [3]. Electricity consumption data are a time series since it is gathered at time intervals. In forecasting time series, past observations of the same variable are gathered for analysis to develop a model to be used for predicting future values [4]. Most real-life time series datasets including weather, electricity consumption, and oil field production have the sequential property of time. So finding an effective technique for trend forecasting continues to be a

challenge [5], [6]. Owing to its unique properties, time series forecasting accuracy is one of the most challenging problems in data mining [7]. Time series data can show diverse components such as trends, seasonality, and jumps [8]. Single models that have been used for time-series forecasting have been found deficient in forecasting accuracy [4]. Hybrid models have been identified as the best in terms of forecasting accuracy [4], [9]. In a comprehensive review by [2], it was discovered that hybrid methods are the most suitable and accurate methods for forecasting time series. Parallel-series hybrid structures could also provide more precise and promising results compared to only hybrid methods [2]. Time series data such as electricity consumption could be composed of linear and nonlinear components, and a single model may not provide an accurate forecast autoregressive integrated moving average (ARIMA) has been widely used in forecasting the linear components of time series [4], [9], and it is widely acknowledged that artificial neural networks (ANNs) model the nonlinear components of time series more accurately [10]. ARIMA is a traditional statistical method used in time series forecasting [11]. Effective techniques capable of splitting the electricity consumption data into linear and nonlinear components to be modeled differently and combined for an effective forecast, are therefore desired.

Motivated by the possibility of modelling linear and non-linear components of time series data separately to obtain more accurate predictions, we propose a hybrid autoregressive integrated moving average-deep long short term memory (ARIMA-DLSTM) model based on discrete fourier transform (DFT) to forecast daily electricity consumption. The proposed ARIMA-DLSTM model is employed to obtain one-step-ahead, one-week-ahead, and 10-days-ahead predictions. Using root mean square error (RMSE), mean absolute percentage error (MAPE), and mean absolute error (MAE), the proposed model is compared with six reference models. The proposed ARIMA-DLSTM model achieved a lower error measure than the reference models, inferring that it outperforms the referenced models in prediction accuracy because the electricity consumption data were decomposed more accurately.

The rest of the paper is organized as follows, in section 2, related work is discussed, and section 3 contains the proposed ARIMA-DLSTM model. Section 4 describes the details of the experimental setup followed by the experimental results which are presented and discussed in section 5. The paper is concluded in section 6.

2. RELATED WORK

A number of approaches have been presented to overcome the challenges associated with time-series forecasting, in general, and electricity consumption time series forecasting in particular. These proposed methods can be classified as statistical and soft computing techniques. ARIMA and artificial neural networks (ANN) models are explored by [9] and a novel hybrid ARIMA-ANN model is devised for the forecasting of time series data. The volatile nature of time series is explored by decomposing the time series data into linear and nonlinear components using an MA filter that achieved better prediction accuracy than individual ARIMA and ANN models and the hybrid ARIMA-ANN model presented in [4]. In [10], an ANN model based on a multilayer perceptron (MLP) with a backpropagation training algorithm, which is used as the neural network topology was proposed to predict the electricity consumption of Turkey. Tangent-sigmoid and pure-linear transfer functions were selected in the hidden and outer layer processing elements, respectively. A novel deep learning forecasting model that exploits the ability of convolutional layers to extract useful knowledge and the efficacy of long short-term memory (LSTM) layers for the identification of short-term and long-term dependencies for accurate forecasting of gold prices was presented by [12]. They contend that the use of LSTM layers together with additional convolutional layers could enhance forecasting performance. In financial time series, accurately forecasting stock prices is a challenging task because there are drawbacks, such as the unsuitability of some models that do not follow statistical assumptions and stock data that are very noisy. In [13], a hybrid time series adaptive network based on a fuzzy inference system (ANFIS) that uses empirical mode decomposition was proposed to address these drawbacks. Time series could be purely linear, purely nonlinear, or hybrid and any proposed model should be able to recognize this pattern for better prediction. In [14], a new hybrid approach using a combination of linear and nonlinear exponential smoothing models from the innovation state space (ETS) with artificial neural networks (ETS-ANN) that enhance the possibility of picking up the varied combinations of linear and/or nonlinear components inherent in time series was proposed.

To predict short-term residential electricity consumption in China, a hybrid holt-winters (HW) method and an extreme learning machine (ELM) network was proposed by [15]. The original data are decomposed into linear and nonlinear components using a moving average filter. The HW method is used to model the linear component, while ELM is used to forecast the nonlinear model, and the results are combined to predict 15-minute electricity consumption. Further, partitioning and interpolation (PI) based ARIMA-ANN model [16] outperformed MA filter based ARIMA-ANN technique in forecasting time series data. In [17],

a novel time series prediction model termed SeriesNet, capable of completely learning time series data features at different intervals was presented. The model is capable of learning multi-range and multi-level features from time series and improves generalization by adopting residual learning and batch normalization, resulting in a higher predictive ability. In [18], a technique based on deep learning to deal with the challenges of forecasting big data time series was also proposed. The deep forward neural network of the H2O big data analysis framework is used where the problem at hand is split into forecasting sub-problems. A hybrid system that looks for an appropriate function to combine linear and nonlinear model forecasts was presented by [19]. An ARIMA model is used on the linear component, and MLP and support vector regression model, two intelligent nonlinear models, are used on the nonlinear component.

ANNs have been used widely in energy prediction for a long time. To improve the prediction accuracy for electricity consumption in the short term, a type of optimized ANN model for predicting hourly building electricity consumption was presented in [19]. A deep learning-based framework for electricity demand forecasting that considers long-term historical dependencies was also presented in [20]. Cluster analysis is initially executed on the electricity consumption data for all months, and season-based segmented data are generated. Load trend characterization is performed to obtain a deeper understanding of metadata by evaluating each of the clusters. Further, they trained LSTM multi-input and output models to predict the demand for electricity depending on the season, day, and interval. To improve the prediction results, they incorporated the concept of moving-window-based active learning. An LSTM-based short-term time-phased electricity consumption prediction model with an attention mechanism was also proposed by [21]. The technique assigns a weight coefficient to the input sequence data. Then, the output data of each cell of the LSTM are computed using the forward propagation technique. The backpropagation technique was used to calculate the error between the real and predicted values. To forecast electricity consumption accurately, a technique of probability density forecasting based on the least absolute shrinkage and selection operator and the quantile regression neural network (LASSO-QRNN) was presented in [22]. To forecast precisely the total and industrial electricity consumption of China, a modified gray prediction model that combines a new initial condition and a rolling mechanism was proposed in [23]. They employed a particle swarm optimization algorithm to generate parameters that can be determined optimally in accordance with various features of the input data. An integrated approach for the 24 hour-ahead prediction of electricity consumption in buildings using the hilbert–huang transform (HHT), regrouping particle swarm optimization, and ANFIS was proposed by [24]. There are complexities and uncertainties in the electricity system. A self-adaptive gray fractional weighted model for the accurate prediction of the city of Jiangsu's electricity consumption was also presented by [25].

The short-falls of ARIMA in accurately predicting time series data that have a blend of linear and nonlinear components compelled facebook to introduce prophet, which is an open-source prediction model based on decomposable models for data with trends, seasonality, and holidays. It is reputed to forecast time series more accurately by employing simple, intuitive parameters and can include custom seasonality and holidays [26]. The use of an improved particle swarm optimization in designing a hybrid prediction model for future prediction of electricity consumption can significantly improve prediction accuracy [27]. According to [28], forecasting electricity consumption in households is challenging, they thus developed an enhanced support vector regression model with the capability of forecasting household electricity consumption through several intervention approaches. A support vector machine (SVM) model was also created in [29] to accurately forecast energy consumption in hotel buildings. The capability and precision of supervised machine learning techniques was assessed in [30] and Gaussian process regression (GPR) was found to offer better predictions of energy consumption of office buildings. One of the most effective methods employed in the classification of time series data is recurrent neural network (RNN) [31], and has been used in forecasting stock prices [32]. ANNs have also been used in the improvement of forecasting for the prices of gold [33]. Convolutional neural networks has been proven to offer slightly better prediction accuracy compared to LSTM [34]. Time series data related to especial finance could exhibit features of volatility, it could also in addition, be non-linear and non-stationary, as it is susceptible to economic factors that are external [35]. Notwithstanding the numerous models proposed to accurately predict electricity consumption, enhancement of the prediction performance of electricity consumption models is still required. Several presented models are single models that are either statistical, such as ARIMA, or soft computing, such as ANN, which in the estimation of [4] do not generate accurate predictions when used individually. The hybrid models presented also require greater accuracy, as the techniques employed degrade performance when decomposing the data into linear and nonlinear components in all the perused literature. Therefore, more enhanced techniques should be discovered that accurately decompose electricity consumption time series data into its linear and nonlinear components, and robust statistical and soft computing models could be hybridized for more reliable predictions.

In summary, a significant number of research on enhancing time series prediction accuracy has been explored. All the research explored have confirmed that single models lack the capability to produce accurate forecast since they assume that the data are purely linear or purely nonlinear and hybrid models are required. Hybrid models require the decomposition of the data into linear and nonlinear components. The decomposition techniques continue to present challenges as quality of predictions are degraded due to ineffective decomposition techniques. To ensure a more accurate forecast of electricity consumption, superior decomposition techniques that do not degrade the quality of predictions are thus desired.

3. PROPOSED MODEL

In this paper, a new hybrid ARIMA-DLSTM prediction model based on DFT decomposition is presented. Electricity consumption time series have trends and seasonality; consumption for industry is very high during particular times and days, such as peak working hours, and very low at other times, such as break times and holidays. The consumption of domestic users is higher on weekends and holidays when they are home, while their consumption is low when they are at work. This implies that electricity consumption data are composed of linear and nonlinear components. Industrial and domestic electricity have the same characteristics. Techniques must be employed to segregate precisely the linear and nonlinear components of the data to model them for prediction. In the hybrid models presented by [4] and [36], the assumption is that the data are a sum of linear and nonlinear components. However, the data are not decomposed into components of linear and nonlinear; instead, an ARIMA model is applied directly to the data, and the residual error is modeled as a nonlinear component, which degrades performance. The methodology of the proposed ARIMA-DLSTM electricity consumption prediction model is shown in Figure 1(a), Figure 1(b) also shows the architecture of the DLSTM recurrent network.

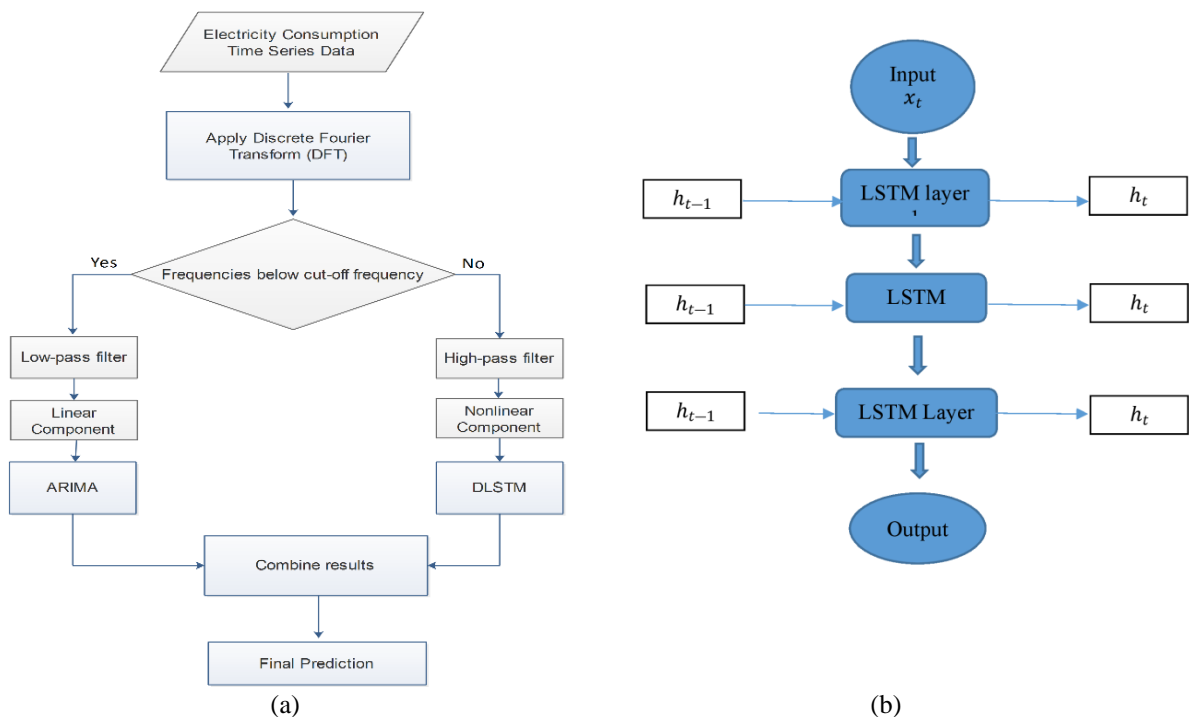


Figure 1. These figures are; (a) Proposed prediction method, (b) DLSTM recurrent network architecture

Using DFT, the electricity consumption time series is converted into Fourier coefficients. First, a low-pass filter is used to isolate Fourier coefficients that are below the cut-off point. The low Fourier coefficients are the linear components of the data. ARIMA, which is reputed to perform very well on linear time series data [4], [14] is used to model the linear component of the data. A high-pass filter is subsequently employed to obtain Fourier coefficients above the cut-off, which represent the high Fourier coefficients. The high Fourier coefficients are nonlinear components of the data. DLSTM, which is very efficient in modeling nonlinear data, is used on the nonlinear component that includes the low Fourier coefficients and the residual error from the ARIMA modeling. The results from both parts are combined to provide the final prediction.

3.1. Converting the electricity consumption data into Fourier coefficients using DFT

DFT is the discrete form of the Fourier transform and converts a signal or discrete sequence from a representation of the time domain to its representation in the Fourier coefficient domain [37]. The equivalent of the continuous Fourier transform is the DFT for signals known at N instants and separated by sample times, T , which is a finite data sequence. The DFT coefficients are obtained as follows:

$$F[n] = \sum_{k=0}^{N-1} f[k]e^{-j\frac{2\pi}{N}nk} \tag{1}$$

N represents the number of observations of the data. $F[n]$ represents the DFT of the sequence $F[k]$. The (1) can be written in matrix form as:

$$\begin{bmatrix} F[0] \\ F[1] \\ F[2] \\ \vdots \\ F[N-1] \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & \dots & 1 \\ 1 & W & W^2 & W^3 & \dots & W^{N-1} \\ 1 & W^2 & W^4 & W^6 & \dots & W^{N-2} \\ 1 & W^3 & W^6 & W^9 & \dots & W^{N-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W^{N-1} & W^{N-2} & W^{N-3} & \dots & W \end{bmatrix} \begin{bmatrix} f[0] \\ f[1] \\ f[2] \\ \vdots \\ f[N-1] \end{bmatrix} \tag{2}$$

where $W = \exp(-\frac{j2\pi}{N})$.

3.2. Application of the low-pass and high-pass filter in data decomposition

The electricity consumption data are converted into Fourier coefficients after the application of the DFT. A low-pass filter is used to filter the linear component of the data, while a high-pass filter is used to filter out the nonlinear component of the data. The low-pass filter selects signals with a Fourier coefficient lower than the selected cut-off, allowing them to pass, and Fourier coefficients higher than the cut-off are blocked. A high-pass filter does the opposite of the low-pass filter, as it filters signals with a Fourier coefficient higher than the cut-off, and Fourier coefficients lower than the cut-off point are then blocked. We contend that Fourier coefficients that are lower than a cut-off point are linear and those higher are nonlinear. The low-pass filter is implemented mathematically as:

$$\hat{F}(\omega) = \begin{cases} 1 & \text{if } |\omega| \leq \omega_c, \\ 0 & \text{if } |\omega| > \omega_c, \end{cases} \tag{3}$$

The specified cut-off Fourier coefficient is denoted as ω_c . The amplitude spectrum of the filter is the unity of $|\omega| < \omega_c$; further, for higher Fourier coefficients, the filter has an amplitude spectrum of zero. The high-pass filter is implemented with the reverse of (3).

3.3. Application of IDFT to convert the data back to the time domain

After the application of DFT to the electricity consumption data, the data will be in Fourier coefficients, and it has to be converted back into its original form using the IDFT. The IDFT of is represented as:

$$F[k] = \frac{1}{N} \sum_{n=0}^{N-1} f[n]e^{+j\frac{2\pi}{N}nk} \tag{4}$$

3.4. Modeling the linear component with ARIMA

ARIMA is a linear modeling technique where the specified time series data are checked first for stationarity. Time series electricity consumption data have trends and seasonality, and, therefore, are not stationary. Where data are not stationary, a differencing operation has to be performed. After each differencing, if the data still exhibit non-stationarity, differencing is again performed until the data are finally stationary. When the differencing is performed d times, the order of integration of the ARIMA technique is d . To predict the time series using ARIMA, the general equation is given as:

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \tag{5}$$

where y_t is the actual value

e_t is the random error at time t

θ_i ($i = 1, 2, \dots, p$) and θ_j ($j = 1, 2, \dots, q$) represent the model's parameters

p and q represent the order of the model and the values are integers

if $q = 0$ then (5) turns into an AR model with the order p

if $p = 0$ in (5) turns into an MA model with the order q . A suitable model order ($p; q$) is determined in building an ARIMA model.

3.5. Deep long short-term memory

The LSTM architecture comprises an input gate, output gate, cell state, and forget gate. The first step is to decide which information the forget gate f_t should throw away from the cell state and is represented by:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

The next stage is to decide on the information to be retained in the cell state. This stage comprises two parts: the input gate layer, i_t , which is a sigmoid layer the values of which are updated, and a new candidate value \tilde{C}_t , which is created by a \tanh layer and is added to the state. This is mathematically represented by:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (7)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (8)$$

The prior cell state C_{t-1} is updated into a new cell state, C_t which is represented as:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (9)$$

The final output is based on a cell state, which is a filtered version. The cell state to be output is decided by a sigmoid layer that is run. DLSTM is similar to other neural networks and requires that data be within the network's activation function. The hyperbolic tangent \tanh is the LSTM default activation function and the values of the output fall between -1 and 1, which is the ideal range in the case of time series data [5]. The cell state is then placed through \tanh and it pushes the values from -1 to 1, which is then multiplied by the output of the sigmoid gate, and the component decided on is output as follows:

$$o_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (10)$$

$$h_t = o_t * \tanh(C_t) \quad (11)$$

From (10) and (11), the input weight, recurrent weight, and bias are presented by the LSTM. To improve general performance, an effective method is to increase the depth of the neural network [38]. Deep recurrent network architectures have impressive learning abilities [39]. In the proposed DLSTM, several LSTM blocks are stacked, as shown in Figure 1(b). They are connected in a deep recurrent network to derive the advantages of a single LSTM layer. The purpose of stacking many LSTMs hierarchically is to build the features at the lower layers for the separation of the factors of variations in the data used for input and subsequently combine them at the higher layers. Such deep architecture will better generalize because the representation is more compact than the shallow representation [40].

From the architecture in Figure 1(b), the input h_t is assigned to LSTM layer 1 at time t with h_{t-1} , which is the previous hidden state of the initial LSTM layer, and transfers to the second layer. The hidden state h_t is used by layer 2 of the LSTM together with the prior hidden state h_{t-1} for the computation h_t of layer 2 of the LSTM and moves afterward to the subsequent stage until the compilation of the last LSTM layer. Part of the task is processed by each layer and is passed to the next layer until the last layer is reached. To model the nonlinear part of the electricity consumption data using DLSTM, the sequence data are initially loaded, which results in a cell array, and each element is a single time step. The data are restructured to a row vector to have a mean of zero and a unit variance for better fitting by standardizing the training data. The test data are also standardized during the forecast using the same parameters as those used for the training data. The predictors and responses are subsequently prepared. To predict a sequence's time steps, the responses are listed as the training sequences, and the values are shifted by a single step. The LSTM network at each step of the sequence learns to predict the subsequent value of the time step. Therefore, the predictor is the training sequence without the last step.

4. EXPERIMENTAL SETUP

The experimental settings that were implemented for both the proposed and reference models are presented in this section. A brief overview of the reference models and the ideal criteria to be relied on for comparing their performance against the proposed model are also presented.

4.1. Hybrid ARIMA–DLSTM prediction model

ARIMA is used to model the reconstructed linear part of the electricity consumption data; the residuals and the nonlinear part are modeled using DLSTM. The generated predictions from ARIMA and DLSTM are aggregated to attain the total forecast. The time series denoted y_t , comprised linear (low) and nonlinear (high) Fourier coefficients, which is:

$$y_t = L_t + H_t \quad (12)$$

H_t represents the nonlinear (high) Fourier coefficient part and L_t represents the linear (low) Fourier coefficient part. L_t is made stationary as a linear function, and, as shown in (13), modeled using ARIMA. H_t is expressed as a nonlinear function, and DLSTM is used for its modeling, as shown in (14).

$$\hat{L}_t = f(L_{t-1}, L_{t-2}, \dots) \quad (13)$$

$$\hat{H}_t = q(H_{t-1}, H_{t-2}, \dots, H_{t-N}) + e_t \quad (14)$$

q signifies the nonlinear function and e_t signifies the model error. The forecast value \hat{Y}_t is shown in (15):

$$\hat{Y}_t = \hat{L}_t + \hat{H}_t \quad (15)$$

In selecting the best fit model in ARIMA, the akaike information criterion (AIC) is used, which suggests that the best model is ARIMA (7, 0, 0). The electricity consumption data are divided into 70% for training and 30% for testing, and because only the next day's forecast is desired, one-step-ahead prediction is employed. The LSTM network architecture is then defined by constructing an LSTM regression network. The network structure is [A, B, C, D]. A represents the input layer, its size is denoted by B, which feeds into an LSTM layer with B neurons, which subsequently feeds into another layer of LSTM with C neurons and then feeds into a normal layer of D neurons that is fully connected to a linear activation that is used in the subsequent step's forecast. To prevent gradient explosion, the threshold of the gradient is set to 1. After the LSTM network is trained with the training options indicated, the time steps for the future are predicted individually, and the state of the network is updated for each forecast. The prior forecast is used as an input function in forecasting future steps.

4.2. Reference models

To achieve a fair evaluation, the proposed hybrid ARIMA–DLSTM model based on DFT decomposition is compared against reference models, which are single statistical or single machine learning models, and hybrid statistical and machine learning models. The reference models are selected based on the impact factor of the journals in which they are published, their citations, and the citations of the authors in similar work undertaken in other papers, the quality of models they were compared with, and their year of publication. The reference models used are as follows:

4.2.1. ARIMA model

The concept of the ARIMA model is explained in section 2. The ARIMA comparison signifies a statistical comparison. The statsmodel library in Python was used to implement the ARIMA program. A grid search that explores diverse combinations of the identified parameters (p, d, q) is used. A new ARIMA model is fit for each parameter combination and the AIC is then used to select the best combination. The AIC measures how well the model fits the data, considering the overall complexity of the model.

4.2.2. Hybrid ETS-ANN model for time series forecasting [14]

This model ensures that purely linear, purely nonlinear, or a combination of linear and nonlinear components are modeled accurately. In their method, the ETS is initially applied to the time series data. This increases the chances of finding linear patterns where they exist. To calculate the residual error, the predictions from the ETS model are deduced from the original series. ANN is used to model the residual error sequence, which is assumed to be nonlinear. The final predictions are obtained through a combination of the predictions from the ETS and ANN.

4.2.3. Modeling and prediction of Turkey's electricity consumption using ANN [10]

An MLP neural network topology and a backpropagation training algorithm are used in this model. Tangent-sigmoid and pure-linear data respectively are selected in the hidden and output layer processing elements.

4.2.4. Electricity consumption prediction based on LSTM with attention mechanism [21]

In this work, an LSTM with an attention mechanism is presented for short-term time-phase electricity consumption modeling. The attention mechanism is initially used to allocate coefficients to the sequence data of the input. The value of the output of each LSTM cell is calculated using the forward propagation method. The backpropagation method is used to compute the error between the actual and predicted values. Each gradient's weight is computed in accordance with the corresponding error term, and, to reduce the error, the model's weight is updated by the gradient descent direction.

4.2.5. A hybrid prediction model for residential electricity consumption using Holt–Winters and extreme learning machine [15]

A hybrid model based on the holt-winters (HW) method and an ELM network are used to predict residential electricity consumption in the short term. An MA filter is used to decompose the original data in components of a stationary linear and fluctuant nonlinear residual. The linear prediction model is generated by the HW method and is used to predict the linear component. The ELM builds a nonlinear model for predicting residential electricity consumption by using linear forecast results, nonlinear residuals, and original data as inputs.

4.2.6. Facebook's Prophet

Prophet is a decomposable time series model that has three model parts: trend, seasonality, and holidays. They are represented as follows:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \quad (16)$$

where $g(t)$ denotes the trend function for modeling non-periodic changes in the value time series, $s(t)$ is the periodic change (i.e., seasonality), $h(t)$ denotes the effect of holidays that occur with irregular schedules, and ϵ_t represents the error term that accounts for any uncommon changes that are not accommodated by the model.

4.3. Measurement of model prediction performance

To measure the performance of the proposed model and compare it against the reference models; RMSE, MAPE, and MAE reputed to be effective in measuring prediction errors [41] are computed. The RMSE measures the error between the predicted and observed values. The smaller the error, the better will be the prediction. RMSE is computed as follows:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (P_i - O_i)^2}{n}} \quad (17)$$

The MAPE usually expressed as a percentage generally states the accuracy of prediction as a ratio and is defined by the formula:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{O_i - P_i}{O_i} \right| * 100 \quad (18)$$

The MAE is simply an average of the absolute percentage errors between the actual and predicted values and is computed as follows:

$$MAE = \frac{\sum_{i=1}^n |P_i - O_i|}{n} \quad (19)$$

where P denotes the value predicted, the observed values at time i are denoted by O , and the number of time elements is denoted by n .

5. RESULTS AND DISCUSSION

Electricity consumption data have trends and seasonality and, therefore, are composed of linear and nonlinear components. DFT converts the data into Fourier coefficients, and, based on a cut-off Fourier coefficient, low-pass and high-pass filters are used to decompose the data into linear and nonlinear components. The Fourier coefficients are converted back into their original forms using inverse discrete Fourier transform (IDFT). The linear component is modeled by ARIMA, and the nonlinear component is modeled by DLSTM, which has the capability of modeling nonlinear data very accurately. Household electricity consumption data gathered in France every minute for 47 months from December 2006 to

November 2010, representing 2,075,259 power measurements [42], are used in the evaluation of the model. To ensure accurate predictions, the data were aggregated daily and used for the forecast. The quantitative and visual results of the proposed hybrid ARIMA-DLSTM model along with the reference models are presented in this section. The individual household electricity consumption dataset is used for the proposed and reference models. This dataset represents 2,075,259 power measurements collected in a house situated in Sceaux near Paris, France, for 47 months spanning December 2006 to November 2010 with data taken every minute. The attributes used are as follows:

- date: Date in format dd/mm/yyyy
- time: time in format hh:mm:ss
- global_active_power: represents the electricity consumption in minutes

In analyzing the raw data, some missing values were detected thus the data had to be pre-processed. The forecasting model’s predictive efficiency may be affected if part of the data is missing and the mean, which is one of the methods for filling missing data in Python, is used to fill the missing values. The data are gathered at intervals of one minute. Given the trends and seasonality inherent in electricity consumption data, the ideal forecast should be performed within 24 hours (1 day). The consumption data, which is taken every minute, is converted to hourly data by dividing each value in global-active-power by 60. To affirm the precision of prediction attained by the proposed model against the six single and hybrid models discussed in section 2, a simulation experiment is performed; these models are applied to the individual household electricity consumption data. Figure 2 shows a plot of the daily household electricity consumption time series data.

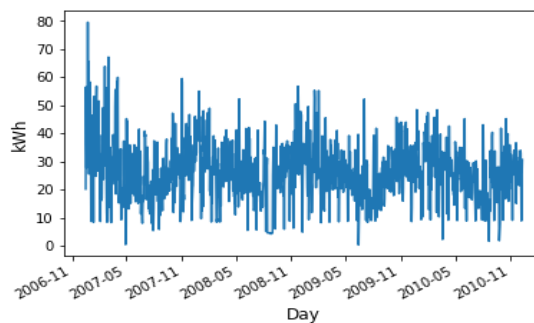


Figure 2 Household electricity time series data

In comparing the accuracy of predictions of the proposed model and the referenced models, three measures are used. They are RMSE, MAPE, and MAE which are depicted in (17), (18), and (19). A good prediction model should have lower error values. The electricity consumption data which is made up of 2,075,259 hourly data points is processed to obtain 1433 points of daily consumption in kWh as depicted in Figure 2. In the experiment, one-day-ahead forecasting with 430 points as forecast horizon, one-week-ahead prediction with 62 points as forecast horizon and 10-day ahead prediction representing 44 points is used. The error performance results are shown in Table 1 and confirms that the prediction accuracy of the proposed model is better than the other models.

Table 1. Performance comparison of models

Model	One-day-ahead			One-week-ahead			Ten-days-ahead		
	RMSE	MAPE %	MAE	RMSE	MAP %	MAE	RMSE	MAPE %	MAE
ARIMA	7.540	35.616	5.719	31.301	15.625	22.424	57.034	26.997	39.077
Prophet (Facebook)	7.232	33.768	5.504	31.089	15.757	22.982	56.792	26.639	37.919
ANN (Kovaklioglu, 2011)	7.054	30.610	5.390	30.738	15.615	22.422	56.804	26.696	38.058
ETS-ANN (Panigrahi and Behera [14])	6.998	32.223	5.331	30.742	15.637	22.554	56.98	26.671	37.698
Holt-Winter and Extreme Learning (Liu et al. [15])	6.933	31.847	5.309	30.571	15.564	22.444	57.135	26.661	37.714
LSTM with Attention Mechanism (Lin et al. [21])	6.918	30.514	5.305	31.377	15.947	23.385	57.426	26.476	36.92
Proposed ARIMA-DLSTM	5.905	28.786	5.292	27.587	12.528	18.382	43.246	24.246	30.613

Prophet was presented by facebook to address the challenges associated with ARIMA's prediction accuracy. From Table 1, Prophet performed better than ARIMA in 1-day ahead, one-week-ahead, and 10-day ahead predictions. ANN outperformed prophet (facebook) and ARIMA in 1-day-ahead, one-week-ahead and 10-days-ahead prediction accuracy. Compared with the other benchmark models, ANN is the best performing single model for one-day-ahead and one-week-ahead predictions, however, for 10-day ahead prediction Prophet outperformed ANN. LSTM with attention seeking proposed in [21], performed better than the other single and hybrid referenced models in one-day-ahead prediction, however, its performance degraded in one-week-ahead and ten-day ahead prediction as the other models outperformed it. The one-day-ahead, one-week-ahead and 10 days-ahead predictions achieved by the benchmark models and the proposed ARIMA-DLSTM model are depicted in Figure 3, Figure 4, and Figure 5 respectively.

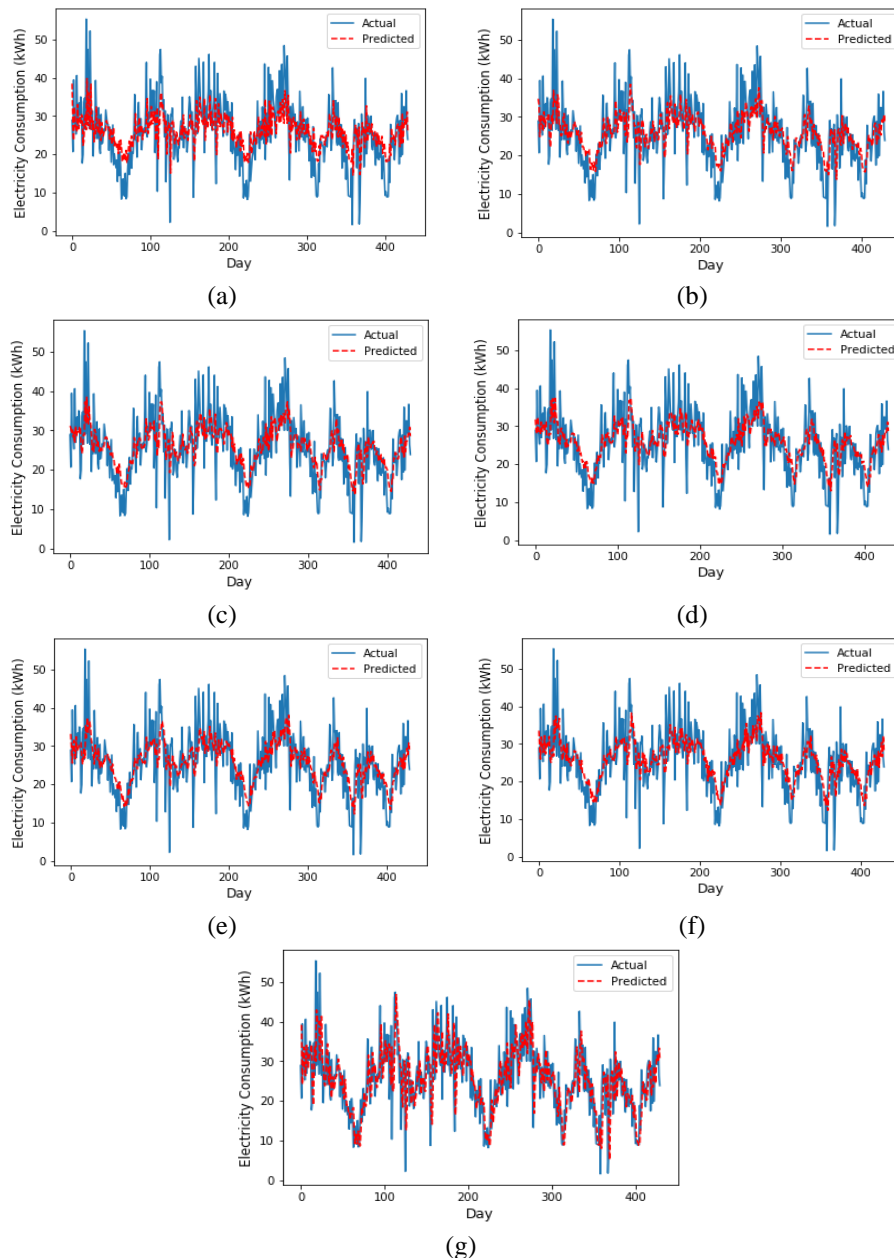


Figure 3. These figures are; (a) ARIMA, (b) Prophet (Facebook), (c) ANN (Kovaklioglu, 2011), (d) ETS-ANN (Panigrahi and Behera [14]), (e) Holt-Winter and Extreme Learning (Liu *et al.* [15]), (f) LSTM with Attention Mechanism (Lin *et al.* [21]). One-day-ahead prediction of electricity consumption in kWh attained by the proposed ARIMA-DLSTM model and the benchmark models

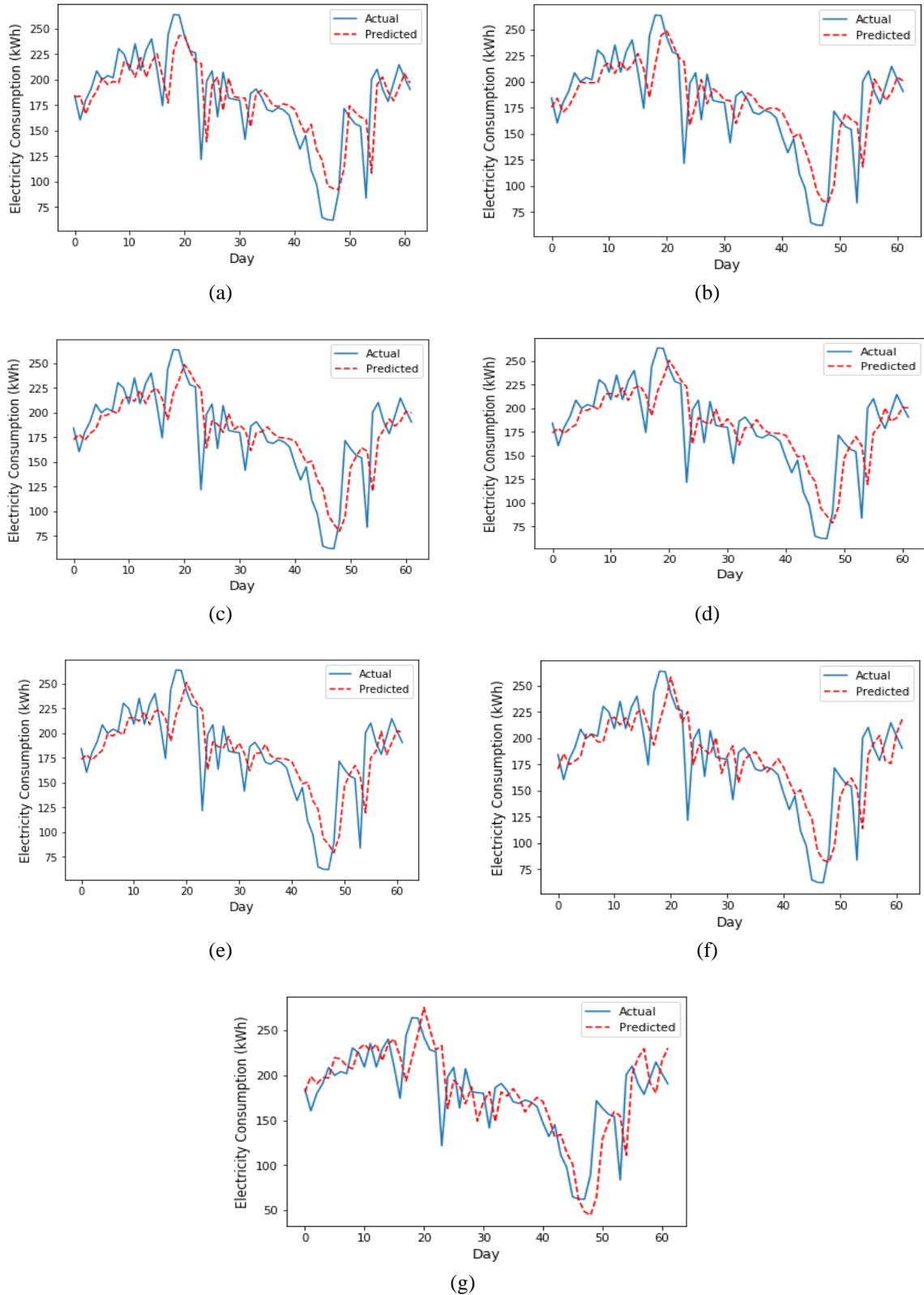


Figure 4. These figures are; (a) ARIMA, (b) Prophet (Facebook), (c) Prophet (Facebook), (d) ETS-ANN, (Panigrahi and Behera [14]), (e) Holt-Winter and Extreme Learning, (Liu *et al.* [15]), (f) LSTM with Attention Mechanism, (Lin *et al.* [21]), (g) Proposed ARIMA-DLSTM. One-week-ahead prediction of electricity consumption in kWh attained by the proposed ARIMA-DLSTM model and the benchmark models

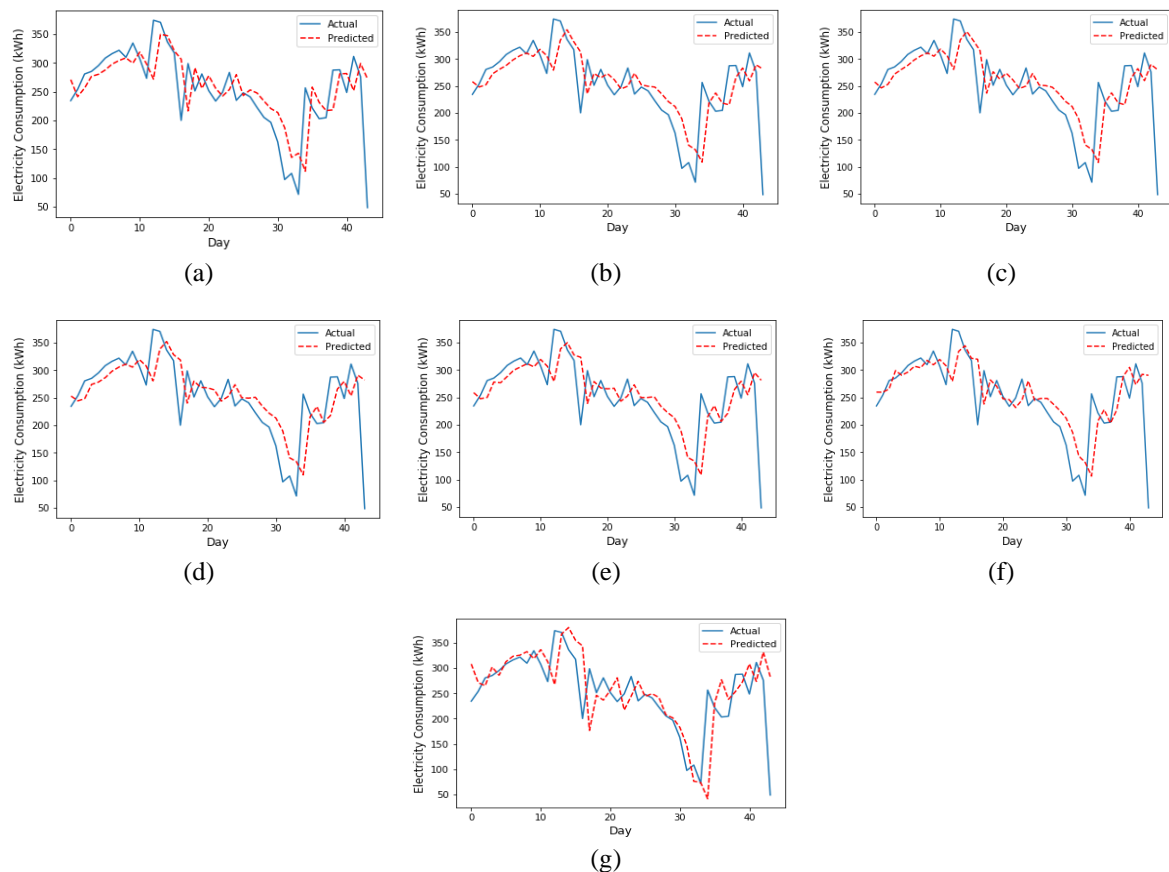


Figure 5. These figures are; (a) ARIMA, (b) Prophet (Facebook), (c) ANN (Kovaklioglu, 2011), (d) ETS-ANN (Panigrahi and Behera [14]), (e) Holt-Winter and Extreme Learning (Liu *et al.* [15]), (f) LSTM with Attention Mechanism (Lin *et al.* [21]), (g) Proposed ARIMA-DLSTM.10-days-ahead prediction of electricity consumption in kWh attained by the proposed ARIMA-DLSTM model and the benchmark models

The proposed model performed better than all the referenced models in one-day-ahead, one-week-ahead and ten-days-ahead prediction. It achieved a lower RMSE, MAPE, and MAE in all the predictions. The one-day-ahead, one-week-ahead, and 10-day ahead predictions are shown in Figure 3, Figure 4, and Figure 5 respectively. It can be observed from the figures that the predicted values closely follow the actual values better than the referenced models.

6. CONCLUSION

A hybrid ARIMA-DLSTM model based on DFT decomposition, which is promising for forecasting the majority of time series problems, was developed. It is widely contended that time series data consist of both linear and nonlinear components. Individual linear models such as ARIMA and nonlinear models such as DLSTM cannot be used to model such data precisely. Therefore, the time series data should be decomposed into linear and nonlinear components with efficient techniques that do not result in performance degradation. In the proposed system, DFT was used to convert the data into Fourier coefficients based on a cut-off; low-pass and high-pass filtering techniques in DFT were used to segregate the low Fourier coefficients (linear component) and high Fourier coefficients (nonlinear component) to be modeled by ARIMA and DLSTM, respectively. Owing to the efficient decomposition technique used, the performance was not degraded. The stacking of more LSTM layers resulted in a more precise prediction of the nonlinear data components. To validate the accuracy of the proposed model, the individual household electricity consumption data in France recorded by the UCI machine learning repository was used on both the proposed and reference models for one-day-ahead, one-week-ahead, and 10-days-ahead prediction. The error measures shows that the proposed model performed remarkably better than the six referenced models. For one-day-ahead prediction and 10-days-ahead prediction, the best MAPE recorded by existing models is 30.514% and

26.476% respectively. The proposed model achieved a better MAPE of 28.786% and 24.246% for one-day-ahead and 10-days-ahead respectively, signifying a remarkable performance. Thus, the proposed model can be utilized to predict other time series tasks. In future work, we expect to extend the prediction of electricity consumption to detect electricity theft based on the predicted consumption.

REFERENCES

- [1] K. Kavaklioglu, "Modeling and prediction of Turkey's electricity consumption using Support Vector Regression," *Applied Energy*, vol. 88, no. 1, pp. 368-375, 2011, doi: 10.1016/j.apenergy.2010.07.021.
- [2] Z. Hajirahimi and M. Khashei, "Hybrid structures in time series modeling and forecasting: A review," *Engineering Applications of Artificial Intelligence*, vol. 86, pp. 83-106, 2019, doi: 10.1016/j.engappai.2019.08.018.
- [3] J. De Gooijer and R. Hyndman, "25 years of time series forecasting," *International Journal of Forecasting*, vol. 22, no. 3, pp. 443-473, 2006, doi: 10.1016/j.ijforecast.2006.01.001.
- [4] G. Zhang, "Time series forecasting using a hybrid ARIMA and neural network model," *Neurocomputing*, vol. 50, pp. 159-175, 2003, doi: 10.1016/S0925-2312(01)00702-0.
- [5] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203-213, 2019, doi: 10.1016/j.neucom.2018.09.082.
- [6] M. Bose and K. Mali, "Designing fuzzy time series forecasting models: A survey," *International Journal of Approximate Reasoning*, vol. 111, pp. 78-99, 2019, doi: 10.1016/j.ijar.2019.05.002.
- [7] Q. Yang and X. Wu, "10 challenging problems in data mining research," *International Journal of Information Technology & Decision Making*, vol. 05, no. 04, pp. 597-604, 2006, doi: 10.1142/S0219622006002258.
- [8] K. Zhang, R. Gençay, and M. Ege Yazgan, "Application of wavelet decomposition in time-series forecasting," *Economics Letters*, vol. 158, pp. 41-46, 2017, doi: 10.1016/j.econlet.2017.06.010.
- [9] C. Babu and B. Reddy, "A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data," *Applied Soft Computing*, vol. 23, pp. 27-38, 2014, doi: 10.1016/j.asoc.2014.05.028.
- [10] K. Kavaklioglu, H. Ceylan, H. Ozturk, and O. Canyurt, "Modeling and prediction of Turkey's electricity consumption using Artificial Neural Networks," *Energy Conversion and Management*, vol. 50, no. 11, pp. 2719-2727, 2009, doi: 10.1016/j.enconman.2009.06.016.
- [11] S. Ho and M. Xie, "The use of ARIMA models for reliability forecasting and analysis," *Computers & Industrial Engineering*, vol. 35, no. 1-2, pp. 213-216, 1998, doi: 10.1016/S0360-8352(98)00066-7.
- [12] I. Livieris, E. Pintelas, and P. Pintelas, "A CNN-LSTM model for gold price time-series forecasting," *Neural Computing and Applications*, vol. 32, no. 23, pp. 17351-17360, 2020, doi: 10.1007/s00521-020-04867-x.
- [13] L. Wei, "A hybrid ANFIS model based on empirical mode decomposition for stock time series forecasting," *Applied Soft Computing*, vol. 42, pp. 368-376, 2016, doi: 10.1016/j.asoc.2016.01.027.
- [14] S. Panigrahi and H. Behera, "A hybrid ETS-ANN model for time series forecasting," *Engineering Applications of Artificial Intelligence*, vol. 66, pp. 49-59, 2017, doi: 10.1016/j.engappai.2017.07.007.
- [15] C. Liu, B. Sun, C. Zhang, and F. Li, "A hybrid prediction model for residential electricity consumption using holt-winters and extreme learning machine," *Applied Energy*, vol. 275, 2020, doi: 10.1016/j.apenergy.2020.115383.
- [16] C. Babu and P. Sure, "Partitioning and interpolation based hybrid ARIMA-ANN model for time series forecasting," *Sādhanā*, vol. 41, no. 7, pp. 695-706, 2016, doi: 10.1007/s12046-016-0508-5.
- [17] M. Shen, Y. Lu, K. Wei, and Q. Cui, "Prediction of household electricity consumption and effectiveness of concerted intervention strategies based on occupant behaviour and personality traits," *Renewable and Sustainable Energy Reviews*, vol. 127, 2020, doi: 10.1016/j.rser.2020.109839.
- [18] J. Torres, A. Galicia, A. Troncoso, and F. Martínez-Álvarez, "A scalable approach based on deep learning for big data time series forecasting," *Integrated Computer-Aided Engineering*, vol. 25, no. 4, pp. 335-348, 2018, doi: 10.3233/ICA-180580.
- [19] D. de O. Santos Júnior, J. de Oliveira, and P. de Mattos Neto, "An intelligent hybridization of ARIMA with machine learning models for time series forecasting," *Knowledge-Based Systems*, vol. 175, pp. 72-86, 2019, doi: 10.1016/j.knosys.2019.03.011.
- [20] J. Bedi and D. Toshniwal, "Deep learning framework to forecast electricity demand," *Applied Energy*, vol. 238, pp. 1312-1326, 2019, doi: 10.1016/j.apenergy.2019.01.113.
- [21] Z. Lin, L. Cheng, and G. Huang, "Electricity consumption prediction based on LSTM with attention mechanism," *IEEE Transactions on Electrical and Electronic Engineering*, vol. 15, no. 4, 2020, doi: 10.1002/tee.23088.
- [22] Y. He, Y. Qin, S. Wang, X. Wang, and C. Wang, "Electricity consumption probability density forecasting method based on LASSO-Quantile Regression Neural Network," *Applied Energy*, vol. 233-234, pp. 565-575, 2019, doi: 10.1016/j.apenergy.2018.10.061.
- [23] S. Ding, K. Hipel, and Y. Dang, "Forecasting China's electricity consumption using a new grey prediction model," *Energy*, vol. 149, pp. 314-328, 2018, doi: 10.1016/j.energy.2018.01.169.
- [24] A. Eseye and M. Lehtonen, "Short-Term Forecasting of Heat Demand of Buildings for Efficient and Optimal Energy Management Based on Integrated Machine Learning Models," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 12, pp. 7743-7755, 2020, doi: 10.1109/tii.2020.2970165.
- [25] X. Zhu, Y. Dang, and S. Ding, "Using a self-adaptive grey fractional weighted model to forecast Jiangsu's electricity consumption in China," *Energy*, vol. 190, 2020, doi: 10.1016/j.energy.2019.116417.
- [26] S. Taylor and B. Letham, "Forecasting at scale," *Facebook Research*, 2021. [Online]. Available: <https://research.fb.com/blog/2017/02/prophet-forecasting-at-scale/>. Accessed: 21 March 2021.

- [27] L. Wen and X. Yuan, "Forecasting the annual household electricity consumption of Chinese residents using the DPSO-BP prediction model," *Environmental Science and Pollution Research*, vol. 27, no. 17, pp. 22014-22032, 2020, doi: 10.1007/s11356-020-08418-8.
- [28] Z. Shen, Y. Zhang, J. Lu, J. Xu, and G. Xiao, "A novel time series forecasting model with deep learning," *Neurocomputing*, vol. 396, pp. 302-313, 2020, doi: 10.1016/j.neucom.2018.12.084.
- [29] M. Shao, X. Wang, Z. Bu, X. Chen, and Y. Wang, "Prediction of energy consumption in hotel buildings via support vector machines," *Sustainable Cities and Society*, vol. 57, 2020, doi: 10.1016/j.scs.2020.102128.
- [30] A. Zeng, H. Ho, and Y. Yu, "Prediction of building electricity usage using Gaussian Process Regression," *Journal of Building Engineering*, vol. 28, 2020, doi: 10.1016/j.jobe.2019.101054.
- [31] Y. Tan, X. Guo, and S. Poh, "Time series activity classification using gated recurrent units," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3551-3558, 2021, doi: 10.11591/ijece.v11i4.pp3551-3558.
- [32] M. Reza Pahlawan, E. Riksakomara, R. Tyasnurita, A. Muklason, F. Mahananto, and R. A. Vinarti, "Stock price forecast of macro-economic factor using recurrent neural network," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 10, no. 1, pp. 74-83, 2021, doi: 10.11591/ijai.v10.i1.pp74-83.
- [33] S. Verma, G. Thampi, and M. Rao, "ANN based method for improving gold price forecasting accuracy through modified gradient descent methods," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 1, pp. 46-57, 2020, doi: 10.11591/ijai.v9.i1.pp46-57.
- [34] H. Bousqaoui, I. Slimani, and S. Achchab, "Comparative analysis of short-term demand predicting models using ARIMA and deep learning," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3319-3328, 2021, doi: 10.11591/ijece.v11i4.pp3319-3328.
- [35] M. Rhanoui, S. Yousfi, M. Mikram, and H. Merizak, "Forecasting Financial Budget Time Series: ARIMA Random Walk vs LSTM Neural Network," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 8, no. 4, pp. 317-327, 2019, doi: 10.11591/ijai.v8.i4.pp317-327.
- [36] 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, doi: 10.1016/j.asoc.2010.10.015.
- [37] S. Winograd, "On computing the discrete Fourier transform," *Mathematics of Computation*, vol. 32, no. 141, pp. 175-175, 1978, doi: 10.1090/s0025-5718-1978-0468306-4.
- [38] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, 2015, doi: 10.1038/nature14539.
- [39] K. Greff, R. Srivastava, J. Koutnik, B. Steunebrink, and J. Schmidhuber, "LSTM: A Search Space Odyssey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2222-2232, 2017, doi: 10.1109/TNNLS.2016.2582924.
- [40] M. Hermans and B. Schrauwen, "Training and analyzing deep recurrent neural networks," *Adv. Neural Inf. Process. Syst.*, pp. 190-198, 2013.
- [41] C. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, 2005, doi: 10.3354/cr030079.
- [42] G. Hebrail and A. Berard, "UCI Machine Learning Repository: Individual household electric power consumption Data Set," [Archive.ics.uci.edu](http://archive.ics.uci.edu), 2021. [Online]. Available: <http://archive.ics.uci.edu/ml/datasets/Individual+household+electric+power+consumption> [Accessed: 20 March 2021].

BIOGRAPHIES OF AUTHORS



Osman Yakubu obtained a B. Sc. computer science degree from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana in 2005, M. Sc. Informatics from the University of Ulster, Belfast, UK in 2007, and a Ph. D in computer science from M. S. Ramaiah University of Applied Sciences, Bangalore, India in 2020. Research interests are internet of things, data science, and biometric systems. Currently lecturing and offering IT consultancy services.



Narendra Babu C. obtained Ph. D (computer science and engineering) from JNT University Anantapur, India. Currently an associate professor in computer science and engineering at M.S. Ramaiah University of Applied Sciences, Bangalore. Research interests include data science, machine learning, and artificial intelligence.