Highway Self-Attention Dilated Casual Convolutional Neural Network Based Short Term Load Forecasting in Micro Grid

[1]Shreenidhi H S and [2]Narayana Swamy Ramaiah
[1,2]Faculty of Engineering and Technology, Jain (Deemed-to-be University), Bengaluru, India.
[1]shreenidhihs.14@gmail.com, [2]narayanaswamy.ramaiah@gmail.com

Correspondence should be addressed to Shreenidhi H S : shreenidhihs.14@gmail.com.

Article Info
Journal of Machine and Comp uting (http://anapub.co.ke/journals/jmc/jmc.html)
Doi: https://doi.org/10.53759/7669/jmc202303033
Received 04 February2023; Revised from 20 May 2023; Accepted 20 June 2023.
Available online 05 October 2023.

Abstract – Forecasting the electricity load is crucial for power system planning and energy management. Since the season of the year, weather, weekdays, and holidays are the key aspects that have an effect on the load consumption, it is difficult to anticipate the future demands. Therefore, we proposed a weather-based short-term load forecasting framework in this paper. First, the missing data is filled, and data normalisation is performed in the pre-processing step. Normalization accelerates convergence and improves network training efficiency by preventing gradient explosion during the training phase. Then the weather, PV, and load features are extracted and fed into the proposed Highway Self-Attention Dilated Casual Convolutional Neural Network (HSAD-CNN) forecasting model. The dilated casual convolutions increase the receptive field without significantly raising computing costs. The multi-head self-attention mechanism (MHSA) gives importance to the most significant time steps for a more accurate forecast. The highway skip network (HS-Net) uses shortcut paths and skip connections to improve the information flow. This speed up the network convergence and prevents feature reuse, vanishing gradients, and negative learning problems. The performance of the HSAD-CNN forecasting technique is evaluated and compared to existing techniques under different day types and seasonal changes. The outcomes indicate that the HSAD-CNN forecasting model has low Mean Absolute Error (MAE), Mean Squared Error (MSE), Mean Absolute Percentage Error (MAPE), and a high R^2.

Keywords – Load Forecasting, Micro Grid, Neural Network, Attention, Energy Management.

I. INTRODUCTION

Weather change and worldwide warming are predictable to result in a rise in the amount of power needed for heating and cooling in the winter and summer [1]. Extreme weather temperature swings would likely cause the peak demand for electricity to rise and decrease, which causes strain on the power grid [2]. Therefore, Renewable Energy sources (RESs) are being incorporated into current power networks in limit the rise in worldwide electrical consumption's effect on greenhouse gas emissions [3]. However, due to intermittent power supplies, such integration presents both economic and technical hurdles [4, 5]. A Microgrid (MG) is a compact grid that may be used to integrate intermittent and inconsistent renewable energy sources into existing distribution systems [6]. It is a collection of geographically dispersed generating sources, regulated loads, and energy storage technology that works together to create a self-sufficient energy system and is extremely economical due to its structure and small scale [7, 8]. In the event of a power shortage or excess exchange [9, 10]. The scientific community has increasingly focused its attention on energy management and load forecasting. However, correct analysis and modelling of energy management and MG operation are difficult due to various internal and external uncertainties in the system [11].

Some of the drawbacks of MG include the high unpredictability of RES, their uncertainties, the balancing of supply and demand, and MG's scalability [12]. High starting costs in terms of money, a lack of suitable regulatory and protection requirements, privacy protection concerns, and control issues are further concerns [13]. Due to these factors, research into energy management systems (EMS) for microgrids has been more concentrated. Due to fluctuating power supply and demand, energy management of a MG becomes more difficult than energy management of traditional power systems with the incorporation of non-dispatchable generators such as wind and solar [14].

As a result, for the system to operate smoothly, the control methods for energy management that were employed in previous decades need to be replaced with new control paradigms and control approaches. For energy production, energy dispatch, and the efficient running of the grid, accurate energy demand forecasting is necessary. Additionally, it can encourage the best possible utilization of renewable resources and lessen the MG's loss of primary energy [15].

A key responsibility of the MG energy management system is short-term load forecasting, especially when various renewable energy sources are connected to the MG. In order to sustain continuous network performance and boost economic returns, short-term forecasting may also be seen as a crucial tool for MG operators. Since solar irradiation and PV production power are correlated, the peak solar irradiation on cloudy days is significantly lower than that on sunny days.

To address this issue, machine learning-based solutions such as long short-term memory (LSTM) [16], CNN [17], fuzzy logic framework, Support Vector Machines [18], and hybrid techniques [19] have been developed and widely employed in short-term load forecasting (STLF). However, the primary causes of unpredictability on the demand side are primarily due to unpredictable loads, which are difficult to anticipate because they depend on several variables like season, weather, weekday, and holidays. Also, a significant source of uncertainty is caused by the distributed PV panels used for renewable energy production. Increased RES adoption increases supply-side uncertainty because of its inherent instability and unpredictable electricity generation. The intermittent PV system output produces fluctuations in the MG's total power that lead to an imbalance between production and consumption. Therefore, it is important to estimate the amount of energy to be delivered for the economic and stable operation of the MG. To manage these kinds of uncertainties, accurate hourly or day-ahead prediction of the load is essential. This motivates us to develop a weather-based short-term load forecasting technique for micro-grid energy management.

The contributions of this research work are as follows:

- Normalization enhances the efficiency of network training by preventing gradient explosions throughout the training phase.
- Forecasting the PV production and load consumption based on weather, day, and load features increased the reliability and power balance in the MG.
- Stacking dilated casual convolutions in the proposed framework reduces the number of learnable parameters and increases the receptive field without significantly raising computing costs.
- The MHSA performs several linear transformations to give importance to the most significant time steps for a more accurate forecast.
- The proposed highway skip network (HS-Net) uses shortcut paths and skip connections to improve the information flow. This speeds up the network convergence and prevents feature reuse, vanishing gradients, and negative learning problems.
- The proposed HSAD-CNN forecasting model is evaluated under various seasonal and day types (working, holiday) and the results are compared with the existing techniques based on metrics such as MAE, MSE, MAPE, and R^2.

II. RELATED WORKS

Arcos-Aviles et al. [20] developed an EMS for a grid-connected electro-thermal MG based on fuzzy logic control (FLC). The EMS is composed of two FLC blocks, each with 25 rules. The EMS predicts the MG's behaviour for the subsequent 12 hours by forecasting the thermal and electrical power balance between consumption and generation at an interval of 15 minutes. Because of the ease with which off-line trained fuzzy controllers can be used, the suggested technique has a low level of computational complexity and may be implemented on low-cost digital platforms.

Al-Gabalawy [21] introduced two energy management algorithms: reinforcement learning (RL) and The MPCLC has improved system dependability. But the RL's utility function can be fine-tuned to produce an improved result.

In order to provide real-time energy dispatch during operational uncertainty, Dong et al. [22] developed an adaptive, optimal fuzzy logic-based EMS. Through a unique meta-heuristic optimization approach, the solution chooses the best fuzzy inference system based on the forecasted data over a predetermined period. This method has high robustness and adaptability. This system perform real-time dispatch and adaptively manage the energy scheduling across a time horizon with much less computational complexity.

In order to address different characteristics of energy consumption and consumer STLF, Muzumdar [23] suggested an accurate and robust model that makes use of methods like LSTM, Support Vector Regression (SVM), and Random Forest (RF) as base predictors. It allocates weights to each predictor dynamically based on the effectiveness of the prediction in order to make the ultimate judgement regarding the forecasting outcome of these predictors. For load forecasting, data samples such as day of the week, hour, month, year, season, etc. are fed into each of the learning models. An ensemble approach is then employed to reach the final decision. LSTM aids in resolving the issue of increasing inconsistency in load. SVR performs well for the aggregated load, whereas RF gives superior stability and resilience against outliers. However, problems like vanishing gradients, feature reuse, and negative learning arise due to the insufficient information flow between the network layers.

An optimum scheduling model for isolated MG is proposed by Li et al. [24] in order to lessen the detrimental effects of renewable energy output and load uncertainties. Initially, a prioritized experience replays automated reinforcement learning (PER-AutoRL) model automatically chooses the best model architecture and optimise hyper-parameters depending on the input data.

Zang et al. [25] introduced a hybrid deep learning model-based day-ahead residential load forecasting approach. LSTM and self-attention are combined to create a hybrid model. Data from each user is pre-processed in two stages, employing multi-source input dimension reconstruction and decomposition. The information from the target and related users is then combined via pooling, in descending order. This method reduces over-fitting by lowering the relative disparities between the model and the data and improving data volume and diversity. However, environmental factors such as humidity, temperature, and rain are not considered in this method, which may degrade the performance.

III. PROPOSED LOAD FORECASTING FRAMEWORK

The proposed HSAD-CNN forecasting framework, as illustrated in **Fig 1.**

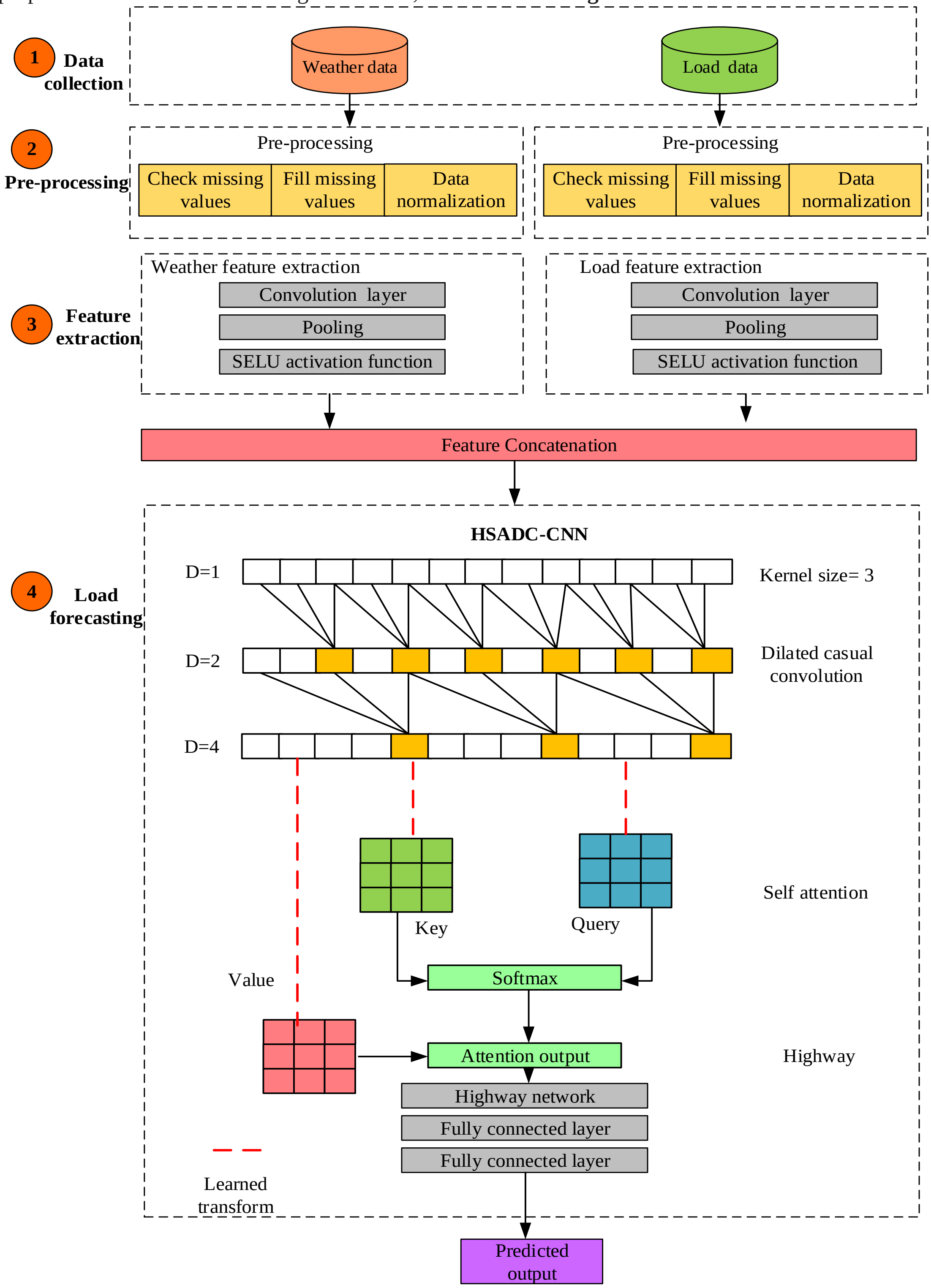


Fig 1. Architecture of the Proposed Load Forecasting Framework

Data Collection

The MiRIS microgrid at the global headquarters of the John Cockerill Group in Seraing, Belgium, is used for the simulation. It is made up of a non-sheddable load, energy storage devices, and PVs. We utilise weather data and data from on-site monitoring for historical load and PV data available at Kaggle platform. It consists of historical weather, load, and PV data with 15 min resolution.

Data Pre-Processing

The weather data and load data are initially gathered. A limited amount of data will inevitable be lost during time series data collection. The steadiness of the time series data will be affected if the missing value in the data is not handled. Therefore, the missing data for the load demand are filled by using the load figure obtained at the same period last week. Moreover, due to the wide numerical range and multiple units of variables, normalizing the input sequences is essential. When the load data is supplied directly into the network during the training process, the computational time increases and the forecast accuracy declines. Using normalised data for training speeds up convergence and enhance model prediction accuracy. Data normalisation entails scaling and mapping data to the range [0, 1]. Normalization can enhance the efficiency of network training by preventing gradient explosions throughout the training phase.
The formula for data normalization is given in Eqn. (1)

$$\hat{z} = \frac{z - z_{mean}}{zmin_{max}} \tag{1}$$

where z signifies the data before normalization, z_{mean} , z_{min}, and z_{max} represents the mean, minimum, and maximum values of the data.

Feature Extraction

The electric load will vary greatly during weak days and holidays due to the major fluctuations in human activity. For example, in the United States, the electric load appears to be lower during Christmas and around Independence Day than during non-holidays. This could be due to many businesses being closed during the holidays. During the Spring Festival in China, the industrial electric load will fall considerably due to the shutdown of large-scale production operations. In contrast, the residential electric load will experience the highest load of the year.

Holidays have atypical hourly load patterns when compared to usual days. Therefore, calendar and weather data have an impact on the load profile at various time resolutions, such as hourly, daily, and yearly. In the feature extraction step, the load, PV, and weather features such as relative humidity, precipitation, solar irradiance, temperature above ground, day type, wind speed, PV output, etc. are extracted by convolution and pooling operations. The Multi time scale CNN (MTS-CNN) extract the local features at multiple time scales. The MTS-CNN's function is to extract features from time series data of load and PV production in similar daily samples at different time scales. Convolution kernels are an excellent neural network model that CNN employs to automatically extract crucial information. During the convolution process, the same convolution kernel acquires a class of related features. Its mathematical representation is defined as,

$$J_l = f\left(J_{l-1} \otimes WT_l + b_l\right) \tag{2}$$

J_l denotes the input of layer l, J_{l-1} denotes the $l-1$ layer's output, WT_l and b_l denotes the weight and bias of layer l, f denotes the Scaled Exponential Linear Unit (SELU) activation function, which is another updated version of the Rectified Linear Unit (ReLU) and has a self-normalizing property.
In order to reduce computation complexity and prevent over-fitting, the pooling layer is used. It comes after the convolution processes, uses data down-sampling to down-sample a large matrix into a small one. The mathematical model for the pooling layer is given in Eqn. (3).

$$J_l = down\left(J_{l-1}\right) \tag{3}$$

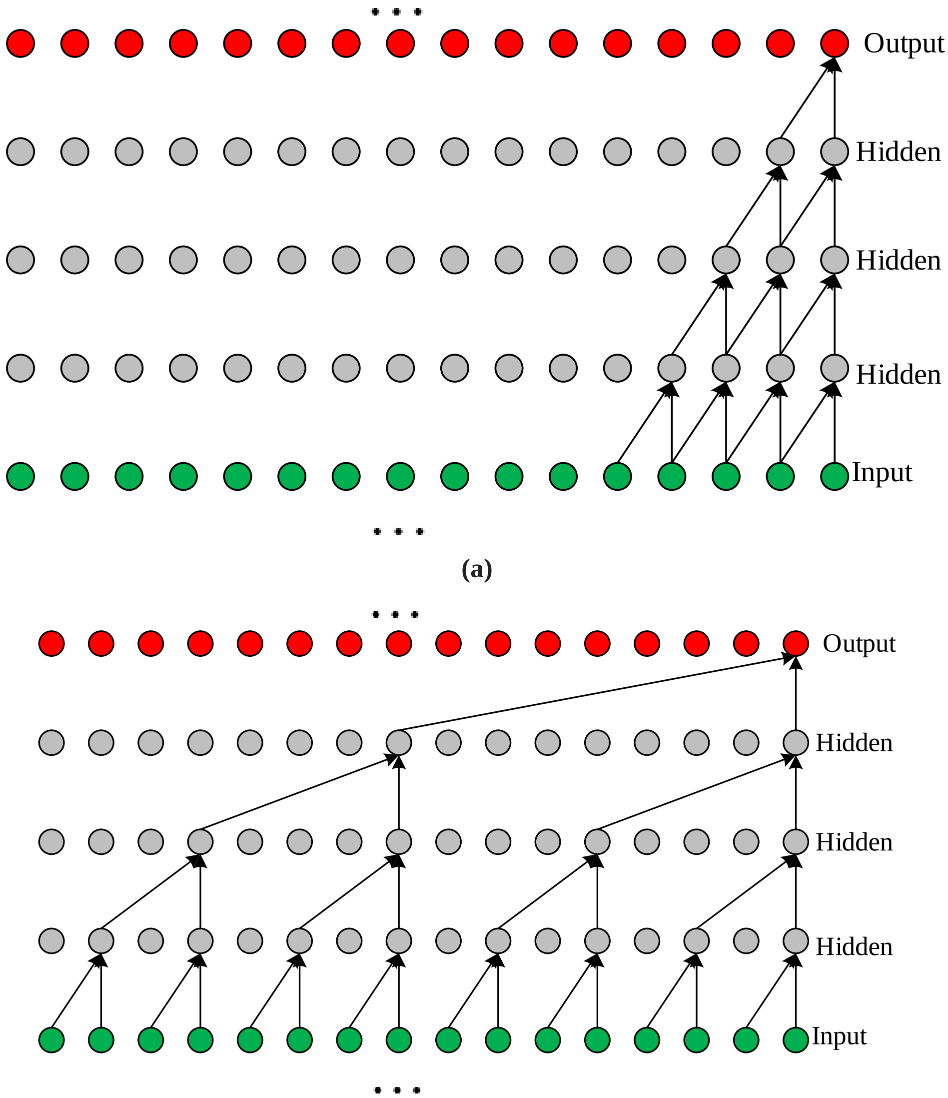
where $down(\;)$ denotes the pooling function and J_{l-1} denotes the features obtained after pooling. Finally, the extracted features are fed into the proposed HSAD-CNN model to forecast the load.

Load Forecasting

The HSAD-CNN forecasting model is made up of three units: dilated casual convolution, multi-head self-attention, and the highway skip network.

Dilated Casual Convolution

The HSAD-CNN receives the output of the MTS-CNN model and captures global temporal dependencies at various time scales and forecast the load and PV production. To forecast the load, a highway self-attention casual convolutional neural network (HSAD-CNN) is proposed in this paper. While attempting to forecasts the future of a time series using its past, we must make sure that the output is independent of future samples. The ordering of sequential input patterns is preserved by causal convolutions, preventing data from the future from leaking into the past. However, to increase the receptive field, causal convolutions need numerous hidden layers or very large filters. Dilated convolutions are integrated into the proposed technique to increase the receptive field without significantly raising the computing cost.



(b)

Fig 2. Illustration convolution (a) Casual Convolution (b) Dilated Casual Convolution

In a dilated convolution, as illustrated in **Fig 2,** the convolution filter is applied over a greater area by bypassing certain input values. Dilated convolutions include a new parameter called the dilation factor, which sets the distance between kernel values. Fig. 3 depicts a conceptual representation of dilated convolution. For example, a 3×3 kernel with a dilation factor 2 does the same operation as a 5×5 kernel while employing just nine parameters. This provides a broader field of view while maintaining the same computational cost.

For an input sequence $z \in R^k$ of one dimension with a filter $h:\{0,\ldots,l-1\}\rightarrow R$, the dilation convolution is defined in Eqn. (4).

$$G(q)=\sum_{p=0}^{l-1} h(p)\cdot z_{q-f\cdot p} \tag{4}$$

where f denotes the dilation factor, q denotes the input sequence element, l denotes the kernel size, $q-f\cdot p$ denotes the past time steps.

When the model depth is increased i.e., $f = 2^n$ at the layer n, the dilation factor f increases exponentially. At each layer $n = 1,\ldots,N$, the dilation factors f is increased by a factor of 2 where N denotes the number of dilated convolution layers. The dilatation factor is given in Eqn. (5).

$$f \in \left[2^0, 2^1, \ldots, 2^{N-1}\right] \quad (5)$$

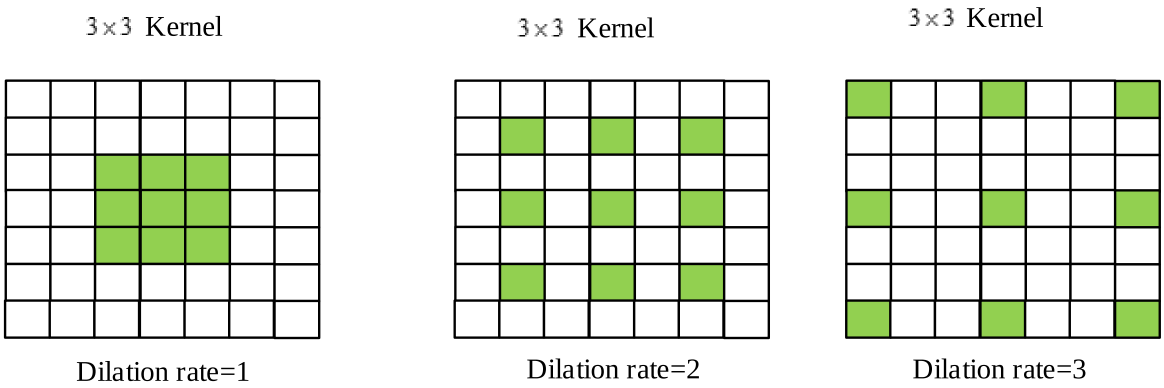


Fig 3. Convolution With Different Dilation Rates

Multi Head Self Attention Layer

Self-attention is employed to allow the network to emphasis on relevant and crucial time periods rather than unimportant time periods. For example, when forecasting the load demand on Sunday, it is important to give importance to the similar holidays compared to working days. Therefore, for each time step, self-attention determines relative weights by taking into account its similarity to all previous time steps. As a result, the attention mechanism gives importance to the most significant time steps for a more accurate forecast.

The traditional attention mechanism can only learn about attention data at a single level. The MHSA, as shown in Fig. 4 performs several linear transformations and learns the attention representation of the text on the input feature matrix to get more detailed information. It is an arrangement of multiple self-attention mechanisms. Each self-attention mechanism (linear or dense layer) has a key matrix M_K, value matrix M_V, and a query matrix M_K. The global feature vector matric H serves as the initial value for M_V, M_K, and M_Q which is depicted in Eqn. (6).

$$M_V = M_K = M_Q = H \quad (6)$$

Scaled Dot-product Attention (SDA) is the central idea of the attention process. SDA calculates the similarities by computing the dot product of M_Q & M_K, then dividing by $\sqrt{d_{M_Q}}$ so that the result is not too huge. Here, d_{M_Q} denotes the size of matrix M_Q. The output is then multiplied by the matrix M_V and normalised using the softmax function to get the attention. The operation of SDA is performed using Eqn. (7).

$$SDA(M_V, M_K, M_Q) = soft\,max\left(\frac{M_Q M_K{}^T}{\sqrt{d_{M_K}}}\right) M_V \quad (7)$$

The multi-head attention mechanism works by using various parameters $W_j^{M_V}$, $W_j^{M_K}$, and $W_j^{M_Q}$ to conduct linear transformations on the matrices M_V, M_K, and M_Q, and then feed the linear transformation output into the SDA mechanism. As indicated in Eqn. (8), $head_j$ represents the computation result.

$$head_j = SDA\left(M_V W_j^{M_V}, M_K W_j^{M_K}, M_Q W_j^{M_Q}\right) \quad (8)$$

Here, the heads give attention to the most important features of weather, PV, and load. The final results of $head_1$ to $head_g$ are then concatenated to build a matrix. It is then multiplied by the parameter W to get the attention mechanism result using Eqn. (9).

$$\begin{aligned} head &= MultiHead\left(M_V, M_K, M_Q\right) \\ &= concat\left(head_1, \ldots, head_g\right)W \end{aligned} \tag{9}$$

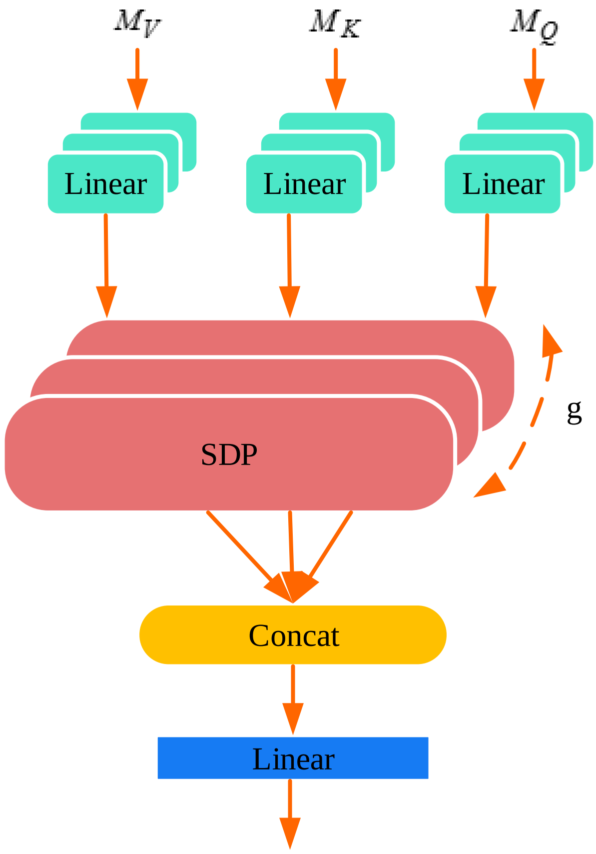


Fig 4. Structure of Multi-Head Attention

Highway Skip Network (HS-Net)

Because of inadequate information flow between network layers, deep architectures suffer from feature reuse, vanishing gradient, and negative learning issues. Highway networks, as illustrated in Fig. 5 reformulate deep architectural design by offering more shortcut paths. These routes use a gating mechanism to govern the flow of information & directly connect the preceding levels with future layers.

Instead of one transformation layer I in typical neural networks, a highway network is made up of blocks that contain two additional transformation layers, as given in Eqn. (10).

$$z = y \cdot Ca\left(y, W_{Ca}\right) + I\left(y, W_I\right) \cdot Tr\left(y, W_{Tr}\right) \tag{10}$$

where $Ca\left(y, W_C\right)$ and $Tr\left(y, W_T\right)$ represents the carry and transform gate transformation respectively. When $Ca = 1 - Tr$, Eqn. (19) can be rewritten as,

$$z = y \cdot \left(1 - Tr\left(y, W_{Ca}\right)\right) + I\left(y, W_I\right) \cdot Tr\left(y, W_{Tr}\right) \tag{11}$$

Convolution layers with ReLU activation, a batch normalisation layer, and a convolution layer with ReLU activation make up the I transformation of the highway block. On the other hand, there is only one convolution layer in the Tr transformation of the highway block.

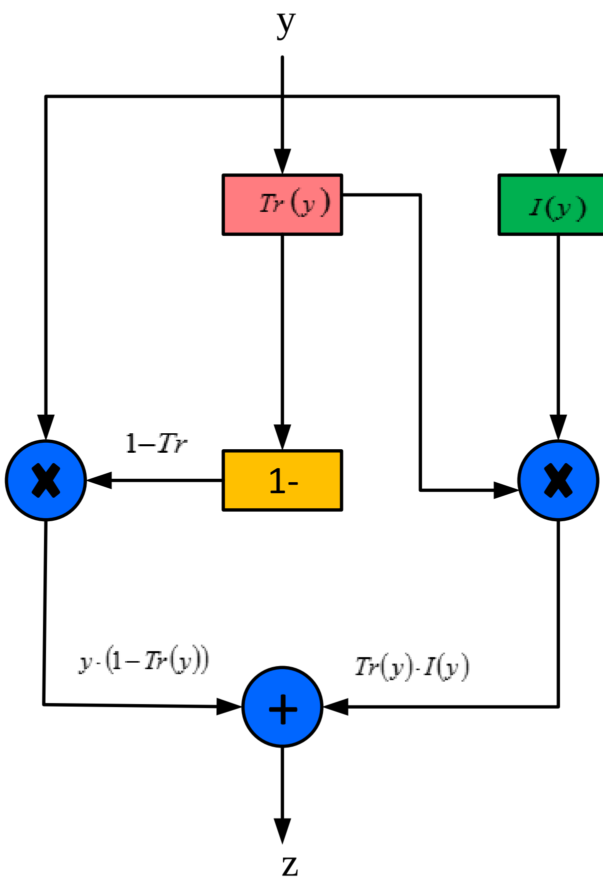


Fig 5. Highway Network Module

Each highway block has skip connections, as shown in Fig. 6, where it skips some of the layers to send its output directly to the last node or successive layers of the model. The results are concatenated with the outputs from all the previous levels and the final output is produced. With time series data, using skip connections enables previous layer features to be taken into account, making the network aware of all data patterns and periodicity, thus improving network convergence. The proposed forecasting network consists of six highway blocks, followed by two fully connected layers to forecast the load and PV production.

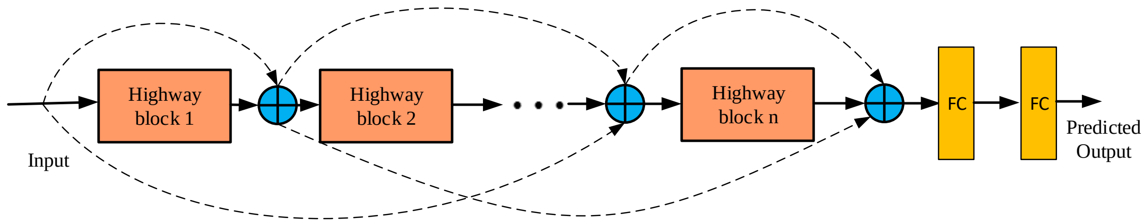


Fig 6. Skip Connection

IV. SIMULATION RESULTS

The proposed forecasting model is simulated on Python platform, and the performance is tested and compared with the existing methods in terms of MSE, MAE, MAPE, and R^2. The proposed model is trained using the Adam optimizer and the MAE loss function.

A set of the parameters given below is utilized in a series of iterative studies to determine for the finest parameters.

- Batch size: 32, 64, 128, and 256
- Epoch: 1-100
- Dropout: 20%, 30%, 40%, and 50%
- Learning rate: 0.0001-0.01

Based on a series of trial-and-error tests, we discovered that the ideal hyper-parameters for the models to converge are provided in **Table 1**.

Table 1. Hyper-Parameter Setting

Parameter	Value
Batch Size	64
Epochs	50
Dropout Rate	20%
Learning Rate	0.001

Dataset Description

This proposed method is simulated on Python platform based on the MiRIS MG at the global headquarters of the John Cockerill Group in Belgium. The historical load, weather, and PV data are available on the Kaggle platform . It includes a 15 min resolution weather forecast that includes air temperature, solar irradiance, etc., generated by the Laboratory of Climatology of the University of Liège. The remaining 20% of the data is utilised for testing, and other 80% is used for training.

Evaluation Metrics

The performance metrics used for evaluating the proposed method are MAE, MSE, MAPE, and R^2. The obtained results are compared with existing methods[32-34] such as LSTM+RF+SVR [24], XGB-LGBM-MLP [26], TCN-GRU [28], CNN+LSTM [28], Hybrid GA-CNN [29], VMD-VAE-LSTM [30], and edRVFL [30].

The MSE calculates the errors between forecasted and actual values. The lower the MSE value, the more robust the forecasting model. It is computed using Eqn. (12).

$$MSE = \frac{1}{K}\sum_{i=1}^{K}\left(\hat{L}_i - L_i\right)^2 \tag{12}$$

where K denotes the number of data points, $\hat{L}_i$ denotes the forecasted load and L_i denotes the actual load.

The MAE help prevent the issue of mutual cancelling of errors. As a result, the MAE accurately reflects the real prediction error. It is computed using Eqn. (13).

$$MAE = \frac{1}{K}\sum_{i=1}^{K}\left|\hat{L}_i - L_i\right| \tag{13}$$

The MAPE reflects the relative difference between actual values and absolute errors. The MAPE is a significant evaluation metric in regression problems. It is computed using Eqn. (14).

$$MAPE = \sum_{i=1}^{K}\left|\frac{\hat{L}_i - L_i}{L_i}\right| \times \frac{100}{K} \tag{14}$$

R^2 is computed using Eqn. (15).

$$R^2 = 1 - \frac{\sum_{i=1}^{K}\left(\hat{L}_i - L_i\right)^2}{\sum_{i=1}^{K}\left(L_i - \overline{L}\right)^2} \tag{15}$$

where $\overline{L}$ denotes the average actual load.

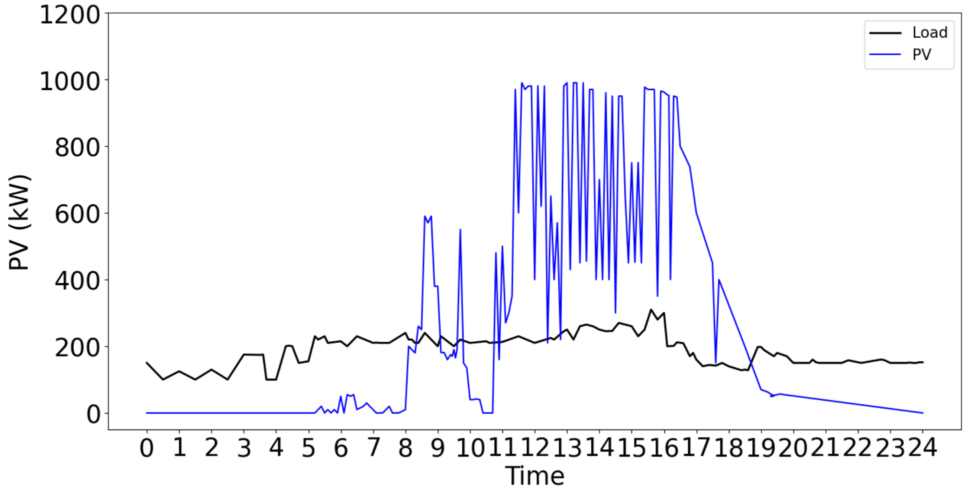
Comparative Analysis

The implied forecasting model's performance has been evaluated under a pair of circumstances.: (i) day type and (ii) seasons.

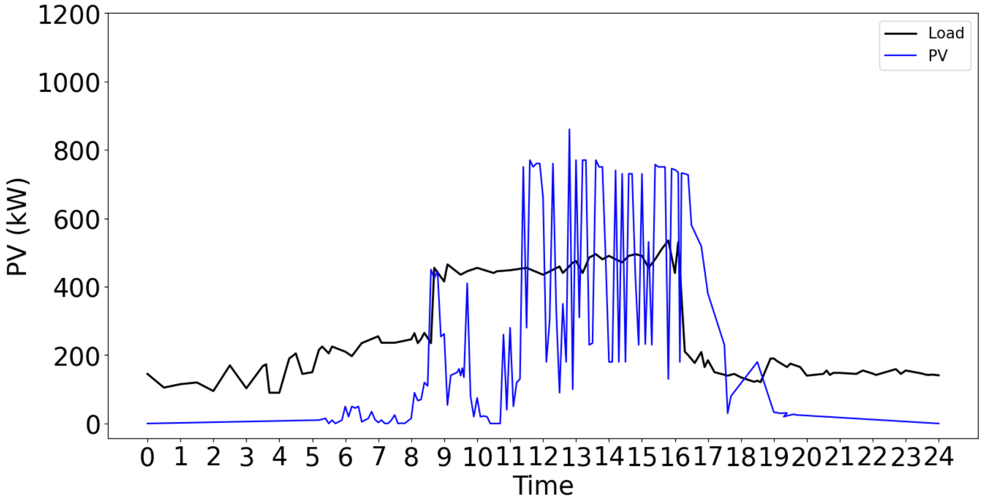
Comparative Analysis on Different Day Type

Fig 7 gives the actual PV production and consumption data for 24-hour time period under two different scenarios: a working day **(**June 12, 2019) and a holiday (June 16, 2019). The ambient temperature, particularly in the winter or summer, has an impact on load consumption. Moreover, Fig. 7 shows that the load on holidays is substantially different from the load on weekdays. As a result, in order to increase the forecast accuracy, the essential factors, such as day type, must be taken into account as input features for accurate load forecasting. In **Fig 8,** the PV forecast accuracy of the proposed and the existing methods are observed under two different day types. As seen in **Fig 8 (a)**, there is a more error in PV forecast by the existing methods such as LSTM+RF+SVR [25], XGB-LGBM-MLP [27], TCN-GRU [28], CNN+LSTM [29], Hybrid GA-CNN [30], VMD-VAE-LSTM [30], and edRVFL [31]. This shows that the existing methods did not accurately forecast

PV production. Since the existing methods are unable to accurately forecast the energy throughout the day, it must import energy at the last minute to balance the MG. This results in inefficient operational planning and energy management. On the other hand, the error between the actual and the forecasted PV production by the proposed HSAD-CNN model is low compared to the existing methods. This enhances the performance of energy management.

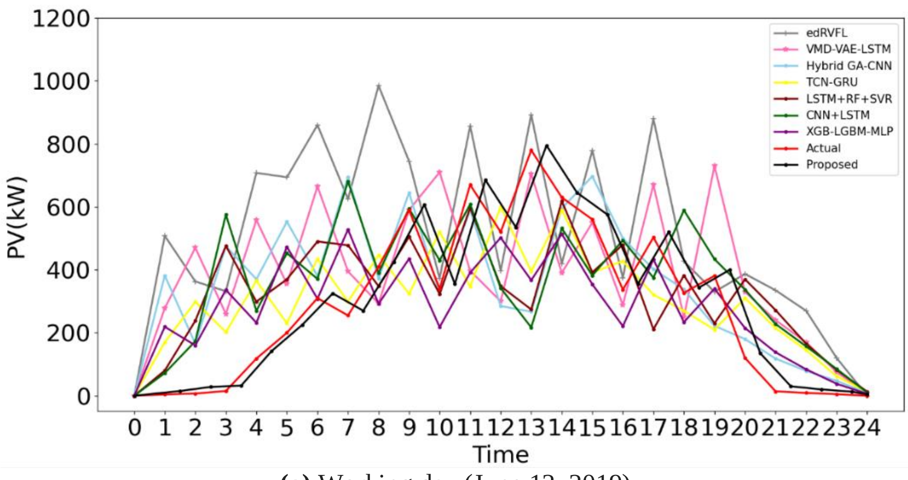


(a) Working day (June 12, 2019)

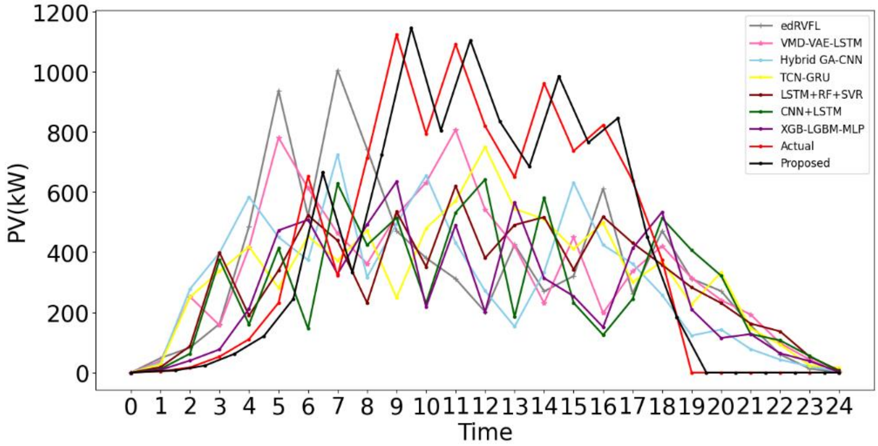


(b) Holiday (June 16, 2019)

Fig 7. PV Production and Load Data



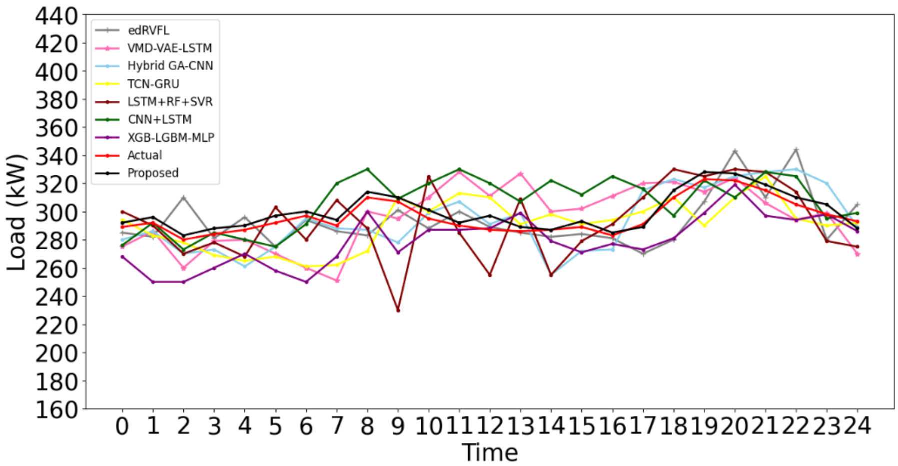
(a) Working day (June 12, 2019)



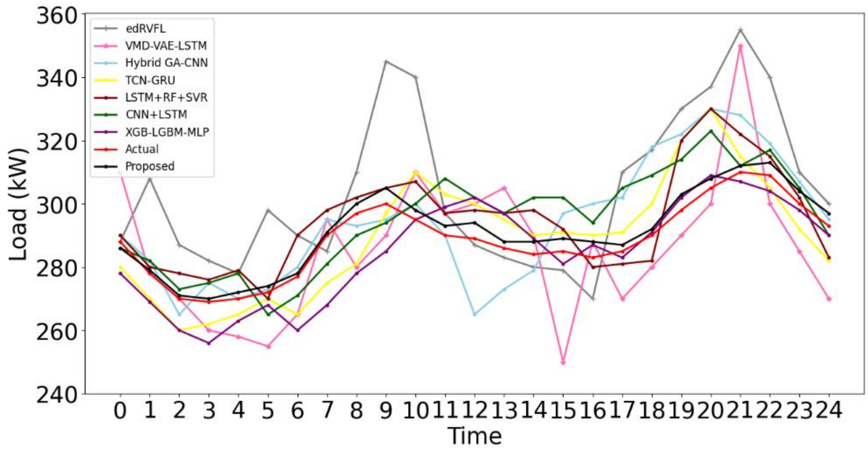
(b) Holiday (June 16, 2019)

Fig 8: PV Forecast

As seen in **Fig 9 (a)**, the load forecast is tested on two different scenarios: working day and holiday. From **Fig 9 (a)**, it is clear that both the proposed and most of the existing models are capable of forecasting the actual load demand on working day. However, in **Fig 9 (b),** the load forecast error on holiday by the existing techniques is more compared to that of working day. This is because, the existing methods did not consider day features which is crucial for accurate load forecasting.



(a) Working day (June 12, 2019)



(b) Holiday (June 16, 2019)

Fig 8. Load Forecast

Table 2. Comparative Analysis on Working Day (June 12, 2019)

Method/Metrics	MSE (kW)	MAE (kW)	MAPE (%)	R^2
XGB-LGBM-MLP [26]	2.575	2.982	12.69	0.883
CNN+LSTM [28]	1.388	1.975	11.69	0.924
LSTM+RF+SVR [24]	1.798	1.935	11.32	0.893
TCN-GRU [29]	0.983	1.275	8.35	0.936
Hybrid GA-CNN [29]	0.698	0.987	7.36	0.949
VMD-VAE-LSTM [30]	0.065	0.097	5.35	0.967
edRVFL [31]	0.132	0.278	4.98	0.974
Proposed	**0.024**	**0.095**	**3.47**	**0.987**

Table 3. Comparative Analysis on Holiday (June 16, 2019)

Method/Metrics	MSE (kW)	MAE (kW)	MAPE (%)	R^2
XGB-LGBM-MLP [26]	5.527	5.934	16.69	0.783
CNN+LSTM [28]	5.357	6.937	12.69	0.824
LSTM+RF+SVR [24]	3.571	3.977	13.37	0.875
TCN-GRU [29]	3.689	4.078	10.35	0.910
Hybrid GA-CNN [29]	1.675	1.933	9.78	0.921
VMD-VAE-LSTM [30]	2.877	2.967	8.34	0.927
edRVFL [31]	0.997	1.678	10.35	0.954
Proposed	**0.098**	**0.134**	**4.71**	**0.980**

Comparative Analysis on Different Seasons

Table 4 to **Table 7** gives the evaluation results of the proposed HSAD-CNN forecasting model and the existing models for four different seasons, such as spring, summer, autumn, and winter. A good performance network should have low MSE, MAE, MAPE, and high R^2. The proposed HSAD-CNN forecasting model has MSE, MAE, MAPE of 0.963 kW, 0.983 kW, 3.96 % respectively, and R^2 of 0.993 kW for the spring season. The results show that the proposed HSAD-CNN model has the least error and highest R^2 compared to the existing techniques, whereas the XGB-LGBM-MLP [29] method has the least performance with high MSE, MAE, and MAPE of 10.251 kW, 11.368 kW, and 10.39 % respectively, and low R^2 of 0.953 kW. Similarly, the HSAD-CNN forecasting model is also tested under other seasons such as summer, autumn, and winter. The results from Table 4–Table 7 show that the proposed HSAD-CNN model has also performed well in other seasons compared to the existing methods such as LSTM+RF+SVR [27], XGB-LGBM-MLP [29], TCN-GRU [30], CNN+LSTM [31], Hybrid GA-CNN [30], VMD-VAE-LSTM [30], and edRVFL [31]. This is because, the proposed HSAD-CNN model not only uses day features but also uses load as well as weather features to forecast the load. This reduces the forecasting error and improves the performance. Moreover, **Table 8** shows that the proposed HSAD-CNN model has less computational time compared to the existing techniques. The introduction of highway skip net in the proposed method improves the convergence speed by skipping the unimportant data patterns, thus reducing the time complexity.

Table 4. Comparative Analysis on Spring

Method/Metrics	MSE (kW)	MAE (kW)	MAPE (%)	R^2 (kW)
XGB-LGBM-MLP [26]	10.251	11.368	10.39	0.953
CNN+LSTM [28]	9.663	10.357	9.36	0.956
LSTM+RF+SVR [24]	7.394	8.369	8.39	0.962
TCN-GRU [29]	7.963	7.998	8.11	0.967
Hybrid GA-CNN [29]	4.589	4.963	5.39	0.983
VMD-VAE-LSTM [30]	1.398	1.986	4.39	0.986
edRVFL [31]	1.678	1.897	4.62	0.979
Proposed	**0.963**	**0.983**	**3.96**	**0.993**

Table 5. Comparative Analysis on Summer

Method/Metrics	MSE (kW)	MAE (kW)	MAPE (%)	R^2 (kW)
XGB-LGBM-MLP [26]	5.789	5.897	4.77	0.927
CNN+LSTM [28]	4.978	5.175	4.38	0.963
LSTM+RF+SVR [24]	4.675	4.978	2.37	0.974
TCN-GRU [29]	6.785	6.971	2.78	0.979
Hybrid GA-CNN [29]	2.736	2.976	0.97	0.983
VMD-VAE-LSTM [30]	1.093	1.273	0.91	0.989
edRVFL [31]	0.089	0.097	0.67	0.990
Proposed	**0.037**	**0.036**	**0.41**	**0.994**

Table 6. Comparative Analysis on Autumn

Method/Metrics	MSE (kW)	MAE (kW)	MAPE (%)	R^2 (kW)
XGB-LGBM-MLP [26]	13.787	14.017	12.98	0.938
CNN+LSTM [28]	12.982	13.287	11.97	0.939
LSTM+RF+SVR [24]	8.963	9.436	7.97	0.944
TCN-GRU [29]	8.247	8.937	7.08	0.951
Hybrid GA-CNN [29]	7.987	7.169	6.97	0.959
VMD-VAE-LSTM [30]	2.687	2.823	1.79	0.972
edRVFL [31]	2.373	2.412	1.44	0.978
Proposed	**0.967**	**0.974**	**0.91**	**0.983**

Table 7: Comparative Analysis on Winter

Method/Metrics	MSE (kW)	MAE (kW)	MAPE (%)	R^2 (kW)
XGB-LGBM-MLP [26]	10.251	11.368	10.39	0.943
CNN+LSTM [28]	12.689	13.987	9.54	0.953
LSTM+RF+SVR [24]	11.379	12.397	9.01	0.951
TCN-GRU [29]	8.369	8.398	5.39	0.955
Hybrid GA-CNN [29]	6.387	7.687	4.17	0.960
VMD-VAE-LSTM [30]	5.397	5.974	4.07	0.962
edRVFL [31]	4.398	5.312	3.46	0.973
Proposed	**1.793**	**1.983**	**1.01**	**0.981**

Table 8: Computational Complexity

Method	[26]	[28]	[24]	[27]	[29]	[30]	[31]	**Proposed**
Time (s)	26.258	12.74	10.25	2.99	7.17	15.77	2.30	**1.01**

V. CONCLUSION

In this paper, a HSAD-CNN forecasting model is proposed to capture long-term dependencies, capture time series trends and seasonality, and ultimately deal with the vanishing/exploding gradient problem. The HSAD-CNN forecasting model is made up of dilated casual convolutions enhanced with highway skip connections and normalization enable for the learning of long-term dependencies. The information flow is improved, and problems such as feature reuse, vanishing gradient, and negative learning issues are overcome by HS-net. The performance evaluation show that the proposed method has less MSE, MAE, MAPE, and high R^2 compared to the existing forecasting models. These improved results indicate that the HSAD-CNN model is able to forecast load even under varying climatic, seasonal, and day variation. Thus, the performance of the energy management can be improved by the proposed approach.

Data Availability

The Data used to support the findings of this study will be shared upon request.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding was received to assist with the preparation of this manuscript.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1]. J. Cronin, G. Anandarajah, and O. Dessens, "Climate change impacts on the energy system: a review of trends and gaps," Climatic Change, vol. 151, no. 2, pp. 79–93, Aug. 2018, doi: 10.1007/s10584-018-2265-4.

[2]. H. C. Bloomfield et al., "Quantifying the sensitivity of european power systems to energy scenarios and climate change projections," Renewable Energy, vol. 164, pp. 1062–1075, Feb. 2021, doi: 10.1016/j.renene.2020.09.125.

[3]. A. Q. Al-Shetwi, M. A. Hannan, K. P. Jern, M. Mansur, and T. M. I. Mahlia, "Grid-connected renewable energy sources: Review of the recent integration requirements and control methods," Journal of Cleaner Production, vol. 253, p. 119831, Apr. 2020, doi: 10.1016/j.jclepro.2019.119831.

[4]. X. Liang and C. Andalib -Bin- Karim, "Harmonics and Mitigation Techniques Through Advanced Control in Grid-Connected Renewable Energy Sources: A Review," IEEE Transactions on Industry Applications, vol. 54, no. 4, pp. 3100–3111, Jul. 2018, doi: 10.1109/tia.2018.2823680.

[5]. B. Marinescu, O. Gomis-Bellmunt, F. Dorfler, H. Schulte, and L. Sigrist, "Dynamic Virtual Power Plant: A New Concept for Grid Integration of Renewable Energy Sources," IEEE Access, vol. 10, pp. 104980–104995, 2022, doi: 10.1109/access.2022.3205731.

[6]. M. F. Roslan, M. A. Hannan, P. J. Ker, and M. N. Uddin, "Microgrid control methods toward achieving sustainable energy management," Applied Energy, vol. 240, pp. 583–607, Apr. 2019, doi: 10.1016/j.apenergy.2019.02.070.

[7]. N. Nasser and M. Fazeli, "Buffered-Microgrid Structure for Future Power Networks; a Seamless Microgrid Control," IEEE Transactions on Smart Grid, vol. 12, no. 1, pp. 131–140, Jan. 2021, doi: 10.1109/tsg.2020.3015573.

[8]. F. Rodríguez, A. M. Florez-Tapia, L. Fontán, and A. Galarza, "Very short-term wind power density forecasting through artificial neural networks for microgrid control," Renewable Energy, vol. 145, pp. 1517–1527, Jan. 2020, doi: 10.1016/j.renene.2019.07.067.

[9]. F. Zheng, X. Lin, Y. Lin, Y. Zhang, and Y. Zhang, "Design of a Novel Hybrid Control Strategy for ES Grid-Connected Inverter for Smooth Microgrid Transition Operation," IEEE Access, vol. 7, pp. 171950–171965, 2019, doi: 10.1109/access.2019.2955713.

[10]. G. Wang, X. Wang, and X. Gao, "Improved Seamless Switching Control Strategy for AC/DC Hybrid Microgrid," IEEE Access, vol. 9, pp. 55790–55801, 2021, doi: 10.1109/access.2021.3071821.

[11]. E. Samadi, A. Badri, and R. Ebrahimpour, "Decentralized multi-agent based energy management of microgrid using reinforcement learning," International Journal of Electrical Power & Energy Systems, vol. 122, p. 106211, Nov. 2020, doi: 10.1016/j.ijepes.2020.106211.

[12]. F. S. Al-Ismail, "DC Microgrid Planning, Operation, and Control: A Comprehensive Review," IEEE Access, vol. 9, pp. 36154–36172, 2021, doi: 10.1109/access.2021.3062840.

[13]. D. A. Gadanayak, "Protection algorithms of microgrids with inverter interfaced distributed generation units—A review," Electric Power Systems Research, vol. 192, p. 106986, Mar. 2021, doi: 10.1016/j.epsr.2020.106986.

[14]. W. Violante, C. A. Canizares, M. A. Trovato, and G. Forte, "An Energy Management System for Isolated Microgrids With Thermal Energy Resources," IEEE Transactions on Smart Grid, vol. 11, no. 4, pp. 2880–2891, Jul. 2020, doi: 10.1109/tsg.2020.2973321.

[15]. S. Aslam, H. Herodotou, S. M. Mohsin, N. Javaid, N. Ashraf, and S. Aslam, "A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids," Renewable and Sustainable Energy Reviews, vol. 144, p. 110992, Jul. 2021, doi: 10.1016/j.rser.2021.110992.

[16]. S. H. Rafi, Nahid-Al-Masood, S. R. Deeba, and E. Hossain, "A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network," IEEE Access, vol. 9, pp. 32436–32448, 2021, doi: 10.1109/access.2021.3060654.

[17]. H. J. Sadaei, P. C. de Lima e Silva, F. G. Guimarães, and M. H. Lee, "Short-term load forecasting by using a combined method of convolutional neural networks and fuzzy time series," Energy, vol. 175, pp. 365–377, May 2019, doi: 10.1016/j.energy.2019.03.081.

[18]. A. Yang, W. Li, and X. Yang, "Short-term electricity load forecasting based on feature selection and Least Squares Support Vector Machines," Knowledge-Based Systems, vol. 163, pp. 159–173, Jan. 2019, doi: 10.1016/j.knosys.2018.08.027.

[19]. H. H. H. Aly, "A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid," Electric Power Systems Research, vol. 182, p. 106191, May 2020, doi: 10.1016/j.epsr.2019.106191.

[20]. D. Arcos-Aviles et al., "An Energy Management System Design Using Fuzzy Logic Control: Smoothing the Grid Power Profile of a Residential Electro-Thermal Microgrid," IEEE Access, vol. 9, pp. 25172–25188, 2021, doi: 10.1109/access.2021.3056454.

[21]. D. Klutz, "Neurofeedback for Cognitive Enhancement, Intervention and Brain Plasticity," Journal of Biomedical and Sustainable Healthcare Applications, pp. 45–55, Jan. 2023, doi: 10.53759/0088/jbsha202303005.

[22]. W. Dong, Q. Yang, X. Fang, and W. Ruan, "Adaptive optimal fuzzy logic-based energy management in multi-energy microgrid considering operational uncertainties," Applied Soft Computing, vol. 98, p. 106882, Jan. 2021, doi: 10.1016/j.asoc.2020.106882.

[23]. A. A. Muzumdar, C. N. Modi, M. G. M, and C. Vyjayanthi, "Designing a Robust and Accurate Model for Consumer-Centric Short-Term Load Forecasting in Microgrid Environment," IEEE Systems Journal, vol. 16, no. 2, pp. 2448–2459, Jun. 2022, doi: 10.1109/jsyst.2021.3073493.

[24]. Y. Li, R. Wang, and Z. Yang, "Optimal Scheduling of Isolated Microgrids Using Automated Reinforcement Learning-Based Multi-Period Forecasting," IEEE Transactions on Sustainable Energy, vol. 13, no. 1, pp. 159–169, Jan. 2022, doi: 10.1109/tste.2021.3105529.

[25]. H. Zang et al., "Residential load forecasting based on LSTM fusing self-attention mechanism with pooling," Energy, vol. 229, p. 120682, Aug. 2021, doi: 10.1016/j.energy.2021.120682.

[26]. M. Massaoudi, S. S. Refaat, I. Chihi, M. Trabelsi, F. S. Oueslati, and H. Abu-Rub, "A novel stacked generalization ensemble-based hybrid LGBM-XGB-MLP model for Short-Term Load Forecasting," Energy, vol. 214, p. 118874, Jan. 2021, doi: 10.1016/j.energy.2020.118874.

[27]. H. Shi, L. Wang, R. Scherer, M. Wozniak, P. Zhang, and W. Wei, "Short-Term Load Forecasting Based on Adabelief Optimized Temporal Convolutional Network and Gated Recurrent Unit Hybrid Neural Network," IEEE Access, vol. 9, pp. 66965–66981, 2021, doi: 10.1109/access.2021.3076313.

[28]. S. H. Rafi, Nahid-Al-Masood, S. R. Deeba, and E. Hossain, "A Short-Term Load Forecasting Method Using Integrated CNN and LSTM Network," IEEE Access, vol. 9, pp. 32436–32448, 2021, doi: 10.1109/access.2021.3060654.

[29]. Y. Hong, Y. Chan, Y. Cheng, Y. Lee, J. Jiang, and S. Wang, "Week-ahead daily peak load forecasting using genetic algorithm-based hybrid convolutional neural network," IET Generation, Transmission & Distribution, vol. 16, no. 12, pp. 2416–2424, Mar. 2022, doi: 10.1049/gtd2.12460.

[30]. Y. Yang, Z. Wang, Y. Gao, J. Wu, S. Zhao, and Z. Ding, "An effective dimensionality reduction approach for short-term load forecasting," Electric Power Systems Research, vol. 210, p. 108150, Sep. 2022, doi: 10.1016/j.epsr.2022.108150.

[31]. R. Gao, L. Du, P. N. Suganthan, Q. Zhou, and K. F. Yuen, "Random vector functional link neural network based ensemble deep learning for short-term load forecasting," Expert Systems with Applications, vol. 206, p. 117784, Nov. 2022, doi: 10.1016/j.eswa.2022.117784.

[32]. S. Surana, K. Pathak, M. Gagnani, V. Shrivastava, M. T. R, and S. Madhuri G, "Text Extraction and Detection from Images using Machine Learning Techniques: A Research Review," 2022 International Conference on Electronics and Renewable Systems (ICEARS), Mar. 2022, doi: 10.1109/icears53579.2022.9752274.

[33]. M. T R, "An Efficient Ensemble Method Using K-Fold Cross Validation for the Early Detection of Benign and Malignant Breast Cancer," International Journal of Integrated Engineering, vol. 14, no. 7, Dec. 2022, doi: 10.30880/ijie.2022.14.07.015.

[34]. B. N. . V. S., A. . Kumar Bhagat, C. . C., B. . Sulaiman Alfurhood, A. . Pratap Singh, and M. . T. R., "Blockchain Based De-Duplication Analysis of Cloud Data with Data Integrity using Policy Based Encryption Technique in Cloud Storage", *Int J Intell Syst Appl Eng*, vol. 11, no. 3s, pp. 161–164, Feb. 2023.