Received: February 8, 2024. Revised: March 5, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

A CNN and BiLSTM Fusion Approach Toward Precise Appliance Energy Forecasts

Khush Attarde[1] **Javed Sayyad**[1]*

[1]*Department of Robotics and Automation, Symbiosis Institute of Technology (SIT), Symbiosis International (Deemed University) (SIU), Lavale, Pune, Maharashtra 412115, India*
* Corresponding author's Email: jksayyad23@gmail.com

Abstract: Optimizing the energy grid management will enhance the effective and efficient use of generated energy. Accurate energy estimations allow power-generating firms to use dynamic energy management strategies to maintain stability across the smart grid. This paper presents a hybrid predictive modelling approach for forecasting the energy consumption of household appliances by combining 1D Convolutional Neural Networks (1D-CNN) with Bi-Directional Long Short-Term Memory (BiLSTM). The hybrid architecture combines CNNs' spatial feature extraction capabilities with BiLSTMs' sequential memory retention to comprehensively understand appliance use patterns. The proposed model is trained and validated on a dataset of household energy consumption, demonstrating superior performance compared to individual CNN or BiLSTM models. By incorporating both spatial and temporal data, energy consumption forecasts become more accurate and adaptable, making them highly appropriate for real-time applications and demand-side management. Utilizing the Analysis of Variance (ANOVA) F-measure feature selection technique and BiLSTM with a 1D-CNN hybrid deep learning model, the Root Mean Square Error (RMSE) of 1.745 and Coefficient of Determination (R^2 score) of 0.997.

Keywords: ANOVA, Bi-directional long short-term memory, Convolutional neural networks, Energy forecasting, Hybrid model, Machine learning.

1. Introduction

With the world's growing population, urbanization, and technological advancements increasing, energy demand has enhanced exponentially. Efficient energy management at the household level is essential to address the broader issue of resource sustainability. It's also a way for individuals to contribute to global efforts to reduce carbon emissions to prevent climate change and environmental concerns. Integrating renewable energy sources with efficient energy management facilitates the transition to cleaner and greener energy alternatives. It also promotes economic stability by reducing utility expenses and enhancing overall energy stability.

As the demand increases, predicting the energy consumption of household appliances at regular intervals helps homeowners make knowledgeable decisions about their energy usage patterns, leading to significant cost savings. It also allows for optimizing the energy efficiency of specific appliances, leading to substantial reductions in utility costs. Understanding weather and climate data in energy management models helps when and how specific appliances consume energy, enabling homeowners to adjust their usage patterns dynamically [1]. Usage patterns, such as load shifting, can help reduce electricity rates during off-peak hours and eventually benefit in cost-saving, sustainability, etc. Routine energy consumption predictions align with sustainability goals, enabling households to participate actively in global energy conservation [2].

Optimizing household energy usage involves multiple factors influencing total consumption patterns, from the house's physical characteristics to its occupants' behaviours. Various aspects contribute to the overall energy landscape, like the size and layout of a house, insulation and efficient windows and doors, and the number and types of appliances in a household. External elements such as weather and climate data also influence the efficiency and sustainability of energy utilization within a home. Smart home systems with Internet of Things (IoT) enabled sensors allow real-time monitoring and data collection [3].

Temperature, pressure, humidity, wind speed, visibility, and dewpoint are vital parameters that directly impact energy consumption. Temperature, for instance, directly correlates with heating and cooling demands. In colder climates, households may require more energy to maintain warmness, while warmer climates require increased cooling efforts [4]. Daily temperature fluctuations and extreme weather events influence a household's energy needs. Pressure variations affect the efficiency of gas-powered appliances, and humidity levels influence the perceived comfort and efficiency of cooling systems. Wind speed and visibility can impact solar panel efficiency, affecting the performance of renewable energy systems. Dewpoint, the temperature at which air becomes saturated and dew forms, is a crucial factor for understanding moisture levels in the air. High dewpoints can result in discomfort, potentially leading to increased reliance on Heating, Ventilating, and Air Conditioning (HVAC) system systems.

The proposed method combines, a 1D-CNN with a BiLSTM to enhance energy consumption prediction. It involves the sequential application of 1D-CNN followed by BiLSTM layers. The open-source Appliances Energy Dataset, available on the UCI-ML Repository, is used to evaluate the approaches. Despite the importance of physical room characteristics and homeowner behaviour as factors influencing energy consumption, they are challenging to quantify numerically. The time series dataset used encompasses the temperature and relative humidity of different rooms in a house and various external environmental attribute around a house. A detailed description of the dataset is presented in Section 3.1.

An essential aspect of this analysis is the examination of factors influencing energy usage. The usage of feature engineering methodologies mostly determines the accuracy of prediction using various Machine Learning (ML) models. This challenge extends to Deep Learning (DL) models like CNN and Long Short Term Memory (LSTM) making researchers delve into multiple feature extraction and selection techniques specifically for time series data. Incorporating an Analysis of Variance (ANOVA) feature selection proves particularly beneficial as it excludes unnecessary and redundant features.

The contribution of this paper is listed as follows -

1) ANOVA F-Measure technique for feature selection to ignore redundant features
2) Hybrid approach featuring 1D-CNN and BiLSTM.
3) Enhancing the prediction of the energy consumption of household appliances.

Further in this paper, in Section 2, the literature review and previous related work are analyzed and discussed thoroughly. In Section 3, the proposed methodology and its components are discussed briefly. In Section 4, the performance is evaluated, and the results are discussed briefly with comparison. Section 5 concludes the paper with a conclusion and future scope.

2. Related work

2.1 Review of available dataset

There are different datasets aimed at forecasting the residential household energy load or consumption. These datasets range from multiple input parameters like internal and external surrounding parameters, meter readings [5, 6], and many more. The dependency of these parameters for predicting energy consumption also depends on the dataset variability and inter-dependency with the output variable. The proper understanding of the input parameters and output parameters can be achieved well using various complex and simple models for forecasting.

The Individual Household Electric Power Consumption (IHEPC) dataset [7] consists of meter readings, relative world average power, and many different parameters related to household energy data. The dataset [8] related to energy consumption was collected in London, where the input parameters are read from the smart meters. The Dutch Residential Energy Dataset (DRED) [9] is the Netherlands distribution comprised of numerous sensors that monitor environmental parameters, consumption of electricity, and the number of individuals living.

The total energy of the appliances used in a household is collected concerning the surrounding internal and external variables, which will be inputted for the prediction of energy consumption Datasets like the Domestic Electricity Demand Dataset of Individual Appliances in Germany (DEDDIAG) [10] and [11] are collected with the help of micro-processes at the residential houses. The proposed strategy [12] highlights the relationship between the energy consumption of appliances and environmental variables. Understanding the energy consumption behavior of appliances due to outside and inside environmental parameters is recorded and studied in this dataset by researchers. This dataset is selected for further exploration and this dataset description can be seen in Section 3.1.

2.2 Review of ML/DL models

Over the years, researchers have dedicated significant efforts to predicting energy consumption using a diverse range of ML/DL approaches. These studies have involved testing ML algorithms on datasets containing previously collected sensor data, including Support Vector Regressor (SVR) and Decision Tree Regressor (DTR) [13]. Artificial Neural Networks (ANN) have also been explored for their efficacy in predicting energy consumption trends. Additionally, the LSTM model, a type of Recurrent Neural Network (RNN), has been applied to forecast energy usage. These diverse approaches have undergone testing across various datasets, all characterized by time series data capturing energy consumption patterns.

Several papers have been undertaken to scrutinize and analyse the electricity consumption in a residential building. Researchers [14] use a fuzzy network model for the prediction of energy consumption. Further, it also discusses the models using an energy management system for saving electricity. Authors [15] used the neural fuzzy stem to enhance and monitor energy efficiency.

Researchers used an ANN as the basic DL model to predict the energy in residential buildings in British Columbia [16]. To design a retrofit model, researchers in [17] used the ANN for the prediction task. In [18] the researchers utilized the Multi-Layer Perceptron (MLP) and SVR techniques for forecasting energy consumption. For a Tsanas and Xifara dataset [19] they achieved the Root Mean Square Error (RMSE) of 2.626. SVR also performs better in their experimentation. For SVR, the RMSE of 3.4 has been achieved by authors [20].

Time series analysis has been performed on the dataset to predict the energy consumption using the Autoregressive Integrated Moving Average (ARIMA) model [21, 22] Electricity consumption of a hospital is analyzed in the [23], which uses the ARIMA model for the forecasting. Due to over-fitting, the performance of neural networks cannot be improved further by just adding layers to that dataset. This study [24] proposes a deep learning model and RNN fusion that batches a set of data into an input pool.

Hybrid models have also been implemented on the various energy prediction datasets. These hybrid models, such as CNN-LSTM [17], CNN-Gated Recurrent Unit (GRU) [18], Multi-information Fusion Deep Learning (MFDL) [19], and many more, have been implemented by the researchers. Compared to prior studies, hybrid CNN and LSTM models from [25] have yielded good results but still need improvement. They have not yet been able to be considered as better outcomes in that domain. Two tiers of information extraction are proposed where the lower and higher-level LSTM and CNN are implemented, respectively.

In [26], researchers have been applied to a dataset aiming to forecast household energy prediction in an individual house. In this study [27], the proposed structure's primary two phases include training and data refinement. CNN features are taken from the input data set and fed into GRU during the training phase, where GRU can learn and adapt the sequential model. GRU-based models are more straightforward and contain fewer gradient flow gates than LSTM-based models, reducing their volatility. Due to CNNs' ability to extract representative features and their efficacy for forecasting results.

This study [20] has integrated an attention mechanism with the RNN-based type models like GRU, LSTM, etc. It was found that longer input patterns do not typically contribute to improved accuracy; accuracy falls as the prediction range increases. This has been implemented on the dataset to predict the load.

Many researchers have experimented with hybrid models containing a BiLSTM with different neural networks for the prediction of the time series data. These time series data applications range from stock prediction and more. On the dataset related to energy consumption, it was explored less. Authors [28] conducted experiments to compare the performance of the CNN-BiLSTM hybrid model applied on IHEPC for prediction. To forecast electric energy consumption, the proposed structure performs more effectively than different approaches over a range of performance criteria, including narrow and long-term periods.

In [29], research introduces an advanced hybrid approach that integrates a CNN with a Multiple-layer BiLSTM approach for efficiently predicting energy consumption in power systems. With the incorporation of a three-layer BiLSTM model, this approach focuses on power management. It eliminates irregularities and enhances the quality of the data. Improved data sequences are processed by a DL network to allow effective prediction-making. The last phase produces probability metrics by comparing the actual and expected data. This hybrid

BiLSTM model has been applied to IHEPC [7] for prediction.

For the selected appliance energy dataset, a stacked model comprises LSTM and BiLSTM for prediction able to capture the temporal relationships, effective computational time and complexity, etc. The sequence component of time series data is the focus of LSTM and BiLSTM models, which indicates significant spatial patterns are ignored. They can also absorb information slowly while they do it step-by-step. In [30] different ML and DL models have been applied to household appliance data for forecasting energy consumption where they employed the Pearson Correlation Coefficient (PCC) for the selection of the enhanced features. PCC measures simple linear relationships between continuous variables but may not fully capture complex relations between input and target variables showing researchers to consider variables and research questions to determine the most appropriate method for feature selection. They had tested over a wide range of models. In a study [31], the LSTM model with minmax scaling data to [-1,1] and selection of data based on covariance is implemented on the dataset. For the selected dataset the experimental results can be seen in Section 4.

2.3 Summary of literature review

The primary focus of the discussion is the application of different hybrid DL/ML models to predict the energy consumption of appliance datasets. However, these datasets show an evident absence of use of the room's surroundings and surrounding environment. More specifically, there is a lack of research and exploration on datasets that include features like temperature and relative humidity from various rooms in a home. This offers a chance for more study and advancement in modelling and predicting energy usage, which can be found in selected dataset and according to previous related work a hybrid DL model is proposed to enhance the prediction efficiency.

3. Proposed methodology

In this section, the proposed methodology and its respective tools and techniques are discussed thoroughly. The generalized flow of the proposed methodology can be found in

3.1 Dataset description

In this paper, an open-access Appliances Energy Consumption (AEC) dataset from the (University of California, Irvine Machine Learning) UCI ML

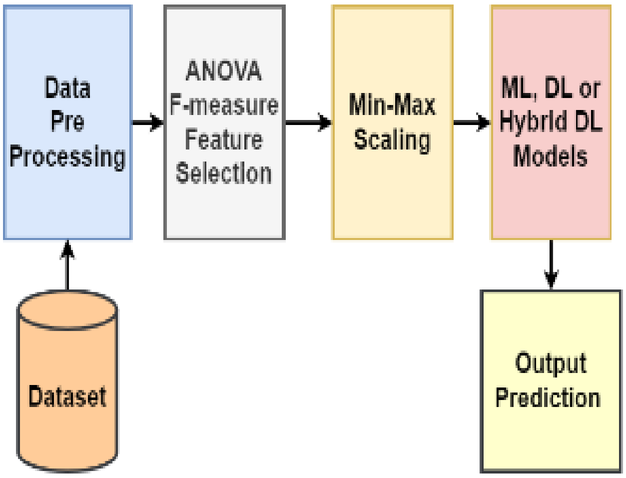

Figure .1 Generalized Block Diagram of Proposed Methodology

Repository, is used to evaluate a proposed methodology and predict the total energy consumption of household appliances. In this dataset, total energy consumption depends on various features like temperature and relative humidity of rooms in a house and temperature, relative humidity, wind speed, and dewpoint. This dataset is collected by Luis Candanedo using a Zigbee module for collecting a house's temperature and relative humidity. The external features around the house are collected from the nearest weather station, Chievres Airport, Belgium [11].

The researchers from Reliable Prognosis collect and download hourly data from the weather station. Apart from logged data, the author [11] has also included two non-dimensional random variables. The features in the dataset with proper units are tabulated in [11] A dataset is collected from 11/01/2016 to 27/05/2016 at 10-minute intervals. Among the total energy consumption of appliances, this dataset also includes the total energy consumption of light, which was ignored here in this study.

3.2 Data pre-processing

Data pre-processing plays a crucial role in enhancing the efficiency of ML models by optimizing computational power and ensuring the focus on the most relevant information.

3.2.1. Data visualization

A thorough dataset analysis is essential for selecting appropriate techniques for data pre-processing. The dataset comprises a total of 28 features and includes 19,735 data points for each feature. To ensure data integrity, null values were examined to remove such values from the dataset. It

Received: February 8, 2024. Revised: March 5, 2024.

Table 1. Feature Set in a dataset

Parameter	Description	Unit	I/O
T_1	Kitchen's temperature	°C	Input Variable - Inside House Parameters
T_2	Living Room's temperature		
T_3	Laundry Room's temperature		
T_4	Office Room's temperature		
T_5	Bathroom's temperature		
T_6	Ironing Room's temperature		
T_7	Outside building's temperature		
T_8	Room 1's temperature		
T_9	Room 2's temperature		
RH_1	Kitchen's relative humidity	%	
RH_2	Living Room's relative humidity		
RH_3	Laundry Room's relative humidity		
RH_4	Office Room's relative humidity		
RH_5	Bathroom's relative humidity		
RH_6	Ironing Room's relative humidity		
RH_7	Outside building's relative humidity		
RH_8	Room 1's relative humidity		
RH_9	Room 2's relative humidity		
T_{out}	Temperature from Weather Station	°C	Input variable - Outside House Parameters
Press_mm_Hg	Pressure from Weather Station	mm-Hg	
RH_{out}	Relative Humidity from Weather Station	%	
Windspeed	Wind Speed from Weather Station	m/s	
Visibility	Visibility from Weather Station	km	
rv_1, rv_2	Non-dimensional Random Variable		Input Variable
Date	Date and Timestamp of the data collected		Input variable
Appliances	Total Energy consumption of an Appliance	Wh	Output variable

was observed that the repetition of values there within the dataset as it is measured by the sensor within a 10-minute interval. The measured values of features in the dataset also show wide-ranging variations.

Given these characteristics, it becomes necessary to implement proper scaling techniques and properly select features to enhance the dataset's quality for subsequent ML/DL model training. An overall appliance energy consumption value distribution according to time instances can be found in Figure. .

3.2.2. Data scaling

The Minimum-Maximum (MinMax) Scaler is designed to standardize and shrink the data within a specified range, often from 0 to 1. To bring all features to a consistent and comparable scale, eliminating biases that may occur due to variations in the original feature magnitudes. The shape of the data's distribution in the MinMax Scale tried to be kept intact, and it is beneficial in scenarios where ML models are sensitive to the scale of input features, enabling more effective learning and improved accurate predictions. The model can capture the underlying patterns and variations in the data. The formula for MinMax scaling can be expressed in Eq. (1) and Eq. (2)

$$X_{[norm]} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

$$X_{[minmaxscaled]} = X_{[norm]}[max - min] + min \tag{2}$$

Where X, X_{max}, X_{min} is the original value, the maximum and minimum value in each feature, respectively; min, max is the minimum and maximum value in a defined range for normalization for transformation, respectively; $X_{[norm]}$ is the fraction for scaling of the features, and $X_{[minmaxscaled]}$ is the transformed MinMax scaled of the features.

3.3 Feature selection technique

The ANOVA F-test measurement is the most statistical feature selection technique for eliminating the redundant features overall. This method assesses whether the means of two or more features are significantly different, making it a valuable tool for identifying features that contribute significantly to the observed variability [32]. The backbone of the ANOVA is to partition the total variability in the dataset into different components, allowing the evaluation of the variability between groups relative to the variability within groups.

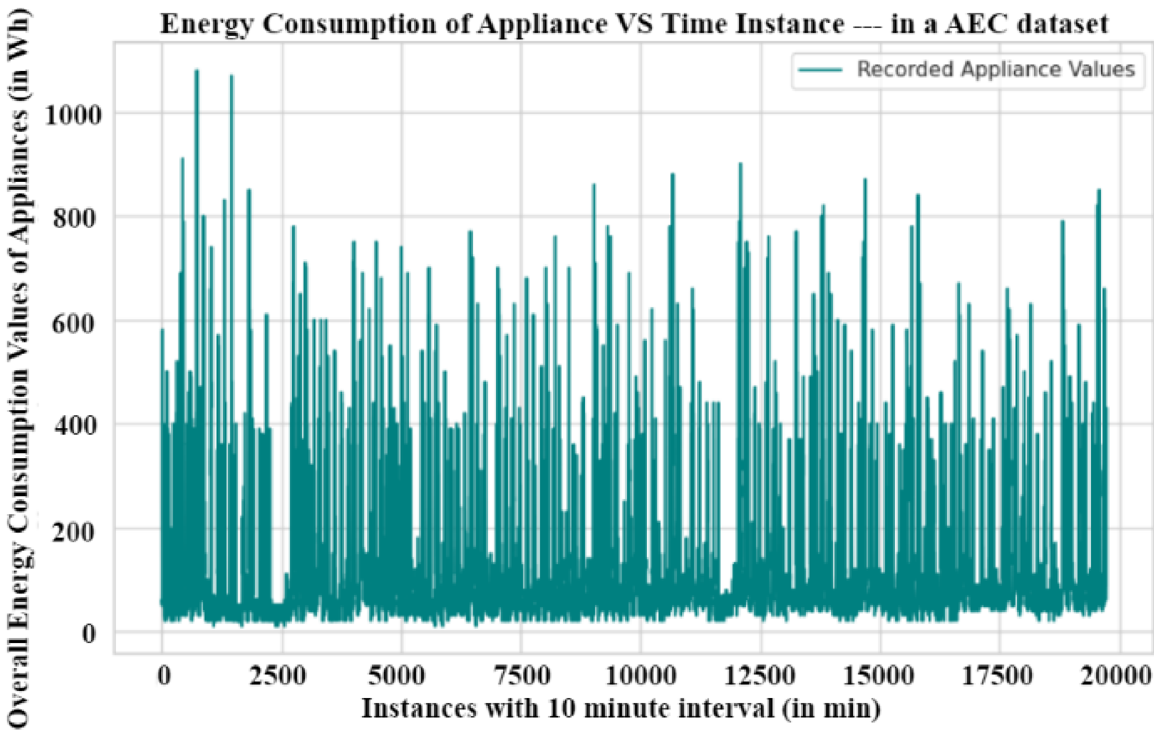

Figure. 2 Energy Consumption of Appliances VS Time Instances in an AEC dataset

If the calculated F-statistic is greater than the critical value, it indicates significant differences between at least two group means. It is applied to feature value by treating them as groups. The F-statistic for each feature helps evaluate whether the means of these feature values significantly differ, making it possible to identify features that contribute significantly to the variance in the target variable. The F-statistic, a key component of the ANOVA F-test, is calculated by Eq. (5).

$$\sigma^2_{[between-grps.]} = \frac{\Sigma_{i=1}^{l} l_i[\bar{x}_i - \mu]}{[N_f - 1]} \quad (3)$$

$$\sigma^2_{[within-features]} = \frac{\Sigma_{i=1}^{N_f} \Sigma_{p=1}^{l_i} [x_{ip} - \bar{x}_i]}{[N - N_f]} \quad (4)$$

$$F - score = \frac{\sigma^2_{[between-features]}}{\sigma^2_{[within-grps.]}} \quad (5)$$

Where l_i is the number of observations in l^{th} group, μ is the total mean of the dataset, $\bar{x}_i$ is the mean of samples in feature 'i', N is the total sample size of the dataset, N_f is the sample size of a respective feature 'f', x_{ip} is the p^{th} observations in i^{th} feature, $\sigma^2_{[between-grps]}$ is the variance between the features Eq. (3), $\sigma^2_{[within-features]}$ is the variance within the features in Eq. (4), *F-score* is an F-measure value used for comparison.

The steps in the ANOVA F-test are listed as follows;

1) The dataset is divided based on the features the target variable represents.
2) For each feature, the F-statistic is evaluated.
3) The significance is obtained by comparing it with a critical value.
4) Features with significant F-measures are considered relevant contributors to the variability in the target variable.

3.4 Deep learning models

This section discusses different DL models such as 1D-CNN, LSTM, and BiLSTM, which will be investigated further to estimate energy usage.

3.4.1. One-dimensional convolutional neural network (1D-CNN)

1D-CNN is a neural network architecture used here for processing sequential data or one-dimensional signals. The architecture of 1D-CNN comprises convolutional layers that scan input data with one-dimensional filters, capturing local patterns and learning hierarchical representations [33]. The first layer of a 1D-CNN typically involves convolutional operations, where filters slide over the input data, extracting features through convolutional and pooling operations. Convolutional layers are crucial for learning spatial hierarchies and recognizing patterns with different levels of abstraction. The subsequent pooling layers downsample the output of the convolutional layers, reducing the spatial dimensions for retaining the most needful features.

Further fully connected layers are incorporated, which enables learning relationships in the data. These layers allow the features learned by the convolutional layers to make predictions based on the extracted representations. Activation functions, such as Rectified Linear Unit (ReLu), are commonly applied after each layer to introduce non-linearity and enhance the model's capacity to capture complex patterns. prediction.

3.4.2. Long short-term memory (LSTM)

LSTM models are a part of RNN architecture used for capturing long-term dependencies in sequential data. LSTM consists of a memory cell that enables the network to store and retrieve information selectively. At the core of the LSTM architecture is the LSTM cell, a fundamental building block that has the model's ability to learn and remember information over varying time scales [34]. For prediction is done by forwarding it to further LSTM cells or fully connected layers.

$$F_{[s]} = \sigma[W_F.X_s + W_F'.H_{s-1} + B_F] \tag{6}$$

$$I_{[s]} = \sigma[W_I.X_s + W_I'.H_{s-1} + B_I] \tag{7}$$

$$C_{[s]} = f[W_C.X_s + W_C'.H_{s-1} + B_C] \tag{8}$$

$$O_{[s]} = \sigma[W_O.X_s + W_O'.H_{s-1} + B_O] \tag{9}$$

$$\tilde{C}_{[s]} = F_s.C_{s-1} + I_s.C_s \tag{10}$$

$$H_{[s]} = O_s.f[C_s] \tag{11}$$

$$Y_{[s]} = W_{Y_H}.H_s + B_Y \tag{12}$$

Where at interval s, X_s, F_s, H_s, I_s, C_s, O_s, and Y_s is the input, output of forget gate, hidden state, output of input gate, cell state, output of output gate and output of model at interval s, *1*respectively; W_F, W'_F, B_F, W_I, W'_I, B_I, W_O, W'_O, B_O and W_C, W'_C, B_C is the weight, hidden layer weight and bias of forget, input, output gate, and cell state respectively, f, σ is the associated activation function; B_Y is the bias associated with the output layer, W_{Y_H} is the weight associated with the hidden output state.

The LSTM cell comprises a cell state, an input gate, a forget gate, an output gate, and various weights and biases. The cell state is used to forward information across time steps. The input gate regulates the flow of new information into the cell state, determining which elements are updated and to what extent. The forget gate decides which information from the previous cell state should be ignored that is irrelevant. The mechanism of the LSTM cell delves around the concept of gates, which control the information flow. Each gate includes sigmoid and Hyperbolic tangent (tanh) activation functions. The tanh function is used to update the cell state. These gates allow LSTMs to capture and remember important patterns across sequences, reducing gradient problems.

The formula used for obtaining an output from the forget gate, input gate, cell state, and output gate can be seen in Eq. (6-9), respectively. For updating the cell state, Eq. (10) is used whereas Eq.(11) and Eq. (12) deal with a final output of the layer.

3.4.3. Bi-Directional LSTM

BiLSTM networks allow the understanding of temporal dependencies by simultaneously processing input sequences in both forward and backward directions. The flow of information in the BiLSTM is a two-way flow of input sequence [35]. This bidirectional processing enables the model to capture complex relationships within the data, contributing to its effectiveness for predictions. This bidirectional processing captures dependencies that may not be clear when considering only one direction, allowing the model to grasp the complex relationships within the data.

Usually, the activation functions used for these gates are the tanh and ReLu functions. This ensures that the values are constrained in the range of [-1, 1], allowing the model to store and learn long-term dependencies in the sequential data. ReLu function can be considered here for incorporating non-linearity in a model, allowing a better understanding of the data to predict the output.

$$Y_{[s]} = f[H_s, H_s'] \tag{13}$$

Where H_s is the output of the forward layer at interval s, H'_s is the output of the backward layer at interval s, Y_s is the overall output of the model at interval s.

The LSTM cell is a building block of the BiLSTM, having specialized gates that regulate the flow of information at each time step. More details on these gates can be found in the Section 3.4.2. The outputs from both the forward and backward LSTM layers are concatenated, a composite representation of the input sequence that integrates information from both directions which is evident from Eq. (13) This concatenated output is particularly important as it allows a better understanding of the temporal dynamics and further layers allow better representation which helps in enhancing the performance and accuracy.

3.5 Hybrid deep learning models

This section, different Combinations of DL models like CNN-LSTM and CNN-BiLSTM are discussed.

3.5.1. CNN-LSTM omdel

This hybrid model sequentially integrates both architectures that combine the strengths of two neural network models, CNN and LSTM. The model can achieve superior performance compared to using either architecture in isolation and is also robust to variations in input data.

3.5.2. Proposed model

The convolutional operations of a lD-CNN are beneficial for extracting short-range dependencies and spatial hierarchies in the data. This makes 1D-CNNs suitable for tasks where recognizing specific patterns is essential. A BiLSTM and 1D-CNN hybrid model architecture that comprises the potential benefits of both RNN and CNN. The BiLSTM enables the handling of long-range dependencies within sequential data as well as processes input sequences in both forward and backward directions simultaneously. The two directional passes allow the network to understand the sequence by considering past and future information at each time step.

$$Z_{[s]} = f_{conv}[W_{conv}.X_s + B_{conv}] \tag{14}$$

$$H_{[s]}^{fwd} = LSTM_{fwd}\left[Z_s, H_{s-1}^{fwd}\right] \tag{15}$$

$$H_{[s]}^{bwd} = LSTM_{bwd}\left[Z_s, H_{s-1}^{bwd}\right] \tag{16}$$

$$H_{[s]} = \left[H_{[s]}^{fwd}, H_{[s]}^{bwd}\right] \tag{17}$$

The BiLSTM layer is followed by another BiLSTM layer, which may be represented mathematically as;

$$H_{[s]}^{fwd} = LSTM_{fwd}\left[H_s, H_{s-1}^{fwd}\right] \tag{18}$$

$$H_{[s]}^{bwd} = LSTM_{bwd}\left[H_s, H_{s-1}^{bwd}\right] \tag{19}$$

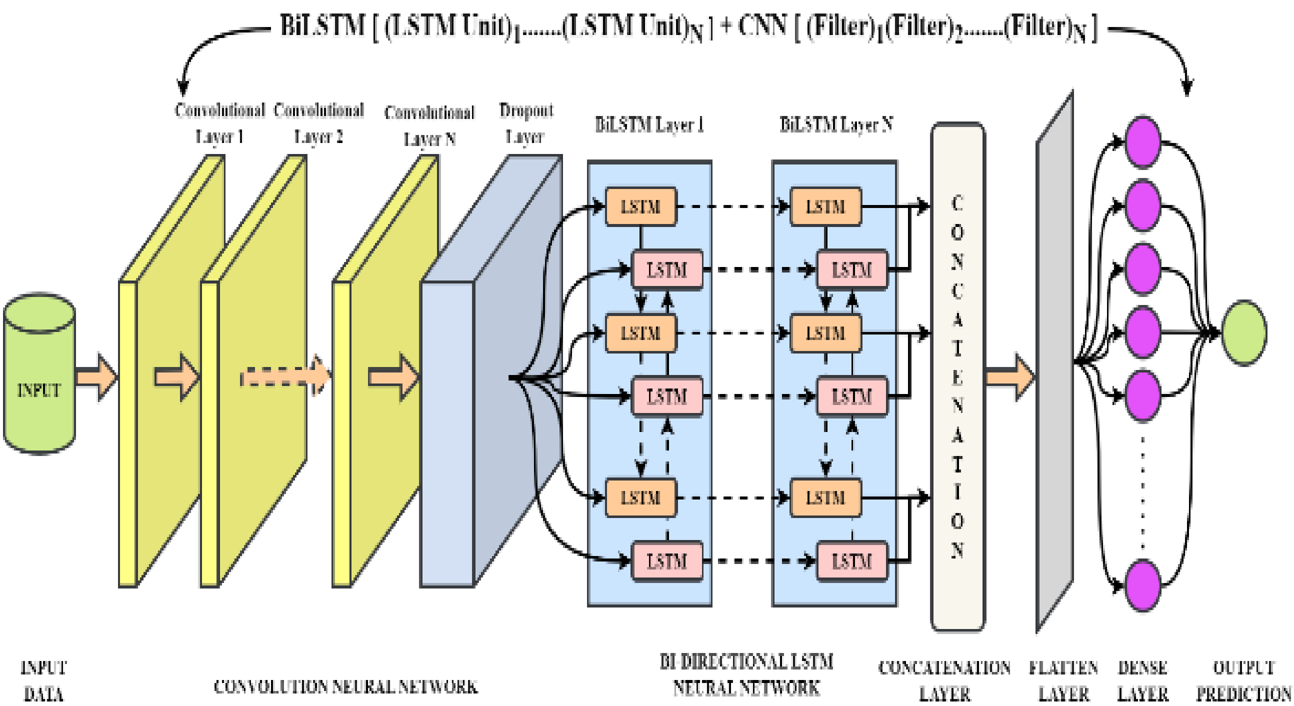

Figure. 3 Architecture of hybrid CNN-BiLSTM model

A Dense layer, which usually takes place before the prediction of the output, is expressed as follows;

$$Y_{[s]} = f^{dense}[W_{dense}.H_s + B_{dense}] \quad (20)$$

Where at interval s, X_s, Z_s is the input, output of convolution layer, f_{conv} is the activation function for convolution layer, f_{dense} is the activation function for dense layer, H_s^{fwd}, H_s^{bwd}, is the hidden states in the forward direction and backward direction, $LSTM_{fwd}$, $LSTM_{bwd}$ is the LSTM operations Eq. (6-12) in the forward direction and backward direction, B_{conv}, B_{dense} is the bias of convolution layer and dense layer.

The proposed hybrid model is trained as shown in Figure. , the weights and biases of both components are simultaneously adjusted. It minimizes a selected objective function also known as the loss function that quantifies the difference between projected and actual outcomes. The effectiveness proposed hybrid model has been demonstrated in further section.

3.6 Performance evaluation parameters

The model's performance is evaluated and compared using several parameters such as Mean Square Error (MSE), RMSE, Coefficient of Determination (R^2 Score), and training time. Maintaining and tuning the hyperparameter is important to obtain better results, which will eventually help better predict the target variables [12]. MSE, RMSE, and R^2 Score can be evidently seen in Eq. (21), Eq. (22), and Eq.(23), respectively. The duration of training time can be influenced by various factors.

$$MSE = \frac{1}{N}\left[Y_i - \widehat{Y}_i\right]^2 \quad (21)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}\left[Y_i - \widehat{Y}_i\right]^2} = \sqrt{MSE} \quad (22)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left[Y_i - \widehat{Y}_i)\right]^2}{\sum_{i=1}^{N}[Y_i - \overline{Y}]^2} \quad (23)$$

$$t_{Total} = \Sigma_{i=1}^{I_{Total}} t_i \quad (24)$$

Where Y_i, $\widehat{Y}_i$, $\overline{Y}$ is the actual, predicted value, and mean of actual value, respectively; N is the total number of samples, t_{Total} is the total training time of the experimentation, I_{Total} is the total number of iterations which are obtained by I_{Total} = [Epoch] x [Time steps per epoch], t_i is a particular iteration time.

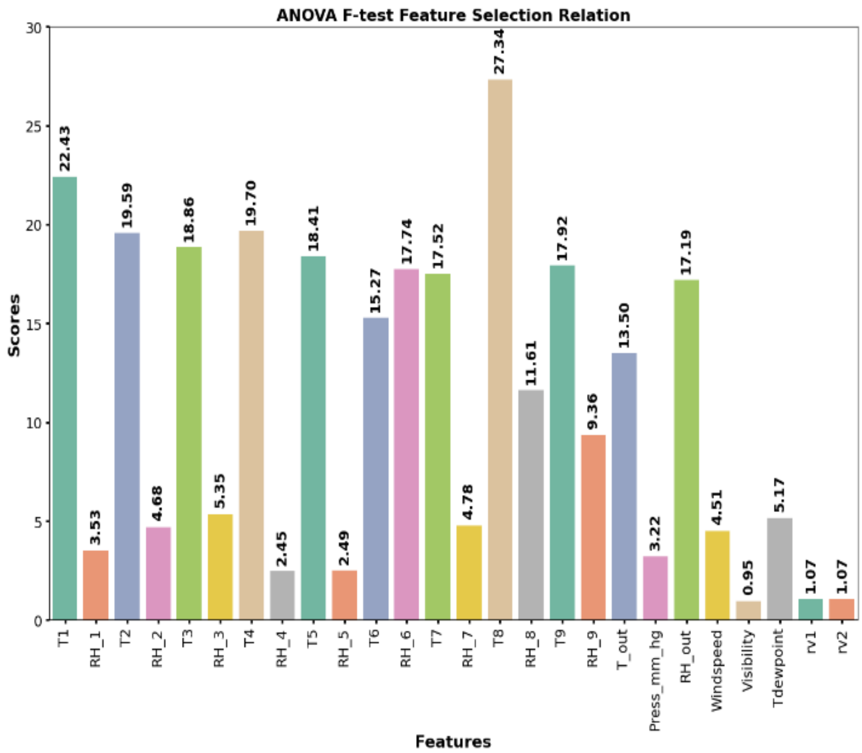

Figure. 4 Feature Importance Graph obtained using ANOVA F-measure technique

Received: February 8, 2024. Revised: March 5, 2024.

4. Results and discussions

4.1 Experimental platform

In this paper, for the experimental purpose Central Processing Unit CPU having processor i5-8250U @ 1.6 GHz with 4 cores and 16 GB RAM is employed for experimentation.

4.2 Feature selection results

By implementing the ANOVA, only the best-contributing features are selected. The ANOVA F-test is used as the scoring function. This scoring tries to maintain the linear relationship between each feature and output variable. A higher F-score suggests that the inclusion of that feature may significantly improve the model's performance in determining the variance in the target variable. For feature selection, the parameter used indicates that all features should be selected based on their scores. The implemented feature selection method statistics can be seen in Figure.. It was viably found that it includes a few features with lower values. These low F-score features need to be dropped. Features like rv_1, rv_2, Visibility, RH_4 and RH_5. It was seen that the random variable was dropped even when different feature selection techniques like PCC [31] and Covariance [32] were implemented. While eliminating it was found that different sets of features are dropped when PCC is applied. The comparison of selected feature on the ML/DL model can be seen in Section 4.3.

4.3 Experimental results and discussion

The model has undergone training for 100 epochs with an early stopping mechanism, where it comprises a patience level of 5 to monitor MSE loss performance. This strategy ensures that the training process halts if no improvement is observed over five consecutive epochs, preventing overfitting. The performance metrics were evaluated up to this epoch value, influencing the overall training time, which is eventually connected with the number of epochs. Various ML/DL models were trained on the dataset for improved predictions. In The application of DL models, including LSTM, CNN, and BiLSTM, is implemented for proper understanding and prediction which can be seen in エラー! ブックマークが自己参照を行っています。. The LSTM enables a long-term understanding of the features whereas the CNN exhibited a short-term understanding of spatial features. However, the LSTM and CNN hybrid model also shows further room for improvement. However, it's essential to note that the computational complexity of these hybrid models increases with the inclusion of filters and units in their layers. As the BiLSTM allows the understanding of long-term features in both forward and backward directions. Furthermore, a combination of BiLSTM and CNN proved to be particularly effective from Table 2.

, the experimentation result on the test dataset can be seen, and it is evident that ML models did not significantly contribute to enhancing the prediction of total household appliance energy consumption. The SVR Model shows poor performance in forecasting. Prediction distribution results for the proposed model compared with the inputted values can be seen in エラー! 参照元が見つかりません。, and the enlarged version for a small subset of test data can be seen in エラー! 参照元が見つかりません。(a).

Table 1. Comparison of predicted and real values.

Model Name	MSE	RMSE	R^2 Score	Total Training Time (Sec)
SVR (Linear kernel)	901.861	30.031	0.8856	14 s
SVR (RBF kernel)	5032.9	70.943	0.362	54 s
LSTM $(50)_1(50)_2(50)_3(50)_4$	422.7344	20.5605	0.9514	462
LSTM $(128)_1(64)_2(32)_3$	9.426	3.07	0.987	484
CNN $(64)_1(32)_2$	9.727	3.1188	0.989	126
CNN $(64)_1$+LSTM $(64)_1(32)_2$	5.849	2.4185	0.992	196
BiLSTM $(128)_1$	24.96	4.9959	0.9915	832
BiLSTM $(64)_1$	4.98	3.8704	0.988	64
BiLSTM $(128)_1(128)_2$	35.567	5.9638	0.985	1296
CNN $(128)_1$ + BiLSTM $(128)_1$ [Proposed]	3.06	1.7493	0.996	920

$$f_{ReLu} = max[0, x] \tag{25}$$

The application of DL models, including LSTM, CNN, and BiLSTM, is implemented for proper understanding and prediction which can be seen in エラー! ブックマークが自己参照を行っています。. The LSTM enables a long-term understanding of the features whereas the CNN exhibited a short-term understanding of spatial features. However, the LSTM and CNN hybrid model also shows further room for improvement. However, it's essential to note that the computational complexity of these hybrid models increases with the inclusion of filters and units in their layers. As the BiLSTM allows the understanding of long-term features in both forward and backward directions. Furthermore, a combination

Table 2 provides details on the fine-tuning of input parameters for BiLSTM and CNN, choosing the best possible model through such adjustments of filters and units.

of BiLSTM and CNN proved to be particularly effective from エラー! ブックマークが自己参照を行っています。.

Extensive testing was conducted on various combinations of BiLSTM and CNN architectures, and the best configuration of the units and layers of BiLTM and CNN was selected

Table 2. Performance evaluation of various hybrid BiLSTM-CNN approaches by tuning parameters with ANOVA

Model Name	**MSE**	**RMSE**	**R^2 Score**	**t_{Total} (Sec)**
CNN $(64)_1$ + BiLSTM $(64)_1(64)_2$	15.936	3.992	0.987	735
CNN$(64)_1$ + BiLSTM $(64)_1$	18	4.242	0.987	256
CNN$(128)_1(64)_2$ + BiLSTM$(64)_1$	6.4715	2.5439	0.991	424
CNN$(64)_1(64)_2$ + BiLSTM$(64)_1(64)_2$	4.1205	2.0299	0.994	656
CNN $(128)_1(64)_2$ + BiLSTM $(128)_1$	6.591	2.5673	0.989	1728
CNN$(128)_1$ + BiLSTM $(128)_1$ [Proposed]	3.06	1.7493	0.996	920

This iterative tuning process is aimed at achieving optimal performance and robust predictive capabilities for the specific characteristics of the dataset. Overall, in the DL models, the ReLu activation function is implemented and that can be seen in Eq. (エラー! 参照元が見つかりません。).

In The application of DL models, including LSTM, CNN, and BiLSTM, is implemented for proper understanding and prediction which can be seen in エラー! ブックマークが自己参照を行っています。. The LSTM enables a long-term understanding of the features whereas the CNN

Table 2, in parenthesis, the input parameters like filters and LSTM units consist. The LSTM, CNN, and LSTM-CNN models best-obtained result after hyperparameter tuning is only recorded. It is also seen that by implementing the feature selection technique there is an enhancement in its model's performance. The complexity is also reduced due to this. The MSE loss of the best-performed model can be seen in エラー! 参照元が見つかりません。(b).

4.4 Comparative analysis

Our approach has performed far better than theirs by comparing with other's work on this dataset which can be seen in Section 2. In [31], different ML/DL

exhibited a short-term understanding of spatial features. However, the LSTM and CNN hybrid model also shows further room for improvement. However, it's essential to note that the computational complexity of these hybrid models increases with the inclusion of filters and units in their layers. As the BiLSTM allows the understanding of long-term features in both forward and backward directions. Furthermore, a combination of BiLSTM and CNN proved to be particularly effective from エラー! ブックマークが自己参照を行っています。.

and

models are experimented with that show 65.54 for SVM, 65.64 for Random Forest (RF), 21.36 for LSTM, 64.99 for K-Nearest Neighbour (KNN), and 59.81 for the Extreme Random Forest (ERF) model. LSTM outperforms other models in fitting nature and understanding data distribution, with a 0.97 R^2 score indicating room for improvement in prediction efficacy and better understanding. These other ML models are unable to understand the behaviour of time-series data.

The min-max scaling shows depending on the characteristics of the data, min-max scaling to [0, 1] may be more beneficial for time series data due to the conservation of original values, benefits for

Received: February 8, 2024. Revised: March 5, 2024.

normalization, and such testing is performed using the LSTM model. A covariance-based feature selection also implement that yields the best result having aggregate RMSE of 5. Research in [30] showcased the hybrid LSTM and BiLSTM model unable to understand the complex spatial features which is done by using the CNN.

In [37] models like RF, SVM, and ANN are implemented without feature selection that yielded result of 21.46, 67.71, 96.61 RMSE and 0.239, 0.764, 0.974 R^2 Score.

5. Conclusion

Managing the proper electric energy is necessary for maintaining the stability of the smart grid, and it also contributes towards sustainability. For energy management systems, forecasting of energy consumption plays a crucial role. Analysis of consumption trends helps individuals to contribute to global energy conservation. This paper evaluates prediction parameters using different ML/DL models on the AEC dataset. The ANOVA F-measure technique has been utilized for feature selection to eliminate redundant features. It helps remove features like visibility, random variables, and values having a score of less than 3. The F-score value ranges from 1.07 to 22.42. The best hybrid CNN-BiLSTM model was chosen through hyper-parameter adjustment.

This study performs a comparative analysis of standalone and hybrid DL/ML models. The obtained parameters such as MSE, RMSE, R^2 score, and t_{total} of the proposed model are 3.06, 1.7493, 0.996 and 920 Sec, respectively. For CNN-LSTM, these values are 5.849, 2.4185, 0.992, and 196 Sec, respectively. The result shows a significant improvement while utilizing the hybrid model, which can predict the total energy consumption output accurately. In the future, this can be used to enhance the predictive ability of the energy consumption of appliances, which will help in energy management systems and smart grids, eventually contributing to sustainability by lowering the stress on environmental resources.

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

Conceptualization, methodology, software, writing—original draft preparation; Khush Attarde. Resources, data curation, proofreading, supervision, funding acquisition; Javed Sayyad.

References

[1] R. L. Hu, R. Skorupski, R. Entriken, and Y. Ye, "A mathematical programming formulation for optimal load shifting of electricity demand for the smart grid", *IEEE Trans Big Data*, Vol. 6, No. 4, pp. 638–651, 2016.

[2] M. Afzalan and F. Jazizadeh, "Residential loads flexibility potential for demand response using energy consumption patterns and user segments", *Appl Energy*, Vol. 254, pp. 113693, 2019.

[3] H. W. Rotib, M. B. Nappu, Z. Tahir, A. Arief, and M. Y. A. Shiddiq, "Electric load forecasting for Internet of Things smart home using hybrid PCA and ARIMA algorithm", *International Journal of Electrical and Electronic Engineering & Telecommunications*, Vol. 10, No. 6, pp. 369–376, 2021.

[4] E. Kamel, S. Sheikh, and X. Huang, "Data-driven predictive models for residential building energy use based on the segregation of heating and cooling days", *Energy*, Vol. 206, pp. 118045, 2020.

[5] H. Ziekow, C. Goebel, J. Strüker, and H.-A. Jacobsen, "The potential of smart home sensors in forecasting household electricity demand," In: *Proc. of 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pp. 229–234, 2013.

[6] H. Pombeiro, R. Santos, P. Carreira, C. Silva, and J. M. C. Sousa, "Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks", *Energy Build*, Vol. 146, pp. 141–151, 2017.

[7] Z. Pezeshki and S. M. Mazinani, "Comparison of artificial neural networks, fuzzy logic and neuro fuzzy for predicting optimization of building thermal consumption: a survey", *Artif Intell Rev*, Vol. 52, No. 1, pp. 495–525, 2019.

[8] H. Zhang, H. Feng, K. Hewage, and M. Arashpour, "Artificial neural network for predicting building energy performance: a surrogate energy retrofits decision support framework", *Buildings*, Vol. 12, No. 6, pp. 829, 2022.

[9] F. Ascione, N. Bianco, C. De Stasio, G. M. Mauro, and G. P. Vanoli, "Artificial neural networks to predict energy performance and retrofit scenarios for any member of a building category: A novel approach", *Energy*, Vol. 118, pp. 999–1017, 2017.

[10] A. Moradzadeh, A. Mansour-Saatloo, B. Mohammadi-Ivatloo, and A. Anvari-Moghaddam, "Performance evaluation of two machine learning techniques in heating and cooling loads forecasting of residential buildings", *Applied Sciences*, Vol. 10, No. 11, pp. 3829, 2020.

[11] A. Tsanas and A. Xifara, "Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools", *Energy Build*, Vol. 49, pp. 560–567, 2012.

[12] S. Paudel *et al.*, "A relevant data selection method for energy consumption prediction of low energy building based on support vector machine", *Energy Build*, Vol. 138, pp. 240–256, 2017.

[13] M. Elsaraiti, G. Ali, H. Musbah, A. Merabet, and T. Little, "Time series analysis of electricity consumption forecasting using ARIMA model", In: *Proc. of 2021 IEEE Green technologies conference (GreenTech)*, pp. 259–262, 2021.

[14] P. Chujai, N. Kerdprasop, and K. Kerdprasop, "Time series analysis of household electric consumption with ARIMA and ARMA models", In: *Proc. of the international multiconference of engineers and computer scientists*, pp. 295–300, 2013.

[15] H. Kaur and S. Ahuja, "Time series analysis and prediction of electricity consumption of health care institution using ARIMA model", In: *Proc. of Sixth International Conference on Soft Computing for Problem Solving: SocProS 2016*, Vol. 2, pp. 347–358, 2017.

[16] H. Shi, M. Xu, and R. Li, "Deep learning for household load forecasting—A novel pooling deep RNN", *IEEE Trans Smart Grid*, Vol. 9, No. 5, pp. 5271–5280, 2017.

[17] M. Alhussein, K. Aurangzeb, and S. I. Haider, "Hybrid CNN-LSTM model for short-term individual household load forecasting", *IEEE Access*, Vol. 8, pp. 180544–180557, 2020.

[18] M. Sajjad *et al.*, "A novel CNN-GRU-based hybrid approach for short-term residential load forecasting", *IEEE Access*, Vol. 8, pp. 143759–143768, 2020.

[19] L. Jiang, X. Wang, W. Li, L. Wang, X. Yin, and L. Jia, "Hybrid multitask multi-information fusion deep learning for household short-term load forecasting", *IEEE Trans Smart Grid*, Vol. 12, No. 6, pp. 5362–5372, 2021.

[20] L. Sehovac and K. Grolinger, "Deep learning for load forecasting: Sequence to sequence recurrent neural networks with attention", *IEEE Access*, Vol. 8, pp. 36411–36426, 2020.

[21] W. Kong, Z. Y. Dong, Y. Jia, D. J. Hill, Y. Xu, and Y. Zhang, "Short-term residential load forecasting based on LSTM recurrent neural network", *IEEE Trans Smart Grid*, Vol. 10, No. 1, pp. 841–851, 2017.

[22] G. F. Martin Nascimento, F. Wurtz, P. Kuo-Peng, B. Delinchant, N. Jhoe Batistela, and T. Laranjeira, "GreEn-ER–Electricity consumption data of a tertiary building", *Frontiers in Sustainable Cities*, Vol. 5, pp. 1043657, 2023.

[23] G. Hebrail and A. Berard, "Individual household electric power consumption data set", *UCI Machine Learning Repository*, 2012.

[24] L. DataStore, "Smartmeter energy consumption data in london households", 2018, https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

[25] A. S. N. Uttama Nambi, A. Reyes Lua, and V. R. Prasad, "Loced: Location-aware energy disaggregation framework", In: *Proc. of the 2nd ACM International Conference on Embedded Systems for Energy-efficient Built Environments*, pp. 45–54, 2015.

[26] M. Wenninger, A. Maier, and J. Schmidt, "DEDDIAG, a domestic electricity demand dataset of individual appliances in Germany", *Sci Data*, Vol. 8, No. 1, pp. 176, 2021.

[27] L. Candanedo, "Appliances energy prediction", 2017, https://archive.ics.uci.edu/dataset/374/appliances+energy+prediction.

[28] T. Le, M. T. Vo, B. Vo, E. Hwang, S. Rho, and S. W. Baik, "Improving electric energy consumption prediction using CNN and Bi-LSTM", *Applied Sciences*, Vol. 9, No. 20, pp. 4237, 2019.

[29] F. U. M. Ullah, A. Ullah, I. U. Haq, S. Rho, and S. W. Baik, "Short-term prediction of residential power energy consumption via CNN and multi-layer bi-directional LSTM networks", *IEEE Access*, Vol. 8, pp. 123369–123380, 2019.

[30] D. Syed, H. Abu-Rub, A. Ghrayeb, and S. S. Refaat, "Household-level energy forecasting in smart buildings using a novel hybrid deep learning model", *IEEE Access*, Vol. 9, pp. 33498–33511, 2021.

[31] L. Xiang, T. Xie, and W. Xie, "Prediction model of household appliance energy consumption based on machine learning", *Journal of Physics: Conference Series*, pp. 12064, 2020.

[32] A. Mehmood, K.-T. Lee, and D.-H. Kim, "Energy Prediction and Optimization for Smart Homes with Weather Metric-Weight

Coefficients", *Sensors*, Vol. 23, No. 7, pp. 3640, 2023.
[33] C. Bartusch, M. Odlare, F. Wallin, and L. Wester, "Exploring variance in residential electricity consumption: Household features and building properties", *Appl Energy*, Vol. 92, pp. 637–643, 2012.
[34] J. Gao *et al.*, "An Exploration into the Fault Diagnosis of Analog Circuits Using Enhanced Golden Eagle Optimized 1D-Convolutional Neural Network (CNN) with a Time-Frequency Domain Input and Attention Mechanism", *Sensors*, Vol. 24, No. 2, pp. 390, 2024.
[35] X. Song *et al.*, "Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model", *J Pet Sci Eng*, Vol. 186, pp. 106682, 2020.
[36] H. Zhen, D. Niu, M. Yu, K. Wang, Y. Liang, and X. Xu, "A hybrid deep learning model and comparison for wind power forecasting considering temporal-spatial feature extraction", *Sustainability*, Vol. 12, No. 22, pp. 9490, 2020.
[37] V. Vakharia, S. Vaishnani, and H. Thakker, "Appliances Energy Prediction Using Random Forest Classifier", In: *Recent Advances in Mechanical Engineering: Select Proceedings of ITME 2019*, pp. 405–410, 2021.