

A Novel Energy Consumption Prediction Model Combining Long Short-Term Memory (LSTM) and Fractional Differential Equations (FDLE)



Ahlam Alghanmi^{ID}

Department of Basic Science, Applied College, Taibah University, Medina 41477, Saudi Arabia

Corresponding Author Email: aghanmi@taibahu.edu.sa

Copyright: ©2025 The author. This article is published by IIETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijcmem.130110>

Received: 20 July 2024

Revised: 9 September 2024

Accepted: 19 September 2024

Available online: 31 March 2025

Keywords:

fractional differential equations, mathematical models, LSTM, LSTM-FDLE model, smart distribution

ABSTRACT

Optimized energy generation and smart distribution in a sustainable manner requires accurate prediction of its consumption. However, the prediction of energy demands of households remains a tedious task due to variations in patterns of energy usage. Mathematical models and artificial intelligence (AI), such as smart energy-efficient designs, strategic planning for smart grids, and Internet of Things (IoT)-enabled smart homes, have recently been considered as solutions to these issues. A major issue encountered in energy consumption prediction systems is their restricted prediction horizons, as well as their dependence on one-step predictions. This study, therefore, suggests an innovative model for the prediction of energy demand that uses a long short-term memory (LSTM) and fractional differential equations (FDLE)-based model. The proposed LSTM-FDLE model was trained to predict the collective active power generated by household devices. LSTM's memory and sequential learning capabilities were also explored in the proposed model for comprehending the complex temporal dependencies and trends in energy consumption data. The performance of the proposed model was evaluated on real-world household energy usage data and found to achieve good prediction accuracy; the performance of the model was also better than that of some conventional one-step prediction models. Therefore, better energy generation planning, and optimal distribution systems can be achieved by the longer forecasting period provided by the proposed "LSTM" model.

1. INTRODUCTION

Rapid urbanization, surging human population, and societal demand in the building sector are some of the factors driving the significant increase in household energy demand recently [1]. The global energy usage and the associated greenhouse gas emissions has been influenced by the increasing demand buildings [2]. It is important that buildings should be sustainable and energy-efficient and there is a need that utility companies, customers, and facility managers to understand the trends in building energy usage to improve energy efficiency. Scholars have developed interest in learning more about energy-saving and building energy efficiency [3, 4]. The study and examination of data through the use of AI-based techniques like machine learning (ML) have been eased with the advancement of Industry 4.0. Therefore, the prediction of the energy usage pattern in buildings has become necessary for better energy performance at the operation and maintenance phases of buildings.

The analysis of building energy patterns can be achieved using either the data-driven approach or the physics-based simulation approach [1]. The physics-based method requires the use of the whole building energy simulation tools like DOE-2 and Energy Plus. The users of these tools must provide the heat properties of the building components, such as walls,

roofs, and windows, HVAC (heating, ventilation, and air conditioning) systems, thermal settings, internal loads of occupancy, building geometry, etc. but due to the lack of necessary data and time commitment, these techniques are not mostly useful for energy analysis throughout the operation and maintenance phases [5].

During energy analyses for various buildings, there are numerous problems associated with the physics-based approaches. Hence, data driven approaches were developed to mostly use past data to access building energy performance. The development of AI and IoT enables the use of this method to understand building energy usage patterns. The use of this data-driven method requires access to open data. The use of ML for the estimation of the future pattern of energy usage of buildings has been proposed by many researchers. For instance, the use of ANN [6], decision trees (DT) [7], ANFIS [8], and SVM [9] for the prediction of building energy usage has been proposed. Artificial Neural Networks (ANN) are computational models inspired by the human brain, designed to recognize patterns and make predictions by learning from data through layers of interconnected nodes (neurons) [6]. Decision Trees (DT) are a type of supervised learning algorithm used for classification and regression tasks; they work by splitting the data into subsets based on feature values, forming a tree-like structure where each branch represents a

decision rule [7]. Adaptive Neuro-Fuzzy Inference System (ANFIS) combines neural networks with fuzzy logic to model complex systems by learning from data and handling uncertainty through fuzzy rules, making it effective for control and prediction tasks [8]. Support Vector Machines (SVM) are powerful supervised learning algorithms used primarily for classification and regression; they work by finding the optimal hyperplane that separates data points of different classes in a high-dimensional space, often providing robust results even in cases where the data is not linearly separable [9].

The ever-increasing energy demand can be addressed by the use of energy efficient solutions, which can lower greenhouse gas emissions and improve energy security [2]. Battile et al. (2020) considered the opportunities for energy reduction in buildings by creating energy baselines and energy performance indicators [10]. They observed that the educational facilities might save up to 9.6% of their annual energy costs by considering the developed energy performance indicators. The study also came up with a method for aiding decision-makers in putting urban greening policies into practice by statistically predicting the impact of green initiatives on building energy [11].

Based on the brief survey, The existing prediction models are either based on one or a small number of datasets for the training and testing processes which may impact the generalization of the model results. this is the first research gap of the current study. Besides, until now, only a few works have assessed the performance of models that uses different datasets to predict hourly building energy use which is can consider the second research gap. Therefore, this work intends to evaluate the applicability and efficacy of the DL model and a modified mathematical equation called fractional calculus in the prediction of future building energy usage using numerous experimental datasets. A new model called (LSTM-FDLE) was designed in this work for the forecasting of energy demand through the utilization of two emerging techniques - long short-term memory (LSTM) and fractional differential equations. The proposed LSTM-FDLE model was trained for accurate prediction of the collective active power generated by household appliances. It leverages the memory and sequential learning capabilities of LSTM to capture complex temporal intricacies and trends in energy usage data.

2. THE MODIFIED FRACTIONAL DIFFERENTIAL EQUATIONS -AN OVERVIEW

In the past few years, fractional differential equations (FDE) have grown in popularity as a powerful and well-organized way to study a wide range of scientific and engineering events. Different fields of study conduct research on FDE as it is usable in many areas, such as heat transfer, circuit systems, elasticity, control systems, continuum mechanics, signal analysis, quantum mechanics, biomathematics, social systems, bioengineering, biomedicine, and many more [12, 13].

As of now, there is no method that everyone agrees on that defines fractional calculus. Different meanings of fractional calculus have been made possible by mathematicians' in-depth studies of the issue from various perspectives. The Riemann–Liouville (R–L) definition, the Grünwald–Letnikov (G–L) definition, and the Capotu definition are the three most common ways to explain fractional calculus [14]. The G–L definition is appropriate for use in medical image processing owing to its lesser complexity compared to other definitions

and only requires one coefficient. The L'Hospital's rule has been used as the basis for deriving the 1st, 2nd, and 3rd-order derivatives of the function $f(t)$ as follows:

$$f'(r) = \lim_{g \rightarrow 0} \frac{f(r+h)-f(r)}{g} \quad (1)$$

$$f''(r) = [f'(t)]' \lim_{g \rightarrow 0} \frac{f(r+2h)-2f(r+h)+f(r)}{g^2} \quad (2)$$

$$f'''(r) = [f''(r)]' \lim_{g \rightarrow 0} \frac{f(r+3h)-3f(r+2h)+3f(r+h)-f(r)}{g^3} \quad (3)$$

$$f^{(n)}(r) = \lim_{g \rightarrow 0} g^{-n} \sum_{j=0}^n (-1)^j \binom{n}{j} f(r-jg) \quad (4)$$

The gamma function generates the fractional order, ranging from integer to fraction. The v-order fractional differential of function $f(t)$ is defined as the derivative of order $(n+1)$ on the interval $[a, b]$, where function $f(r)$ has $(n+1)$ -order derivatives.

$$aD_b^\nu f(r) = \lim_{g \rightarrow 0} r^{-\nu} \sum_{j=0}^{\lfloor (b-a)/r \rfloor} (-1)^j \binom{\nu}{j} f(r-hg) \quad (5)$$

where, the integer part of $\frac{b-a}{r}$ is $\left[\frac{b-a}{r} \right]$ and $\binom{\nu}{j} = \frac{\nu^j}{j! (v-j)!}$ is binomial coefficient.

3. PROPOSED METHODOLOGY

The research framework developed for this study is presented in this section; The method begins with a literature review and concludes with an assessment of the research's results. The literature review was aimed at looking for methods already in use for the prediction of household electric power usage. In the suggested methodology, there are 3 stages, and the output of each stage is used as the input for the next. Phase 1 is about the dataset reading while phase two is about pre-processing phase for preparing to train the LSTM- FDLE model. Finally, phase three is about model training and evaluation of the research achievements. Figure 1 depicts in detail the general methodology of the proposed LSTM- FDLE model.

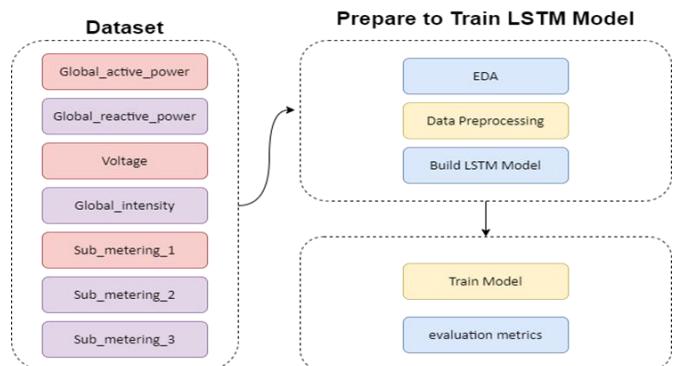


Figure 1. Structural design of the suggested framework

3.1 The utilized dataset

The data utilized in this study called (Household Electric Power Consumption dataset) comprises 2,075,259 data points collected from a residence situated in Sceaux, France, spanning a period of 47 months, commencing in December

2006 and concluding in November 2010 [15]. These measurements were recorded at one-minute intervals. It's worth noting that this dataset might include some gaps, accounting for approximately 1.25% of the total entries. While the dataset contains all calendar timestamps, there are certain timestamps that do not have corresponding measurement values. These missing data points are identified by the lack of a value between two semicolons attribute separators.

The dataset will be used to develop a predictive model for the prediction of the global active power and household energy usage based on the explained parameters.

The EDA of the dataset is provided here [16, 17]; this includes 4 figures that further explained the dataset. There are numerous subplots in Figure 2, each depicting the mean resampled data for a specific variable on a monthly basis. Time (in months) is represented in the horizontal axis, while the

average value of the variable is in the vertical axis. This figure improves understanding of the monthly trends in the data.

Like in Figure 2, the image in Figure 3 consists of multiple subplots but the data representation is on a daily basis. The mean of the resampled data for a specific variable is represented by each subplot; the duration is represented in the horizontal axis (in days) while the average value of the variable is in the vertical axis. This figure provides better understanding of the daily data variations and patterns.

Figure 4 is a representation of the data on an hourly basis where each subplot is a representation of the meaning of the resampled data for a given variable. The time in hours is plotted in the horizontal axis while the average value of the variable is in the vertical axis. This figure provides better understanding of the hourly changes and patterns in the data.

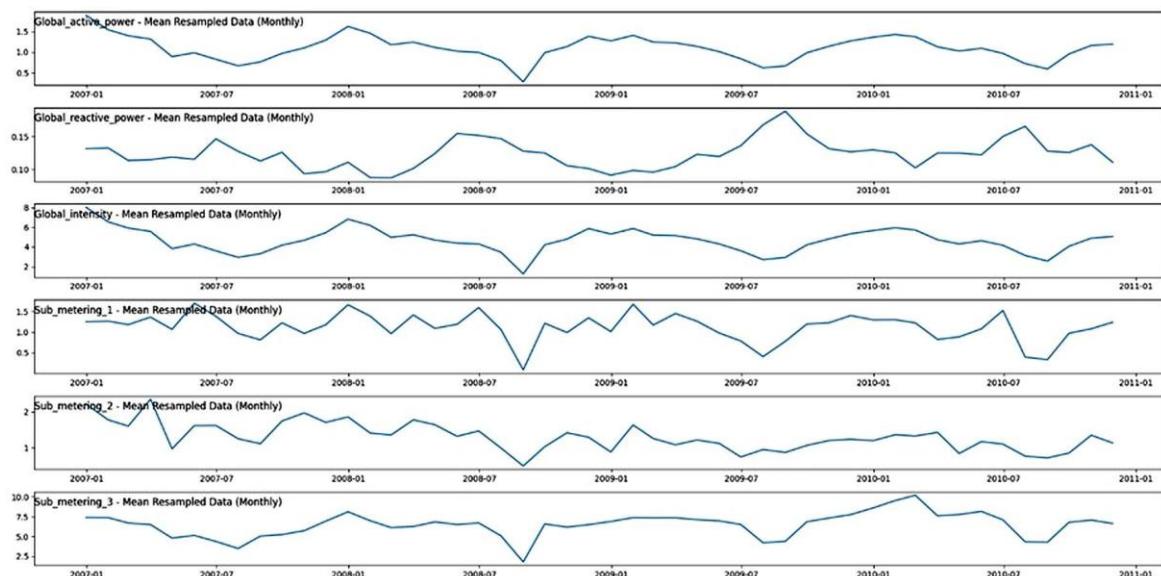


Figure 2. The subplots for the monthly data trends and patterns

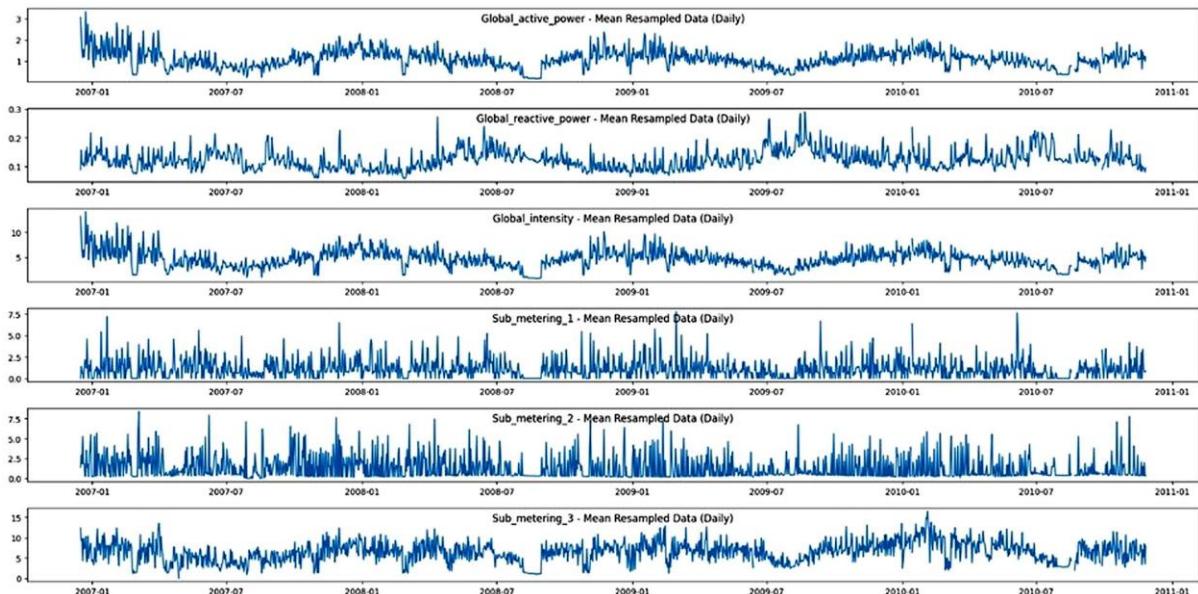


Figure 3. Multiple subplots for the daily data variations and patterns

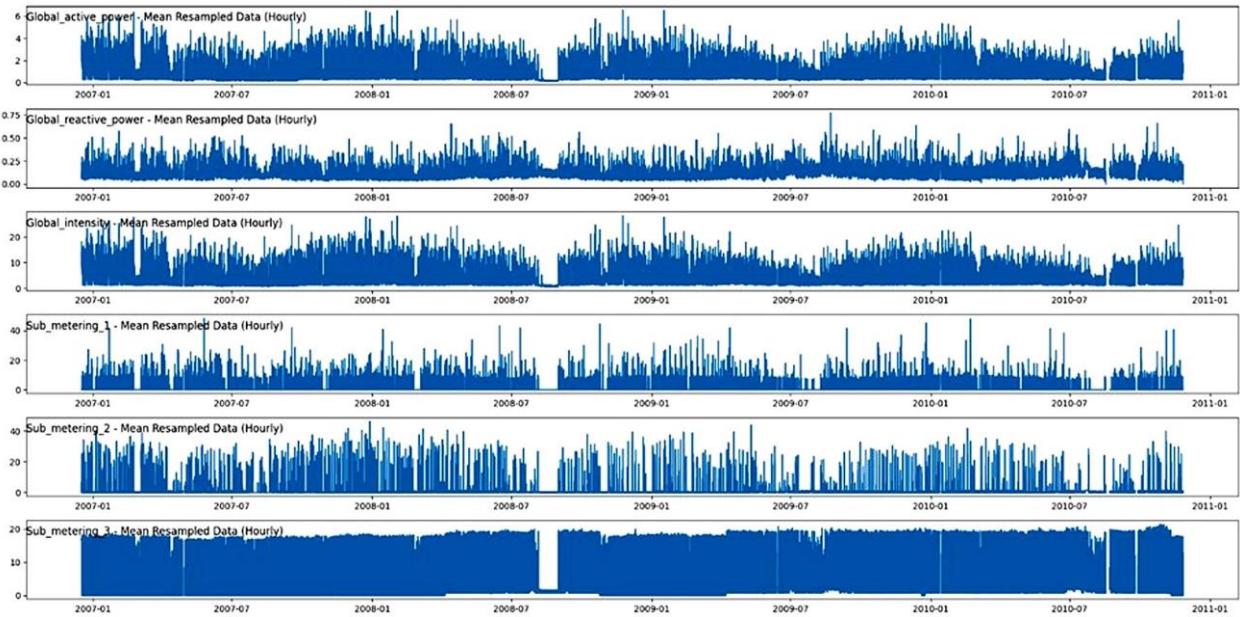


Figure 4. Multiple subplots for the hourly changes and trends in the data

A correlation matrix heatmap is presented using a color map in Figure 5; it depicts the relationship between different parameters in the data. The strength and direction of the correlation is indicated by the values in the heatmap, where a positive value = positive correlation, negative value = negative correlation. Darker colors show strong correlations. This figure provides better view of the relationships between different data variables, as well as a knowledge of the trends, patterns, and correlations in the data.

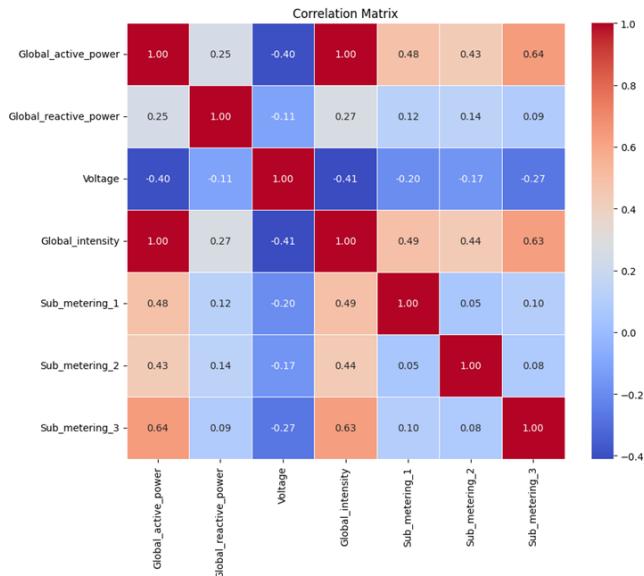


Figure 5. Correlations and dependencies between different data variables

3.2 Data pre-processing

The significance and novelty of using fractional differential equations (FDEs) in this context lie in their ability to model systems with memory and hereditary properties more accurately than classical differential equations. This approach allows for a more nuanced description of dynamic processes, particularly in cases where standard integer-order models fall

short in capturing the complex behaviors of the system. Unlike traditional methods, FDEs provide greater flexibility in adjusting the model to account for non-local interactions and anomalous diffusion, which are common in many real-world applications, such as image processing or signal analysis. There are three important steps in the data preprocessing stage before the training phase of the developed LSTM- FDLE model; the first step is to removal of missing data from the dataset to ensure the completeness of the data for the training process and to avoid potential biases or inaccuracies caused by missing data. The next step is the data standardization step using the MinMaxScaler [18]. This process is mostly used to transform data into a common scale to improve effectiveness when using ML models [19]. The MinMaxScaler was used to scale the data to a specific range, mostly [0, 1] to ensure that the magnitude of all the features is similar. The final step is the preparation of the data as a time series by transforming it into a suitable format for the training process. His process requires the creation of lagged input and output variables that allows the network to learn patterns and dependencies over time. The completion of these preprocessing steps prepares the data for the training of the model on the prepared time series data.

Among the ML algorithms, RNNs have been the most used for sequential data modeling; they are used in many areas like speech recognition, time series data analysis, and natural language processing. The emphasis in this article is on the RNNs, especially LSTM networks and their usage for energy consumption prediction. Efficient energy management, demand projection, and resource allocation requires accurate energy consumption prediction. LSTM leverages the temporal connections and trends in energy usage data to predict the future energy usage pattern. A summary of RNNs and “LSTM” networks is presented here, focusing more on their foundations and their relevance in energy consumption prediction. The RNN is derived from differential equations and introduced from the general form of the state signal evolution equation. The special case of the state signal equation was also presented; the components involved in energy consumption prediction were also detailed.

An RNN was developed in this work using differential equations; the starting point is the definition of $\tilde{s}(r)$ as a d-

dimensional state signal vector, as well as the analysis of the general nonlinear first-order non-homogeneous ODE; this will reveal the evolution of the state signal over time (r).

$$\frac{c\vec{s}(r)}{cr} = \vec{f}(r) + \vec{\phi} \quad (6)$$

where, $\vec{f}(r)$ is a vector-valued function of time, where t is a positive real number. The vector ϕ is a constant with d dimensions. A common representation for $\vec{f}(t)$ is as follows:

$$\vec{f}(r) = \vec{h}(\vec{s}(r), \vec{x}(r)) \quad (7)$$

$\vec{x}(r)$ signifies the d -dimensional input signal vector, and the function $\vec{h}(\vec{s}(r), \vec{x}(r))$ takes vectors as arguments and yields a vector-valued result. The system that emerges from this formulation is encountered in a variety of scenarios within the fields of physics, chemistry, biology, and engineering.

$$\frac{c\vec{s}(r)}{ct} = \vec{h}(\vec{s}(r), \vec{x}(r)) + \vec{\phi} \quad (8)$$

Finally, $\vec{x}(r)$ is derived by gathering all the values of $x(i, r)$ for all possible index components permutations into a column vector as seen in Eq. (8). An example of a particular scenario for $\vec{f}(r)$ in Eq. (9) is:

$$\vec{f}(r) = \vec{a}(r) + \vec{d}(r) + \vec{c}(r) \quad (9)$$

The Additive Model will be analyzed to saturation as explained in Eq. (9). There are three elements of the model which are $\vec{a}(r)$, $\vec{d}(r)$, and $\vec{c}(r)$, and each element has its distinct definition.

$$\vec{a}(r) = \sum_{k=0}^{K_s-1} \vec{a}_k(\vec{s}(r - \tau_s(k))) \quad (10)$$

$$\vec{d}(r) = \sum_{k=0}^{K_r-1} \vec{d}_k(\vec{r}(t - \tau_r(k))) \quad (11)$$

$$\vec{r}(r - \tau_r(k)) = G(\vec{s}(r - \tau_r(k))) \quad (12)$$

$$\vec{c}(r) = \sum_{k=0}^{K_x-1} \vec{c}_k(\vec{x}(r - \tau_x(k))) \quad (13)$$

Using the given definitions, the system that resulted from replacing Eq. (10)-(13) into Eq. (6) and then incorporating the outcome into Eq. (1), can be expressed as follows:

$$\frac{c\vec{s}(r)}{dt} = \sum_{k=0}^{K_s-1} \vec{a}_k(\vec{s}(r - \tau_s(k))) + \sum_{k=0}^{K_r-1} \vec{d}_k(\vec{r}(r - \tau_r(k))) + \sum_{k=0}^{K_x-1} \vec{c}_k(\vec{x}(r - \tau_x(k))) + \vec{\phi} \quad (14)$$

$$\vec{r}(r - \tau_r(k)) = G(\vec{s}(r - \tau_r(k))) \quad (15)$$

The first component: denoted as $\sum_{k=0}^{K_s-1} \vec{a}_k(\vec{s}(r - \tau_s(k)))$, is a combination of up to K_s functions. These functions are obtained by time-shifting the original functions, $\vec{a}_k(\vec{s}(r))$, by specific delay time constants ($\tau_s(k)$). It's worth noting that the term "analog" emphasizes that each $\vec{a}_k(\vec{s}(r))$ depends on the

state signal itself, rather than the readout signal, that is the transformed version of the state signal.

The second component: $\sum_{k=0}^{K_r-1} \vec{d}_k(\vec{r}(r - \tau_r(k)))$, combines up to K_r functions. These functions are generated by applying time shifts $\tau_r(k)$ to the functions $\vec{d}_k(\vec{r}(r))$. These functions originate from the readout signal, which is defined in Eq. (10) and represents a warped (potentially binary-valued) version of the state signal.

The third component: $\sum_{k=0}^{K_x-1} \vec{c}_k(\vec{x}(r - \tau_x(k)))$, represents the external input; it is formed by aggregating up to K_x functions, obtained by time-shifting $\tau_x(k)$ the functions $\vec{c}_k(\vec{x}(r))$. These functions are associated with the input signal.

Consider a case where $\vec{a}_k(\vec{g}(r - \tau_s(k)))$, $\vec{d}_k(\vec{r}(r - \tau_r(k)))$, and $\vec{c}_k(\vec{x}(r - \tau_x(k)))$ are linear functions of \vec{s} , \vec{r} , and \vec{x} , respectively; here, Eq. (9) is a nonlinear DDE with linear coefficients, and these coefficients are represented as matrices:

$$\frac{c\vec{s}(r)}{cr} = \sum_{k=0}^{K_s-1} A_k(\vec{s}(r - \tau_s(k))) + \sum_{k=0}^{K_r-1} B_k(\vec{r}(r - \tau_r(k))) + \sum_{k=0}^{K_x-1} C_k(\vec{x}(r - \tau_x(k))) + \vec{\phi} \quad (16)$$

Moreover, in the case where the matrices A_k , B_k , and C_k are circulant or block circulant, the expression of the matrix-vector multiplication components given in Eq. (11) may be done as convolutions within the confines of the elements of \vec{s} , \vec{r} , \vec{x} , and $\vec{\phi}$. These convolutions are performed over each element indexed by i :

$$\frac{c\vec{s}(i,r)}{dr} = \sum_{k=0}^{K_n-1} a_k(i) * s(i, r - \tau_s(k)) + \sum_{k=0}^{K_r-1} d_k(i) * r(i, r - \tau_r(k)) + \sum_{k=0}^{K_z-1} c_k(i) * x(i, r - \tau_z(k)) + \phi(i) \quad (17)$$

It is important to note that previous research has established a link between a nonlinear dynamical system, as defined in Eq. (14), and the extension of a specific type of neural network. This connection was demonstrated when the functions $\vec{a}_k(\vec{s}(r))$, $\vec{d}_k(\vec{r}(r))$, and $\vec{c}_k(\vec{x}(r))$ were linear operators in Eq. (16), resulting in the Continuous Hopfield Network as a specialized case. A closely related version, described in Eq. (17), which further limits these operators to be convolutional, was found to incorporate the Cellular Neural Network.

Applying the simplifications:

$$\left. \begin{array}{lcl} K_s & = 1 \\ \tau_s(0) & = 0 \\ A_0 & = A \\ K_r & = 1 \\ \tau_r(0) & = \tau_0 \\ B_0 & = B \\ K_x & = 1 \\ \tau_x(0) & = 0 \\ C_0 & = C \end{array} \right\} \quad (18)$$

(Although some of these constraints will be relaxed later), incorporating them into Eq. (16) transforms it into:

$$\frac{c\vec{s}(r)}{cr} = A\vec{s}(r) + B\vec{r}(r - \tau_0) + C\vec{x}(r) + \vec{\phi} \quad (19)$$

Eqs. (16), (17), and (19) are examples of nonlinear first-order non-homogeneous DDEs. To solve these equations, as well as any other versions of Eq. (1), a common numerical

method is to discretize them in time. This entails computing the values of both the timestamp input and state signals, continuing this process for the needed total duration. This results in numerical integration.

$$\frac{r}{dr} = n\Delta T \quad (15)$$

$$\frac{c\vec{g}(r)}{dr} \approx \frac{\vec{g}(n\Delta T + \Delta T) - \vec{g}(n\Delta T)}{\Delta T} \quad (20)$$

$$A\vec{s}(r) + B\vec{r}(r - \tau_0) + C\vec{z}(r) + \vec{\phi} = A\vec{s}(n\Delta T) + B\vec{r}(n\Delta T - \tau_0) + C\vec{z}(n\Delta T) + \vec{\phi} \quad (21)$$

$$A\vec{s}(r + \Delta T) + B\vec{r}(r + \Delta T - \tau_0) + C\vec{z}(r + \Delta T) + \vec{\phi} = A\vec{s}(n\Delta T + \Delta T) + B\vec{r}(n\Delta T + \Delta T - \tau_0) + C\vec{z}(n\Delta T + \Delta T) + \vec{\phi}$$

$$\frac{\vec{g}(n\Delta T + \Delta T) - \vec{g}(n\Delta T)}{\Delta T} \approx A\vec{s}(n\Delta T + \Delta T) + B\vec{r}(n\Delta T + \Delta T - \tau_0) + C\vec{z}(n\Delta T + \Delta T) + \vec{\phi} \quad (22)$$

The calculation can be simplified by setting $\tau_0 = \Delta T$ and using an equal sign to replace the approximation sign in Eq. (22) as follows:

$$\frac{\vec{g}(n\Delta T + \Delta T) - \vec{g}(n\Delta T)}{\Delta T} = A\vec{s}(n\Delta T + \Delta T) + B\vec{r}(n\Delta T) + C\vec{x}(n\Delta T + \Delta T) + \vec{\phi} \quad (23)$$

$$\frac{\vec{g}((n+1)\Delta T) - \vec{g}(n\Delta T)}{\Delta T} = A\vec{s}((n+1)\Delta T) + B\vec{r}(n\Delta T) + C\vec{l}((n+1)\Delta T) + \vec{\phi} \quad (24)$$

$$\vec{s}((n+1)\Delta T) - \vec{s}(n\Delta T) = \Delta T(A\vec{s}((n+1)\Delta T) + B\vec{r}(n\Delta T) + C\vec{l}((n+1)\Delta T) + \vec{\phi}) \quad (25)$$

The result is that all the signals will be converted into sequences and their domain will be represented by the discrete index, 'n.' Thus, there will be a conversion of the Eq. (14) into a non-linear non-homogeneous first-order difference equation as follows:

$$\vec{s}[n+1] - \vec{s}[n] = \Delta T(A\vec{s}[n+1] + B\vec{r}[n] + C\vec{x}[n+1] + \vec{\phi}) \quad (26)$$

$$\vec{g}[n+1] = \vec{g}[n] + \Delta T(A\vec{s}[n+1] + B\vec{r}[n] + C\vec{x}[n+1] + \vec{\phi})$$

$$(I - (\Delta T)A)\vec{s}[n+1] = \vec{g}[n] + ((\Delta T)B)\vec{r}[n] + ((\Delta T)C)\vec{z}[n+1] + (\Delta T)\vec{\phi} \quad (27)$$

Defining:

$$W_s = (I - (\Delta T)A)^{-1} \quad (28)$$

and when we multiply both sides of Eq. (24) by W_s , we obtain: $\vec{s}[n+1] = W_s\vec{s}[n] + ((\Delta T)W_s B)\vec{r}[n] + ((\Delta T)W_s C)\vec{x}[n+1] + ((\Delta T)W_s \vec{\phi})$, and once we shift the index, n, forward by one step, it becomes:

$$\vec{s}[n] = W_s\vec{s}[n-1] + ((\Delta T)W_s B)\vec{r}[n-1] + ((\Delta T)W_s C)\vec{x}[n] + ((\Delta T)W_s \vec{\phi}) \quad (29)$$

$$\vec{r}[n] = G(\vec{s}[n])$$

By introducing two extra weight matrices and a bias vector,

$$W_r = (\Delta T)W_s B \quad (30)$$

$$W_x = (\Delta T)W_s C \quad (31)$$

$$\vec{\theta}_s = (\Delta T)W_s \vec{\phi} \quad (32)$$

the system mentioned above can be converted into the standard canonical RNN configuration:

$$\vec{s}[n] = W_s\vec{s}[n-1] + W_r\vec{r}[n-1] + W_x\vec{x}[n] + \vec{\theta}_s \quad (33)$$

$$\vec{r}[n] = G(\vec{s}[n]) \quad (34)$$

Having demonstrated the basic form of the RNN in Eq. (30), let's transition to the "LSTM" model. LSTM [20] is a robust recurrent neural network purposefully crafted to address the challenges associated with exploding or vanishing gradients. These difficulties tend to arise during the learning of long-term relationships, particularly when there are substantial time lags [20]. To address these challenges, the "LSTM" model incorporates a constant error carousel (CEC) that effectively preserves the error signal of each cell unit. Interestingly, these cells themselves are recurrent networks and have a unique architecture that extends the CEC with additional components, such as the output gate and input gate, to form the memory cell.

There are several components of the typical "LSTM" unit, including an input gate, an output gate, a forget gate, and a cell. The forget gate was introduced by Gers as it was not originally a part of the "LSTM" network; the introduction was to make the network able to reset its state. In the LSTM, the cell can hold data for any number of time periods, and the 3 gates work collectively to control the information flow within the cell. In the subsequent section, when we refer to LSTM, we are specifically addressing the basic version, which is the most widely employed architecture for "LSTM" networks [21]. However, it is important to recognize that being the most popular does not necessarily mean it is the best choice for every situation.

To put it briefly, the "LSTM" design consists of a set of memory blocks that are recurrently connected. The memory blocks maintain their state and oversee the control of information flow over time through the non-linear gating units. A typical "LSTM" block, as shown in Figure 6, includes gates, an input signal represented as $x^{(t)}$, an output represented as $y^{(t)}$, peephole connections, and activation functions. The output of the block forms recurrent connections with both the block input and all the gates.

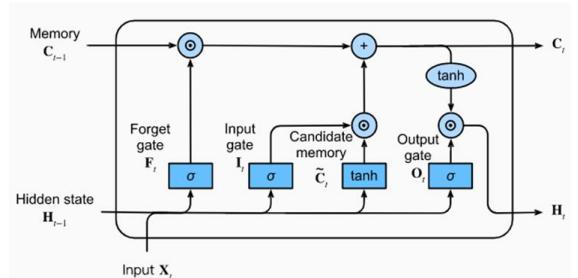


Figure 6. A typical vanilla "LSTM" block architecture

The working principle of the LSTM can better be understood by analyzing a network that comprises M inputs

and N processing blocks. In this recurrent neural system, the forward pass can be explained thus:

Block Input: The major process of this step is updating the component of the block input which involves combining the current input, denoted as $x^{(r)}$, and the output of the “LSTM” unit, $y^{(r-1)}$, from the preceding iteration; this process can be expressed thus:

$$z^{(r)} = g(W_z x^{(r)} + R_z y^{(r-1)} + d_z) \quad (35)$$

Input gate: The process here is the updating of the input gate that requires the incorporation of the output of the preceding “LSTM” unit $y^{(r-1)}$, the current input $x^{(r)}$, and the cell value $c^{(r-1)}$ from the preceding iteration as represented in the following equation:

$$i^{(r)} = \sigma(W_i x^{(r)} + R_i y^{(r-1)} + p_i \odot c^{(r-1)} + d_i) \quad (36)$$

where, \odot is a point -wise multiplication of 2 vectors, W_i, R_i and p_i are the weights related to $x^{(r)}$, $y^{(r-1)}$ and $c^{(r-1)}$, respectively, and d_i is the bias vector related with this specific component.

The “LSTM” layer is responsible for determining the data to be kept in the network's cell states from the previous step, represented as $c^{(r)}$. This process requires the selection of the candidate values $z^{(f)}$ to be added to the cell states, as well as the determination of the activation values $z^{(f)}$ of the input gates.

Forget Gate: The major task in this phase is the determination of the data to be discarded from its prior cell states $c^{(r-1)}$. This is achieved by computing the forget gate's activation values $f^{(r)}$ at time step t in consideration of the current input $x^{(r)}$, the outputs $y^{(r-1)}$, the memory cell's state $c^{(t-1)}$ from the previous time step (t-1), the bias terms b_f associated with the forget gates, and the peephole connections.

$$f^{(r)} = \sigma(W_f x^{(r)} + R_f y^{(r-1)} + p_f \odot c^{(r-1)} + b_f) \quad (37)$$

where, W_f, R_f and p_f are the weights of $x^{(r)}$, $y^{(r-1)}$ & $c^{(f-1)}$, respectively; b_f is the bias weight vector.

Cell: During the cell stage, the calculation of the cell value entails incorporating the block input $z^{(r)}$, input gate $i^{(r)}$, and forget gate $f^{(r)}$ values with the preceding cell value. This process can be expressed as follows:

$$c^{(r)} = z^{(r)} \odot i^{(r)} + c^{(r-1)} \odot f^{(r)} \quad (38)$$

Output Gate: Output Gate: The process here is to determine the output gate by merging the current input $x^{(r)}$, the cell value $c^{(r-1)}$ from the preceding iteration, and the output of the “LSTM” unit $y^{(r-1)}$ as seen in the following relation:

$$o^{(r)} = \sigma(W_o x^{(r)} + R_o y^{(r-1)} + p_o \odot c^{(r)} + d_o) \quad (39)$$

In the given equation, p_o , W_o and R_o are the weights of $c^{(r-1)}$ $x^{(r)}$ and $y^{(r-1)}$, respectively, while b_o denotes the bias weight vector.

Block Output: In the end, the block output is computed by combining the current cell value $c(t)$ with the current output gate value, as illustrated below:

$$y(r) = g(c(r)) \circ o(r) \quad (40)$$

where, σ , g , and h are non-linear activation functions that are

applied to each element.

4. RESULTS AND DISCUSSION

Several metrics were employed for the evaluation of the proposed network in predicting power consumption accurately. These metrics encompass RMSE, MSE, and R2 score.

(1) RMSE: RMSE reflects the mean deviation between the predicted and actual energy consumption values. In this context, an RMSE value of 0.611 suggests an average deviation of 0.611 units between the actual and predicted energy consumption values. A lower RMSE value suggests a higher accuracy of the predictions [22].

(2) MSE: The MSE, with a value of 0.374, indicates that, on average, the squared difference between the predicted power consumption values and the actual values is approximately 0.374 units. A lower MSE value also implies greater accuracy in the predictions [23].

(3) R2 Score: The R2 score for the test data is 0.504. This score indicates the percentage of the variation in power consumption that can be accounted for by the predicted values. A higher R2 score indicates a better alignment of the predictions with the actual values. An R2 score of 0.504 suggests a moderate level of explanatory power, meaning that around 50.4% of the variance in power consumption is explained by the predictions [24].

The training graph illustrates the loss values for the validation and training datasets throughout the training procedure. Loss measures the disparity between the actual and predicted energy consumption values. Within the Figure 7, the training loss curve represents the error on the training set, while the validation loss curve signifies the error on the validation set. The objective is to minimize both the validation and training loss, as it reflects the learning and generalization abilities of the model.

Ideally, the desired outcome is to observe a decreasing trend in both the training and validation loss curves. Such a trend suggests that the model is enhancing its performance over time and becoming proficient at capturing the underlying patterns in the power consumption data.

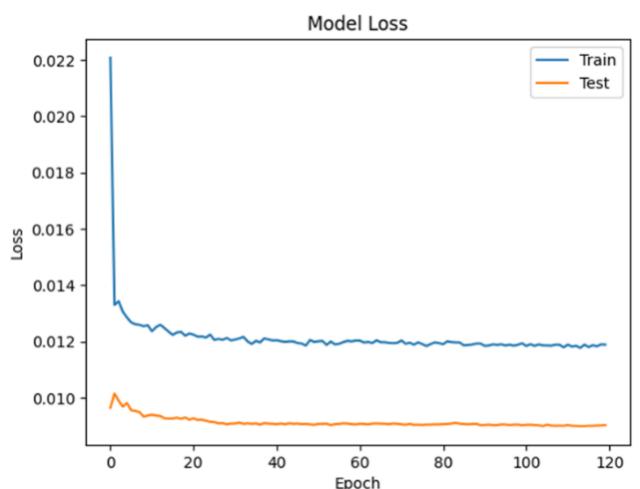


Figure 7. Training model

Figure 8 represents the actual power consumption values and a scatter plot that represents the actual and predicted energy consumption values. The line plot shows the actual

energy consumption values as a series of data points connected by lines. Each data point is marked with an asterisk symbol. The predicted energy consumption values are shown in the scatter plot, where each data point is captured as circles. The y-axis has the power consumption values while the x-axis has the time steps for the first 500 h. This figure enhances better comparison of the predicted and actual power consumption values for better understanding of the model's prediction performance.

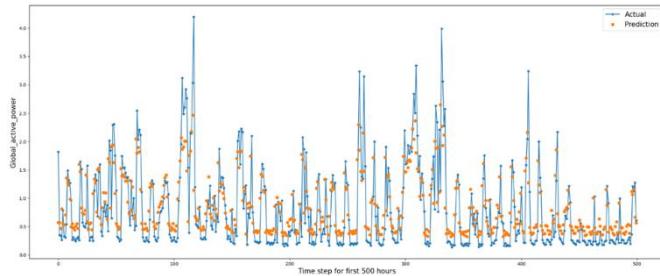


Figure 8. Actual vs predicted power consumption values for the first 500 h

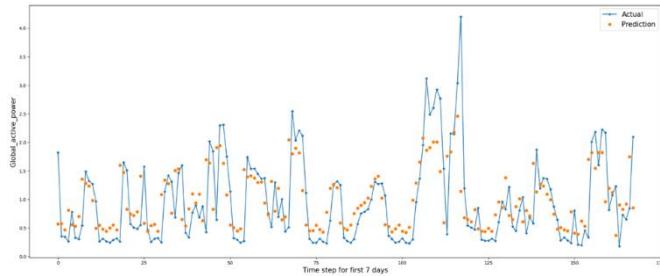


Figure 9. The predicted vs actual power consumption values for the first 7 days

The blue line is the actual energy consumption data for each time step for the period of 7-days; each data point relates to the actual energy usage value at the related time step. The red line are the predicted energy consumption data for each time step over the same 7-day period; each circle relates to the predicted energy usage data at the related time step. The model's prediction performance is visualized better by comparing the actual and predicted data in Figure 9; a closer match between the red and blue lines indicates a stronger model performance. Table 1 presents the RMSE for each day and the RMSE sum for the full 7 days.

Table 1. The RMSE for each day and the RMSE sum

Day	RMSE
1	0.498
2	0.634
3	0.601
4	0.325
5	0.880
6	0.389
7	0.600
Sum	3.927

The RMSE is a way of determining the average difference between the predicted and actual values; lower RMSE value implies higher model accuracy and vice versa.

LSTM- FDLE-based power consumption prediction was conducted in this work; the performance of the developed

method was benchmarked against other methods in terms of the RMSE, MSE, and R² values. In this study, the RMSE value was 0.611, indicating an average difference of 0.611 units between the actual and predicted power consumption values. The MSE was 0.374, suggesting an average squared difference of approximately 0.374 units. The R² score was 0.504, representing a moderate level of explanatory power. The study presented by Cascone et al. [15] used "LSTM" to predict global active energy consumption and achieved an RMSE of 0.617%, indicating a slightly higher deviation than this study [15]. Another study presented by Qin [25] used both a linear regression model and a neural network model to predict power consumption (see Table 2 for the comparison of the performance of the developed model in this study with other methods).

Table 2. Comparison of the performance of the proposed model in this study with other methods

Study	RMSE	MSE	R ² Score
Cascone et al. [15]	0.617%	-	-
Qin [25]	-	-	0.409
Current Study	0.611	0.374	0.504

5. CONCLUSIONS

Energy efficiency in residential homes has become a major concern to facility managers as they strive towards energy consumption minimization during the operation and maintenance of such facilities. The early prediction of energy use patterns of residential buildings may aid building managers in making decisions regarding energy consumption reduction. This paper introduces an innovative model for forecasting energy demand through the utilization of a long short-term memory (LSTM) as well as fractional differential equations (FDLE) based model. The proposed LSTM-FDLE model is trained to predict the collective active power generated by household appliances. By leveraging the memory and sequential learning capabilities of LSTM, the proposed model can capture complex temporal dependencies and patterns in energy consumption data. The efficiency of the suggested LSTM- FDLE based model is evaluated using real-world household energy consumption data. Future works in this regard are intended towards adopting the developed methods for real-time energy consumption prediction, as well as introducing a feedback mechanism for the minimization of the RMSE score of the proposed method for real-world applications.

The proposed method contributions of the current study are as follows:

(1) The issue of making accurate energy demand predictions for residential houses which is considered important for the optimization of energy generation and distribution in a sustainable manner was solved. A novel model that relies on both LSTM and modified Fractional calculus was developed and used to solve the problems of the existing prediction models that mostly rely on one-step prediction.

(2) The effective prediction of household energy consumption using the suggested LSTM-based model was demonstrated; the model reached a good level of accuracy and performed better than most of the existing one-step forecasting approaches.

(3) The possibility of a longer forecasting period leveraging the "LSTM" model was proven as it enables better planning

and optimization of energy production and distribution systems. The accurate prediction of energy demand over an extended period could aid decision-makers in making better resource allocation decisions, load balancing, and infrastructure planning.

REFERENCES

- [1] Amasyali, K., El-Gohary, N.M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81: 1192-1205. <https://doi.org/10.1016/j.rser.2017.04.095>
- [2] Pham, A.D., Ngo, N.T., Truong, T.T.H., Huynh, N.T., Truong, N.S. (2020). Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability. *Journal of Cleaner Production*, 260: 121082. <https://doi.org/10.1016/j.jclepro.2020.121082>
- [3] Guo, Y., Wang, J., Chen, H., Li, G., Liu, J., Xu, C., Huang, R., Huang, Y. (2018). Machine learning-based thermal response time ahead energy demand prediction for building heating systems. *Applied Energy*, 221: 16-27. <https://doi.org/10.1016/j.apenergy.2018.03.125>
- [4] Altamimi, A.S.H., Al-Dulaimi, O.R.K., Mahawish, A.A., Hashim, M.M., Taha, M.S. (2020). Power minimization of WBSN using adaptive routing protocol. *Indonesian Journal of Electrical Engineering and Computer Science*, 19(2): 837-846.
- [5] Aslam, S., Herodotou, H., Mohsin, S.M., Javaid, N., Ashraf, N., Aslam, S. (2021). A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*, 144: 110992. <https://doi.org/10.1016/j.rser.2021.110992>
- [6] Li, Z., Dai, J., Chen, H., Lin, B. (2019). An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. In *Building Simulation*, 12: 665-681. <https://doi.org/10.1007/s12273-019-0538-0>
- [7] Ramos, D., Faria, P., Morais, A., Vale, Z. (2022). Using decision tree to select forecasting algorithms in distinct electricity consumption context of an office building. *Energy Reports*, 8: 417-422. <https://doi.org/10.1016/j.egyr.2022.01.046>
- [8] Nabavi-Pelesarai, A., Rafiee, S., Hosseini-Fashami, F., Chau, K.W. (2021). Artificial neural networks and adaptive neuro-fuzzy inference system in energy modeling of agricultural products. In *Predictive Modelling for Energy Management and Power Systems Engineering*, pp. 299-334. <https://doi.org/10.1016/B978-0-12-817772-3.00011-2>
- [9] Ghimire, S., Nguyen-Huy, T., Deo, R.C., Casillas-Perez, D., Salcedo-Sanz, S. (2022). Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. *Sustainable Materials and Technologies*, 32: e00429. <https://doi.org/10.1016/j.susmat.2022.e00429>
- [10] Batlle, E.A.O., Palacio, J.C.E., Lora, E.E.S., Reyes, A.M.M., Moreno, M.M., Morejón, M.B. (2020). A methodology to estimate baseline energy use and quantify savings in electrical energy consumption in higher education institution buildings: Case study, Federal University of Itajubá (UNIFEI). *Journal of Cleaner Production*, 244: 118551. <https://doi.org/10.1016/j.jclepro.2019.118551>
- [11] Qiao, R., Liu, T. (2020). Impact of building greening on building energy consumption: A quantitative computational approach. *Journal of Cleaner Production*, 246: 119020. <https://doi.org/10.1016/j.jclepro.2019.119020>
- [12] Li, B., Xie, W. (2015). Adaptive fractional differential approach and its application to medical image enhancement. *Computers & Electrical Engineering*, 45: 324-335. <https://doi.org/10.1016/j.compeleceng.2015.02.013>
- [13] Zhang, Y., Yang, L., Li, Y. (2022). A novel adaptive fractional differential active contour image segmentation method. *Fractal and Fractional*, 6(10): 579. <https://doi.org/10.3390/fractfrac6100579>
- [14] Teodoro, G.S., Machado, J.T., De Oliveira, E.C. (2019). A review of definitions of fractional derivatives and other operators. *Journal of Computational Physics*, 388: 195-208. <https://doi.org/10.1016/j.jcp.2019.03.008>
- [15] Cascone, L., Sadiq, S., Ullah, S., Mirjalili, S., Siddiqui, H.U.R., Umer, M. (2023). Predicting household electric power consumption using multi-step time series with convolutional LSTM. *Big Data Research*, 31: 100360. <https://doi.org/10.1016/j.bdr.2022.100360>
- [16] Chatfield, C. (1986). Exploratory data analysis. *European Journal of Operational Research*, 23(1): 5-13. [https://doi.org/10.1016/0377-2217\(86\)90209-2](https://doi.org/10.1016/0377-2217(86)90209-2)
- [17] Barriga, R., Romero, M., Hassan, H., Nettleton, D.F. (2023). Energy consumption optimization of a fluid bed dryer in pharmaceutical manufacturing using EDA (Exploratory Data Analysis). *Sensors*, 23(8): 3994. <https://doi.org/10.3390/s23083994>
- [18] Deepa, B., Ramesh, K. (2022). Epileptic seizure detection using deep learning through min max scaler normalization. *International Journal of Health Sciences*, 6: 10981-10996. <https://doi.org/10.53730/ijhs.v6ns1.7801>
- [19] Gal, M.S., Rubinfeld, D.L. (2019). Data standardization. *NYUL Review*, 94, 737.
- [20] Hochreiter, S., Schmidhuber, J. (1996). LSTM can solve hard long time lag problems. *Advances in Neural Information Processing Systems*, 9.
- [21] Greff, K., Srivastava, R.K., Koutník, J., Steunebrink, B.R., Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10): 2222-2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- [22] Hameed, R.S., Mokri, S.S., Taha, M.S., Taher, M.M. (2022). High capacity image steganography system based on multi-layer security and LSB exchanging method. *International Journal of Advanced Computer Science and Applications*, 13(8): 108-115.
- [23] Taha, M.S., Rahem, M.S.M., Hashim, M.M., Khalid, H.N. (2022). High payload image steganography scheme with minimum distortion based on distinction grade value method. *Multimedia Tools and Applications*, 81(18): 25913-25946. <https://doi.org/10.1007/s11042-022-12691-9>
- [24] Chai, T., Draxler, R.R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE in the literature. *Geoscientific*

Model Development, 7(3): 1247-1250.
https://doi.org/10.5194/gmd-7-1247-2014
[25] Qin, J. (2022). Experimental and analysis on household

electronic power consumption. Energy Reports, 8: 705-709. https://doi.org/10.1016/j.egyr.2022.02.270