STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **13**, April 2025, pp 1578–1594.
Published online in International Academic Press (www.IAPress.org)

Short-Term Load Forecasting Method for Renewable Energy Integration and Grid Stability Using CNN, LSTM, and Transformer Models

Khaoula Boumais *, Fayçal Messaoudi

Artificial Intelligence, Data Science and Emerging Systems Laboratory, National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco

Abstract This study examines the feasibility of combining Morocco's renewable energy plan with artificial intelligence to improve energy management in the industrial sector. Based on Moroccan Law 82-21, which promotes the self-consumption of renewable energy, the study addresses the fundamental difficulty of accurately estimating energy consumption in dynamic industrial environments. This difficulty is addressed using advanced machine learning models such as convolutional neural networks (CNNs), long-term memory networks (LSTMs) and transformers. The results show that deep learning models outperform classical methods such as ARIMA, with transformers and LSTM models excelling at handling erratic and steady energy consumption patterns, respectively.
In particular, hybrid CNN-LSTM architectures provide the highest level of accuracy, with prediction accuracy improved by up to 20%. While improving grid stability and renewable energy integration, this development has the potential to reduce operational costs by up to 30%. This analysis not only supports Morocco's ambitious goal of generating 52% of its electricity from renewables by 2030 but also highlights the critical role of AI-based solutions in creating a sustainable energy future.

Keywords Energy transition, Deep learning, Short-term load forecasting, Renewable energy self-consumption

DOI: 10.19139/soic-2310-5070-2256

1. Introduction

In recent years, renewable energy and artificial intelligence have become increasingly important due to several factors, including rising electricity prices, environmental pollution, and the depletion of fossil fuels. The issue of renewable energy gained momentum towards the end of the last century with the help of international political agreements such as the Kyoto Protocol [1]. Self-generation is at the centre of many activities and discussions around the world. It is a natural trend and an important development for renewable energy and decarbonization of the power sector. in the context of self-generation of electrical energy implementation, several countries around the world have shown the advantages of self-consumption for the prosumer and the grid perspective since it contributes to lower energy consumption from the grid and reduces energy offtake from the grid. Since then, the industrial sector has been directly responsible for 21% and indirectly responsible for 11% of total greenhouse gas emissions globally [2]. In some joint studies, 2000 PV networks for self-consumption of electricity have been implemented in different countries around the world, resulting in a significant impact in achieving 400 GW of energy capacity [3]. In the specific context of Morocco, the government has implemented policies to encourage the adoption of self-generation of renewable energy by industries looking to reduce their carbon footprint and energy costs [5]. In addition, the use of artificial intelligence (AI) such as Machine learning methods for forecasting in energy systems is gaining momentum to optimize energy usage and improve efficiency. The researchers found that the current electricity system is inefficient and that a new system based on local energy is needed to maintain the balance

*Correspondence to: Khaoula Boumais (Email: khaoula.boumais@usmba.ac.ma). Artificial Intelligence, Data Science and Emerging Systems Laboratory, National School of Applied Sciences, Sidi Mohamed Ben Abdellah University, Fez, Morocco.

ISSN 2310-5070 (online) ISSN 2311-004X (print)

between electricity supply and demand as well as power quality in the grid [6]. Considering this scenario, as well as the current digital energy transition and the penetration of renewable energy systems (RES), In this context, the implementation of a smart microgrid refers to a smart grid based on renewable energy and Machine learning (ML) that integrates information and cost-effectively controls infrastructure to enable more reliable, secure, and efficient operation of the electrical system. In this study, we aim to explore the advantages of Moroccan law on using self-consumed renewable energy for industry and ML methods for energy forecasting. It examines the legal framework for these technologies in Morocco and the benefits and challenges associated with their implementation.

2. Related works

2.1. Challenges and strategies for implementing self-generation of electricity for the industry in Morocco

The energy consumption rate by the industrial sector in Morocco is a significant contributor to the country's total energy demand. According to the Ministry of Energy Transition and Sustainable Development, the industrial sector accounts for approximately 40% of the country's total energy consumption, this high rate of energy consumption is primarily due to the increasing demand for energy-intensive manufacturing processes, such as steel production and chemical manufacturing.

2.1.1. Strategies of the Moroccan Code of Self-generation for Industry The government of Morocco has implemented various policies and initiatives to promote energy efficiency and renewable energy in the industrial sector, aiming to reduce the sector's energy consumption and dependency on fossil fuels. The Moroccan Code of Self-Consumption, also known as Law n 82-21, was published in the Official Bulletin on 27 February 2023, which sets out the main objectives of this Code:

- Encouraging industrial companies to generate their electricity from renewable energy sources for self-consumption.
- Allowing consumers to generate their electricity from renewable sources and consume it on-site reduces their reliance on the grid and electricity bills.
- Establishing a legal framework for self-consumption that guarantees the safety, quality, and reliability of the electricity produced and consumed.
- Facilitating the connection of self-consumption systems to the grid and ensuring their compatibility with the existing infrastructure.
- Contributing to developing a sustainable energy sector in Morocco and achieving the country's renewable energy targets of 52% of its electricity from renewable sources by 2030.
- Promoting energy efficiency and reducing the carbon footprint of the industrial sector in Morocco.
- Contributing to the development of a sustainable energy sector in Morocco and achieving the country's renewable energy targets.

This law also allows for net metering, which means that any excess energy produced can be sold back to the grid, up to a maximum of 20% of the annual production, at a fair price to be determined by the ONEE.

2.1.2. Renewable energy capacity for self-generation for industry Moroccan industries can adopt a distributed self-generation strategy to fulfil their energy requirements. This method involves producing electricity on-site using renewable energy sources like solar panels, wind turbines, or biomass plants and utilising it for their consumption. Morocco has benefited from its geographic location and natural surroundings and has become a global leader in renewable energy. For solar capacity, the average incoming solar radiation in Morocco varies between 4.7 and 5.6 kWh/m2/day, and the country's annual sunshine hours range from 2700 in the north to more than 3500 in the south. For these reasons, we have managed to install a capacity of 1770 MW through various projects in different regions, representing 16.57% of the total installed capacity in the Moroccan grid in 2021. For wind capacity, average wind speeds exceed 8 m/s in Morocco's windiest areas, located in the far north of the country, the South Atlantic, and between the Atlas and Rif Mountain ranges. For these reasons, we have managed to install a capacity of 1430 MW through various projects (Tangier II, Boujdour, Tiskrad, Midelt, Taza... [7]) in different regions, representing 13.4% of the total capacity installed on the Moroccan grid in 2021. The Office Chérifien des Phosphates (OCP) group is one of Morocco's largest industrial companies and is known for its significant self-generation of electricity. In 2019, OCP used 87% of its electricity from cogeneration or renewable sources. This includes electricity generated by exothermic processes used in the production of sulphuric acid, which drives turbines to generate electricity. The company's energy mix has shifted significantly towards clean energy, with clean energy accounting for 87% of total electricity usage in 2019, up from 71% in 2017. OCP's chemical processing facilities, such as Jorf Lasfar, even generate a surplus of clean electricity that is sold back to the grid. OCP's commitment to sustainability is exemplified by its ambitious target to produce 100% clean or renewable electricity by 2030. The Group also purchases electricity from wind farms through power purchase agreements, with the Gantour and Phosboucraa mining sites already using 100% renewable energy. These activities demonstrate OCP's commitment to mitigating climate change and promoting a circular economy. [8] According to the latest report from the Ministry of Energy, Morocco's total renewable energy capacity has reached 3.950 MW, while the country's goal is to reach 10.090 MW of renewable energy capacity by 2030 (see Table 1).

Table 1. Total installed renewable energy capacity in Morocco's grid [7].

Installed Power	**2021 Capacity (MW)**	**2030 Forecasting Capacity (MW)**
Renewable Energy Sources	3950	10,090

2.1.3. Energy performance of self-generation The combination of various consumer needs serves as the primary technical and financial justification for distributed or collective self-generation. The consumption and production of energy are based on actual, measured data. The self-consumption and self-sufficiency rates are used to evaluate the energy efficiency of such operations. While the self-sufficiency rate (SSR) describes the production of the electrical needs met by RE production, the self-consumption rate (SCR) refers to the portion of RE-produced electricity that is directly consumed locally [5]:

$$0SCR = \frac{\text{Quantity of RE electricity consumed}}{\text{Quantity of RE electricity produced}} \tag{1}$$

$$SSR = \frac{\text{Quantity of RE electricity consumed}}{\text{Quantity of total (grid with RE sources) electricity consumed}} \tag{2}$$

According to the study realized by O. Nait Mansour et al [9], The total self-consumed electricity generated by using photovoltaic as a renewable energy source in the Souss-Massa site is 15,141,347 kWh/year, thus reducing energy purchases from the grid to 8,381,188 kWh/year, representing a 34% reduction. Figure 1 shows the principle of self-generation for the industrial sector using artificial intelligence as a management solution. By implementing this strategy, Moroccan industries can decrease their dependence on the electricity grid, lower their energy expenses, and reduce their carbon emissions, thereby contributing to a greener environment.

2.2. Different stages of ML implementation for control and methods commonly used for energy forecasting

Intelligent control of electrical energy involves optimizing the use of energy within an industrial facility by monitoring and controlling the consumption of energy from various sources such as the grid, renewable energy sources, and energy storage systems [10, 11]. Machine learning (ML) can be used to create intelligent control systems that optimize energy consumption based on historical data, weather forecasts, and other real-time information. Here are some steps for implementing ML in intelligent control for industrial self-consumption:

1. **Data Collection:** Collect data on energy consumption and production, as well as other relevant information such as weather data, production schedules, and facility occupancy.
2. **Data Pre-processing:** Pre-process the data by cleaning, transforming, and normalizing it to make it suitable for use in ML algorithms.

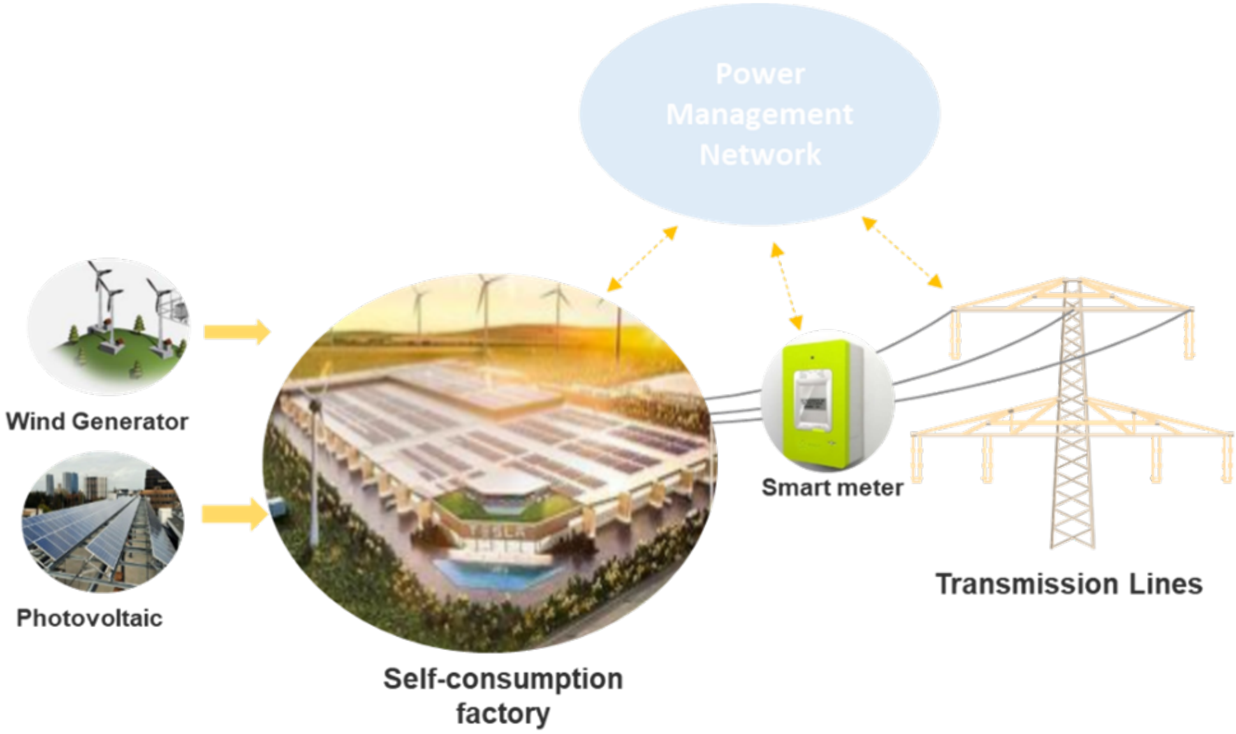


Figure 1. Demonstration of the self-generation industry

3. **Feature Engineering:** Identify and extract relevant features from the data that can be used to train the ML models.
4. **Model Selection:** Choose the appropriate ML model for the problem at hand, such as regression, classification, or clustering.
5. **Model Training:** Train the ML model using the pre-processed data and the selected features.
6. **Model Validation:** Validate the trained model using cross-validation or other techniques to ensure that it is accurate and generalizes well to new data.
7. **Model Deployment:** Deploy the trained model to the control system to make real-time predictions and optimize energy consumption.
8. **Continuous Monitoring and Improvement:** Monitor the performance of the ML model and continuously improve it by updating the model with new data and re-training it when necessary.

In Figure 2, ML aggregation models are widely used to estimate electric load forecasts, consumption forecasts, or peak power demand forecasts. Thus, many researchers have proven that implementing ML for power management in the industry is a solution to the traditional grid problem [12, 22].

The field of short-term load forecasting has undergone significant developments to facilitate the integration of renewable energy sources and ensure the stability of power systems. While traditional models such as ARIMA are highly effective for linear models, they tend to be less robust in the presence of non-linear dependencies. The combination of techniques, such as the ARIMA-SVM approach, can enhance the accuracy of forecasting models by addressing both linear and non-linear components [23, 25]. Deep learning models, such as recurrent neural networks (RNN), long short-term memory (LSTM), and gated recurrent units (GRU), are particularly adept at capturing temporal relationships. In certain tasks, such as offshore wind forecasting, the seasonal autoregressive integrated moving average (SARIMA) model outperforms GRU and LSTM models. Recently, there have been significant advancements in combining convolutional neural networks (CNN) and LSTM to enhance feature extraction and temporal learning, aided by attention processes [26, 28]. Hybrid models such as MTMV-CNN-LSTM and Quantile Regression-CNN-GRU demonstrate superior accuracy across diverse time horizons. In contrast, CGANs facilitate model learning with synthetic data, achieving high accuracy in challenging scenarios such as winter demand peaks (cited in references [29, 31]. The incorporation of multi-task learning and feature separation-fusion algorithms into advanced frameworks has demonstrated exceptional performance in integrated energy systems, with accuracy exceeding 98% and minimal relative errors. These advances not only enhance the precision of predictions but also address the challenges posed by dynamic energy systems, thereby facilitating more judicious decision-making. The combination of adaptive methods with robust data augmentation techniques enables effective resource allocation, improves grid resilience and facilitates sustainable energy transitions [32, 34].

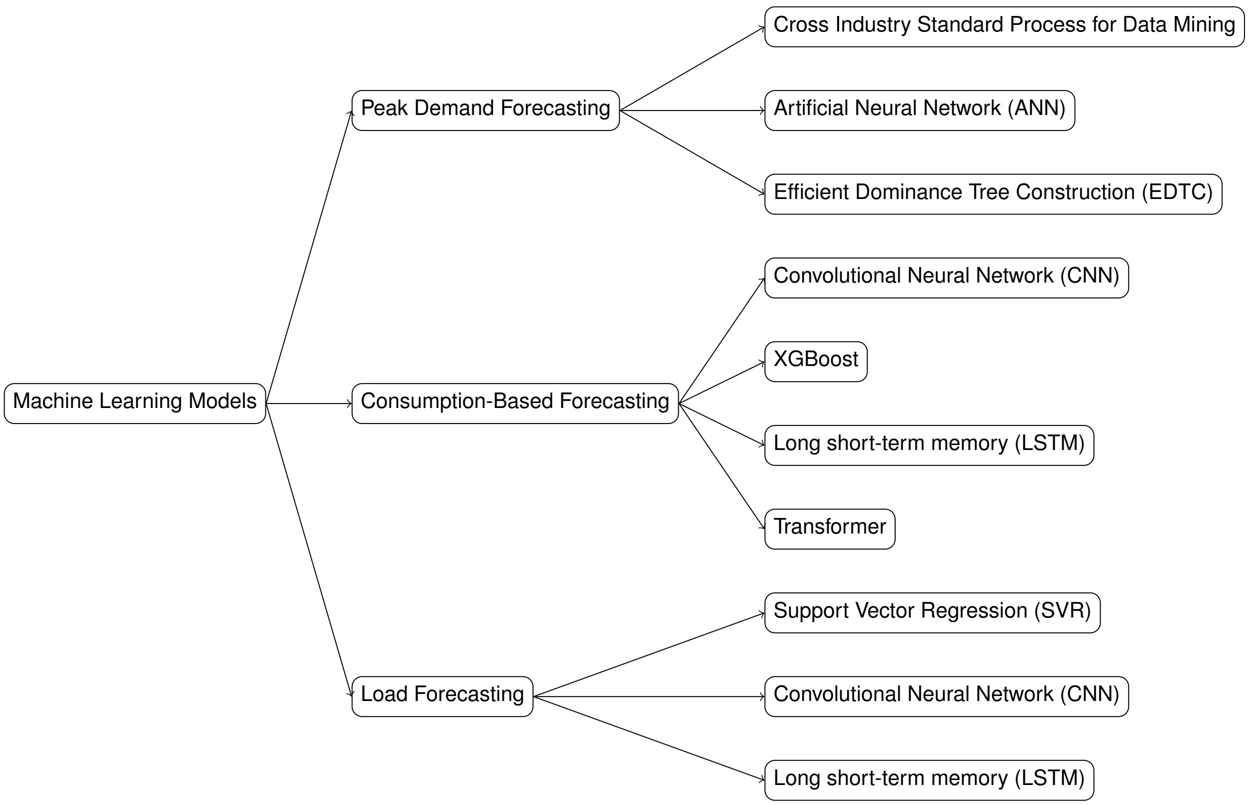


Figure 2. Different Machine Learning Models for Energy Forecasting

3. Proposed method

The following study focuses on industrial load forecasting in Morocco, with a particular emphasis on the use of deep learning models for energy demand management and the incorporation of renewable energy into short-term forecasting approaches. In Figure 2, we analyse the performance of deep learning models such as CNN, LSTM, and Transformer for forecasting power demand.

3.1. Data pre-processing

The dataset underwent numerous pre-processing stages before its utilisation in the training of the models. Initially, the dataset was examined for missing values, but none were identified. Temporal information (such as day, week, and time) was transformed into cyclic features by applying sine and cosine transformations, thereby enabling the models to capture the periodicity inherent to temporal data. Meteorological data, including temperature, humidity, wind speed, and flux measurements, were scaled using the MinMaxScaler to ensure all features aligned on the same scale, facilitating model convergence.

3.2. System Architectures

The study employs three deep learning models: CNN, LSTM, and Transformer, each tailored to specific consumption patterns and load forecasting difficulties.

a) CNN Architecture

The CNN architecture was meant to capture short-term regional consumption patterns. It employs a Conv1D layer to detect trends and patterns in historical load data and maximum pooling layers to minimise the dimensionality of recovered features. Activation layers have been added to generate non-linearities, allowing the model to recognise complicated patterns driven by weather and exceptional occurrences. The result is processed by fully connected layers, which produce load demand predictions. The Convolutional Neural Network (CNN) is designed to capture short-term patterns in electrical power consumption. The main components of the architecture are as follows:

Input Layer The input to the CNN is a time series of power consumption data:

$$\mathbf{X} = [x_1, x_2, \dots, x_T] \in \mathbb{R}^T \tag{3}$$

where T is the number of time steps.

Convolutional Layer (Conv1D) The Conv1D layer detects local patterns in the time series using a kernel of size K. The convolution operation is defined as:

$$y_t = \sigma\left(\sum_{i=1}^{K} w_i x_{t+i-1} + b\right) \tag{4}$$

where:

- y_t is the output at time step t,
- w_i are the kernel weights,
- b is the bias term,
- $\sigma(\cdot)$ is the activation function (e.g., ReLU).

Max Pooling Layer To reduce dimensionality, a max pooling operation is applied with a window of size P:

$$y_t^{\text{pool}} = \max\{y_t, y_{t+1}, \dots, y_{t+P-1}\} \tag{5}$$

Fully Connected Layer The pooled features are passed through a fully connected layer:

$$\hat{y} = \sigma\left(\sum_{i=1}^{N} w_i z_i + b\right) \tag{6}$$

where z_i are the flattened features, and w_i are the weights.

Output Layer The final prediction of load demand is obtained through the output layer:

$$\hat{y}_t = \sum_{i=1}^{M} w_i^{\text{out}} z_i + b^{\text{out}} \tag{7}$$

where M is the number of neurons in the final layer.

b) LSTM Architecture

The Long Short-Term Memory (LSTM) network is designed to capture long-term dependencies in time series data, making it suitable for forecasting electrical power consumption. The main components of the architecture are as follows:

Input Layer The input to the LSTM is a time series of power consumption data:

$$\mathbf{X} = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^T \tag{8}$$

where T is the number of time steps.

Forget Gate The forget gate determines which part of the previous cell state should be forgotten. It is computed as:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{9}$$

Input Gate The input gate decides how much of the new information should be added to the cell state:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{10}$$

Output Gate The output gate controls what part of the cell state should be output as the hidden state for the next time step:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{11}$$

where:

- W_f, W_i, W_o are the weight matrices,
- b_f, b_i, b_o are the bias terms,
- h_{t-1} is the hidden state from the previous time step,
- $\sigma(\cdot)$ is the sigmoid activation function.

C) Transformer Architecture

The Transformer model is designed to handle sequential data by employing self-attention mechanisms, making it effective for power consumption forecasting. The main components of the architecture are as follows:

Input Layer The input to the Transformer is a time series of power consumption data, which is typically embedded into a higher-dimensional space:

$$\mathbf{X} = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^T \tag{12}$$

where T is the number of time steps.

Positional Encoding Since Transformers do not inherently capture the order of sequences, positional encodings are added to the input embeddings:

$$\text{PE}(pos, 2i) = \sin\left(\frac{pos}{10000^{\frac{2i}{d}}}\right), \quad \text{PE}(pos, 2i+1) = \cos\left(\frac{pos}{10000^{\frac{2i}{d}}}\right) \tag{13}$$

where pos is the position and d is the dimensionality of the embeddings.

Self-Attention Mechanism The self-attention mechanism computes attention scores based on query, key, and value representations:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{14}$$

where Q, K, and V are the query, key, and value matrices derived from the input, and d_k is the dimensionality of the keys.

Multi-Head Attention To capture multiple relationships, multi-head attention is utilized:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_h)W^O \tag{15}$$

where each head is computed as:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V) \tag{16}$$

and W_i^Q, W_i^K, W_i^V are the learned projection matrices.

Feed-Forward Network Each output from the attention layer passes through a feed-forward network:

$$\text{FFN}(x) = \text{ReLU}(xW_1 + b_1)W_2 + b_2 \tag{17}$$

Output Layer Finally, the outputs are transformed into predictions through a linear layer:

$$\hat{y} = W_y h + b_y \tag{18}$$

where h is the output from the last layer, and W_y and b_y are the weights and bias for the output layer.

3.3. Model Training and Evaluation

All models were trained using the Adam optimizer, with 64 hidden units and 100 epochs. For evaluation, the Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) metrics were used to measure model accuracy. The following equations represent the evaluation metrics:

Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left((y_{\text{pred}_i} - y_{\text{true}_i})^2\right). \tag{19}$$

Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left((y_{\text{pred}_i} - y_{\text{true}_i})^2\right)}. \tag{20}$$

Mean Absolute Percentage Error (MAPE):

$$\text{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{|y_{\text{pred}_i} - y_{\text{true}_i}|}{|y_{\text{true}_i}|}\right) \times 100. \tag{21}$$

Where y_{pred} represents the predicted electricity demand and y_{true} is the actual value of the dataset. Lower values for MSE, RMSE, and MAPE indicate better forecasting accuracy.

4. Experimental results and discussion

This section outlines the experiments conducted to evaluate the effectiveness of the LSTM, CNN, and Transformer models in forecasting energy consumption across three zones of the power network. The results demonstrate that each model is effective at processing the dataset and estimating short-term energy requirements.

4.1. Dataset and Model Parameters

The dataset employed in this study comprises energy consumption statistics for three grid zones in a Moroccan industry, collated at 30-minute intervals between 1 January and 31 December 2023. The dataset, comprising 14 columns and 54,144 rows, contains a wealth of information that is well-suited to the purpose of energy forecasting. The dataset comprises meteorological variables, including temperature in degrees Celsius, humidity in percentage, wind speed in kilometres per hour, general diffuse flux, and diffuse flux. Additionally, it contains temporal attributes such as year, hour, minute, day, weekday, weekend, week number, and month. Furthermore, it includes microgrid electricity consumption data in kilowatts, derived from the manufacturing plant. To prepare the dataset for modelling, MinMaxScaler normalisation was employed to provide consistent scaling between features, which served to enhance the model's capacity to capture underlying patterns. Subsequently, The data was then split into training (80%) and testing (20%) subsets for each zone. The selected hyperparameters, shown in Table 2, were optimized through cross-validation, which supports robust model selection and helps avoid overfitting by ensuring that the model's performance generalizes well across different data splits. The experiments were conducted using the Python 3.7 programming language and the Keras machine learning library on a workstation equipped with an 11th-generation Intel® Core™ i7-1165G7 central processing unit (CPU) operating at 2.80 gigahertz (GHz) and 8 gigabytes (GB) of random-access memory (RAM), running the Windows 11 Pro operating system.

Table 2. Summary of Experimental Results and Model Hyperparameters.

Groups	**Hyperparameters**	**Values**
Experimental Setup	Dataset Size	1 Year (30-minute intervals)
	Features Used	{Year, Hour, Minute, Temperature, Humidity, etc.}
	Target Variables	Power Consumption (Zone 1, Zone 2, Zone 3)
	Data Normalization	MinMaxScaler
	Input Sequence	Reshaped into sequences for LSTM and CNN
Model Hyperparameters	Optimization Algorithm	Adam
	Activation Function	ReLU
	Training Epochs	100 epochs
	Cross-Validation	5-Fold
	LSTM Hidden Units	64
	CNN Kernel Size	2
	Transformer Model Architecture	LSTM-based
	Batch Size	32
	Loss Function	MSE

The combination of hyperparameters and cross-validation assures that the selection of the model is both comprehensive and data-driven. Each model was trained with optimal parameters that balanced computational efficiency and predicted accuracy. The application of cross-validation facilitates the efficient adjustment of model parameters, thereby ensuring the reliability of performance evaluations on diverse subsets of data. The dataset employed in this study, which contains a wide variety of temporal, environmental, and energy parameters, is sufficient to capture the necessary variability for reliable forecasting. The extensive scope of the data provides a robust basis for assessing the performance of the models in relation to the study's objectives, encompassing the primary determinants of power consumption without incorporating new external variables.

4.2. Model performance across microgrid zones

The performance of the LSTM, CNN, and Transformer models was evaluated through comprehensive testing in three distinct power consumption zones, each characterised by a unique set of demand patterns.

As can be seen in Table 3, in Zone 1, which exhibits relatively consistent consumption with moderate fluctuations, both the LSTM and Transformer models demonstrated excellent prediction accuracy, with a mean absolute error (MAE) of 2500 kW. The root mean square error (RMSE) for the LSTM was somewhat lower at 3400 kW in comparison to 3600 kW for the Transformer, with similar mean absolute percentage errors (MAPE) of 11.85% for the LSTM and 11.31% for the Transformer. In contrast, the CNN model exhibited a notable deficiency in performance, with a higher mean absolute percentage error (MAPE) of 3400 kW and a root mean square error (RMSE) of 4300 kW. This indicates an inability to accurately capture the stable consumption trends observed in the data. In Zone 2, which exhibits more volatile consumption patterns, likely due to operational changes or external factors, the Transformer model demonstrated superior performance, with the lowest MAE of 1000 kW, RMSE of 2200 kW, and MAPE of 13.54%. This highlights the effectiveness of the Transformer model in capturing irregular fluctuations. The LSTM model demonstrated satisfactory performance, with an MAE of 1,500 kW. Nevertheless, the RMSE of 2,200 kW indicates potential challenges in accurately forecasting sudden demand surges. The convolutional neural network (CNN) model demonstrated suboptimal performance, exhibiting the highest error rates (mean absolute error (MAE) of 2,700 kW and root mean square error (RMSE) of 3,300 kW). This result suggests that the model is unable to respond adequately to rapid changes in consumption. In Zone 3, which exhibited significant fluctuations due to intermittent peak demands, the LSTM model demonstrated the most robust performance, with an MAE of 1,000 kW and an RMSE of 1,500 kW. This demonstrates that the model was capable of effectively addressing the considerable variability. It is noteworthy that the transformer demonstrated commendable performance, with an MAE of 1,000 kW and an RMSE of 1,700 kW. However, its mean absolute percentage error (MAPE) was considerably higher (26.16%) than that of the long short-term memory (LSTM) model (24.62%), indicating a potential difficulty in dealing with extreme outliers. The convolutional neural network (CNN) model exhibited the poorest performance, with a MAPE of 3,500 kW and an RMSE of 4,500 kW. This illustrates the model's challenges in controlling the significant fluctuations in demand characteristic of this zone.

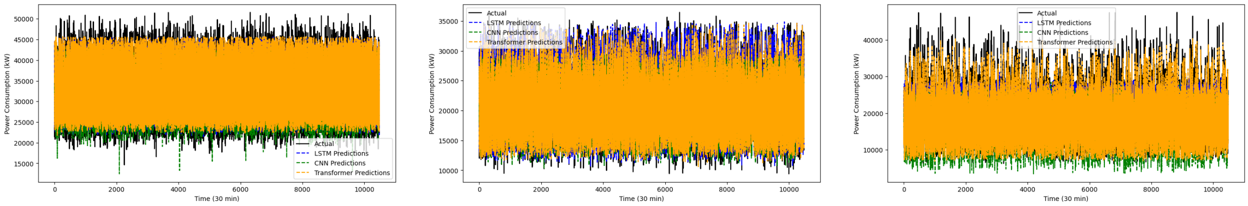
Table 3. Energy consumption forecasting results across three zones.

Model	MAE (kW)	RMSE (kW)	MAPE (%)
	Zone 1		
LSTM	2500	3400	11.85
CNN	3400	4300	17.45
Transformer	2500	3600	11.31
	Zone 2		
LSTM	1500	2200	14.86
CNN	2700	3300	17.93
Transformer	1000	2200	13.54
	Zone 3		
LSTM	1000	1500	24.62
CNN	3500	4500	20.24
Transformer	1000	1700	26.16

Figures 3 and 4 illustrate the efficacy of the models by contrasting the actual and forecasted energy consumption values for the three zones. The analysis indicates that the Transformer model demonstrates particular efficacy in the context of fluctuating consumption patterns, notably in Zone 2. This renders it an efficient alternative for forecasting in areas characterised by frequent fluctuations in demand. Moreover, the LSTM model demonstrates robust performance, particularly in settings with both consistent and highly fluctuating consumption patterns, as observed in zones 1 and 3. In contrast, the CNN model consistently demonstrates suboptimal performance across all zones, underscoring its inadequacy for forecasting industrial power usage. A hybrid strategy that combines the strengths of the LSTM model in stable and extreme settings with the capacity of the Transformer to react to volatility should considerably improve forecast accuracy. Moreover, future developments may be enhanced by the integration of external variables, such as operational programmes and weather forecasts, to enhance responsiveness to sudden changes in demand.

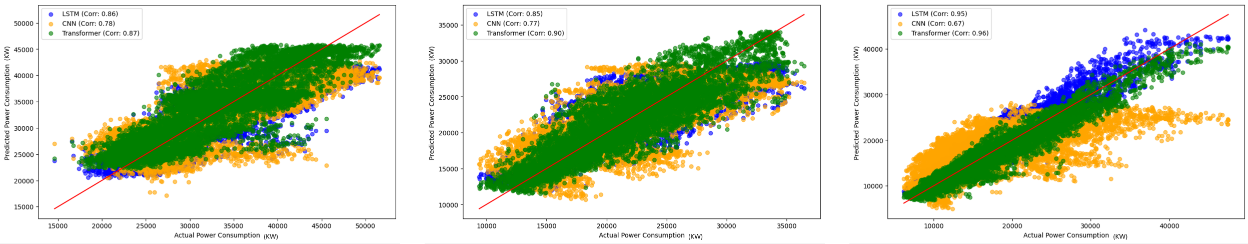
4.3. Error Distribution and Consistency

To provide a comprehensive comparison of model performance, Figures 5, 6, and 7 illustrate the diverse forecasting capabilities of the LSTM, CNN, and Transformer models across three power zones, thereby facilitating a comprehensive assessment of model performance. Figure 5 compares the RMSE and MAE values for the three



(a) Power Consumption (Zone1) (b) Power Consumption (Zone2) (c) Power Consumption (Zone 3)

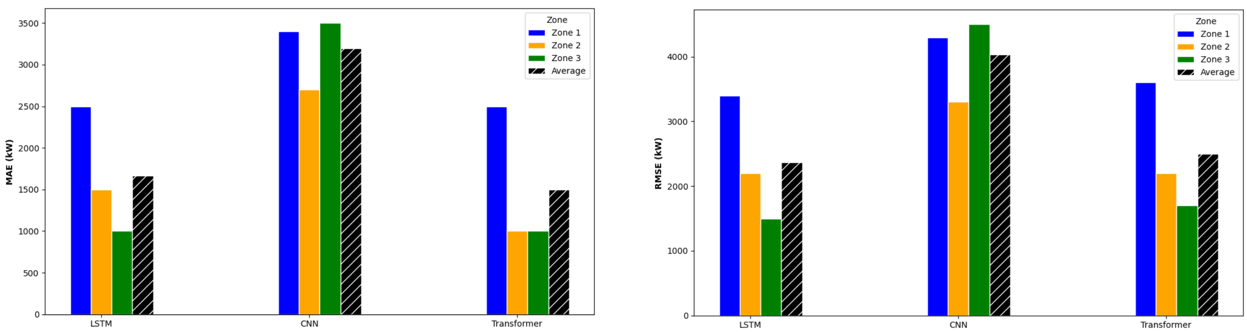
Figure 3. Comparison of Predictions vs Actual Power Consumption for Power Zones



(a) Power Zone 1 (b) Power Zone 2 (c) Power Zone 3

Figure 4. Correlation between the actual and the predicted consumption for each power grid zone.

models. The Transformer model performs well in zone 1, with an MAE of 2,500 kW and an RMSE of 3,600 kW, matching the results of the LSTM (MAE 2,500 kW, RMSE 3,400 kW) and exceeding those of the CNN (MAE 3,400 kW, RMSE 4,300 kW). Similarly, in Zone 2, the transformer model demonstrates superior performance compared to the CNN model. The transformer model achieved an MAE of 1,000 kW and an RMSE of 2,200 kW, while the CNN model exhibited an MAE of 2,700 kW and an RMSE of 3,300 kW. In Zone 3, the LSTM outperforms the transformer in terms of RMSE (1,500 kW vs. 1,700 kW), but both models outperform the CNN, which continually falls behind in all zones.



(a) MAE for different models (b) RMSE for different models

Figure 5. Bar graph of RMSE and MAE for different models.

Figure 6 shows the variability in prediction error across models, using box plots of the mean absolute percentage error (MAPE). In Zone 1 (Figure 6a), the Transformer and LSTM demonstrate narrow interquartile ranges and low median MAPE values (11.31% and 11.85%, respectively), indicating reliable and accurate predictions. In contrast, the CNN exhibits a broader interquartile range and a higher median MAPE (17.45%), indicating a higher level of error variability. In Zone 2, the Transformer continues to demonstrate superior performance, with a lower median MAPE (13.54%) and a smaller interquartile range, indicating its capacity to manage more volatile data

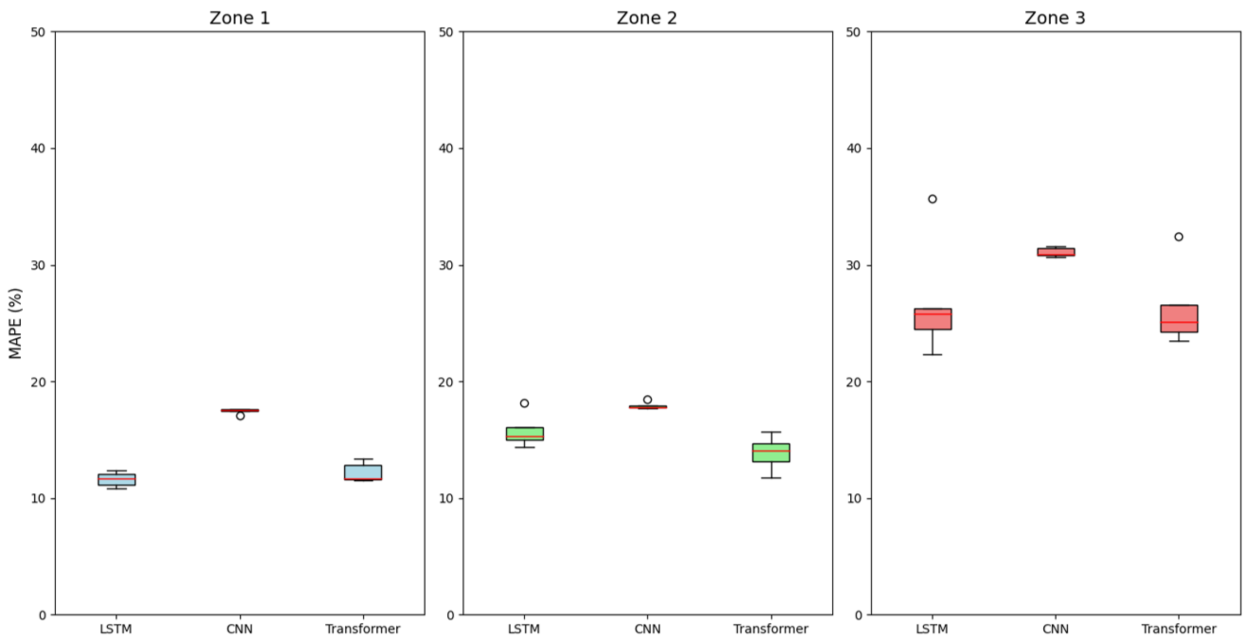


Figure 6. Box Plot of MAPE for different models across three power zones.

than the other models. The LSTM demonstrated a satisfactory level of performance, with a median mean absolute percentage error (MAPE) of 14.86%. However, the performance of the model was not consistent. The convolutional neural network (CNN) exhibited a less favourable performance, with a higher median mean absolute percentage error (MAPE) of 17.93%. In Zone 3, the LSTM model demonstrates greater consistency in its predictions, with a median MAPE of 24.62%. The Transformer model demonstrated a noteworthy performance in this zone, with a median MAPE of 26.16%. Nevertheless, the variability of the Transformer projections is slightly greater. The CNN model demonstrated the poorest performance across all zones.

As illustrated in Figure 7, the training and validation loss curves for the LSTM, CNN and Transformer models elucidate the discrepancies in convergence between the three feed zones. The Transformer and LSTM models demonstrate robust and consistent convergence, with training and validation losses exhibiting close alignment over 100 epochs. This indicates that the models exhibit minimal overfitting and robust generalisation across a range of scenarios. This demonstrates the models' capacity to process the dataset's temporal and meteorological characteristics, including temperature, humidity, and wind speed. The Transformer model demonstrates superior performance due to its capacity to address intricate long-term dependencies, enabling it to discern volatility in zones 1 and 2, where data unpredictability is more pronounced. The LSTM model's capacity to effectively distinguish sequential dependencies allows it to demonstrate competence in all zones, particularly zone 3, where it has a modest RMSE advantage over the transformer. This is due to the great stability of consumption patterns. In contrast, the CNN model demonstrates slower convergence and elevated validation losses, indicating that it encounters difficulties in capturing the temporal complexity of the data. The inferior performance of the CNN model can be attributed to its reliance on geographical rather than temporal models, which are essential for forecasting energy usage. The alignment of losses in the Transformer and LSTM models demonstrates their resilience. However, the CNN model's greater validation losses confirm that it is not suitable for this application. The findings illustrate that both the Transformer and the LSTM models are capable of accurately estimating energy consumption. However, the Transformer exhibits superior performance in terms of data volatility, particularly in power zones 1 and 2. This functionality renders the Transformer the optimal solution for dynamic environments where conditions may fluctuate rapidly. In contrast, the LSTM model demonstrates superior performance in power zone 3, where prediction stability is of paramount importance due to the relatively constant nature of the data.

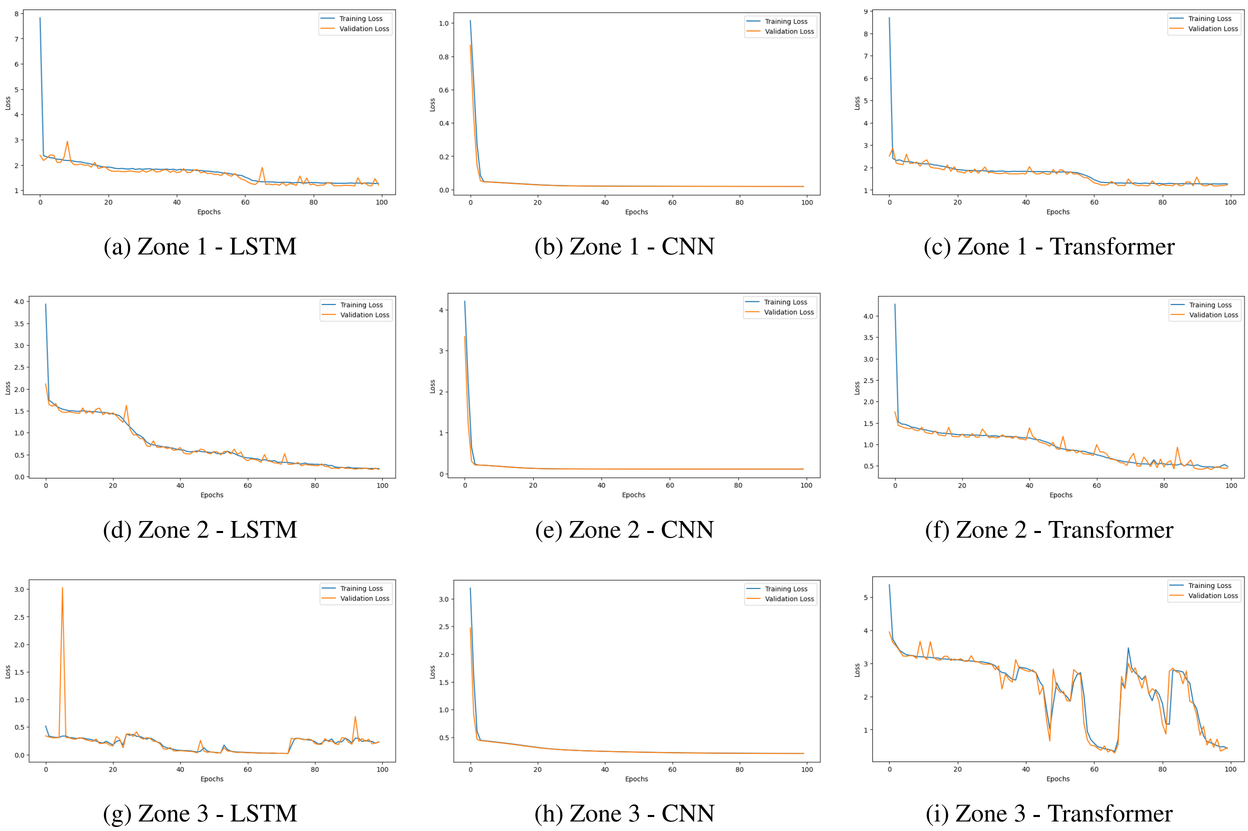
Conversely, the CNN model consistently demonstrates inferior performance across all metrics, indicating its inability to effectively capture the requisite temporal relationships for accurate energy consumption prediction. Consequently, it is unsuitable for this particular use case.

4.4. *Comparative Analysis of Model Efficiency and Limitations with Traditional and Hybrid Models*

The dataset employed in this study comprises statistics on electricity consumption in a range of geographical areas, indicating network users' average energy consumption. To evaluate the model, the average electricity consumption was used as the target variable, which is defined as the average power consumption in three areas to represent a broad consumption pattern.

$$y_{\text{avg}} = \frac{\text{PowerConsumption\_Zone1\_KW} + \text{PowerConsumption\_Zone2\_KW} + \text{PowerConsumption\_Zone3\_KW}}{3} \tag{22}$$

The models tested included Long Term Memory (LSTM), Convolutional Neural Network (CNN), Transformer, CNN-LSTM, Autoregressive Integrated Moving Average (ARIMA), and Exponential Smoothing, to predict future energy consumption. To enhance the efficacy of the model, data normalisation was employed during the learning phase, followed by inverse normalisation during the testing phase, thus facilitating the generation of interpretable error measurements. Table 4 presents the findings for each model, comparing the mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE) for estimating future energy usage. As demonstrated in the table, the convolutional neural network model with long-term memory (CNN-LSTM) exhibited the most optimal performance, with the lowest MAE (2851.70 kW), RMSE (3815.98 kW), and MAPE (11.47%).



(a) Zone 1 - LSTM (b) Zone 1 - CNN (c) Zone 1 - Transformer

(d) Zone 2 - LSTM (e) Zone 2 - CNN (f) Zone 2 - Transformer

(g) Zone 3 - LSTM (h) Zone 3 - CNN (i) Zone 3 - Transformer

Figure 7. The training and validation loss curves for the LSTM, CNN, and Transformer models are shown in Figures (a-c) for zone 1, Figures (d-f) for zone 2, and Figures (g-i) for zone 3. The figures exhibit the models' performance over 100 epochs, demonstrating consistent convergence with minimal overfitting, as evidenced by the near alignment of the training and validation losses. Each model adequately accommodates the varying energy consumption trends in each zone.

Table 4. Load forecasting average results for all models.

Model	MAE (kW)	RMSE (kW)	MAPE (%)
LSTM	3046.18	4041.98	12.76
CNN	4161.78	5113.12	18.40
Transformer	3177.31	4067.59	13.87
CNN-LSTM	2851.70	3815.98	11.47
ARIMA	5381.19	6882.49	20.67
Exponential Smoothing	4731.19	5960.81	19.63

Table 5 presents a comparative analysis of the computational complexity of the models employed for energy consumption prediction. The time required for learning varies considerably between models, with the LSTM and Transformer models requiring the longest (65.38 and 65.63 seconds, respectively). In contrast, the exponential smoothing model exhibited the shortest learning time (0.92 seconds). In terms of inference speed, the convolutional neural network (CNN) exhibited the fastest performance, with an inference time of 0.000056 seconds per sample. This makes it an ideal choice for real-time forecasting applications. In contrast, while the ARIMA and exponential smoothing models are efficient, they may lack the requisite predictive accuracy for larger-scale applications. The LSTM model exhibited the highest memory usage (4190.18 MB), while the exponential smoothing model demonstrated remarkable efficiency (0.05 MB). These findings underscore the intrinsic trade-off between computational cost and predictive capability, underscoring the necessity of selecting models that strike a balance between efficiency and accuracy, tailored to the specific requirements of real-world applications, particularly in the context of large, dynamic datasets.

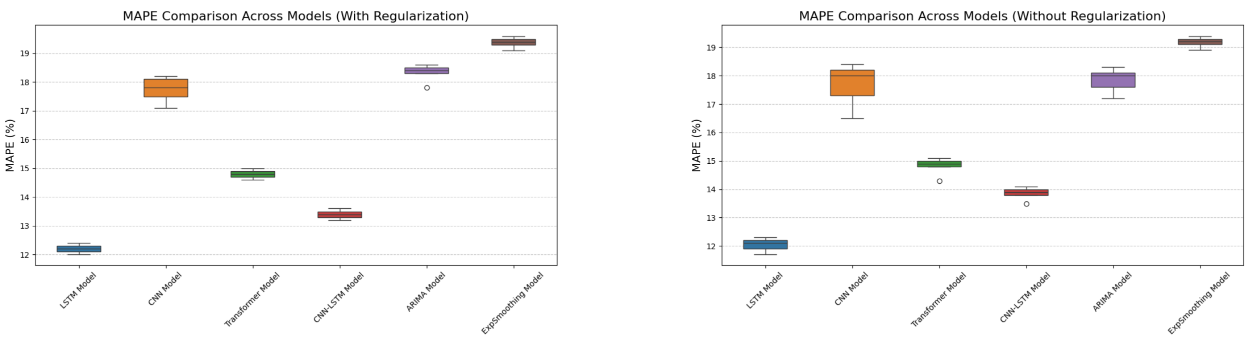
Table 5. Computational complexity analysis of the models: Training Time, Inference Speed, and Memory Usage.

Model	Training Time (seconds)	Inference Speed (seconds/sample)	Memory Usage (MB)
LSTM	65.38	0.000113	4190.18
CNN	19.21	0.000056	2.94
Transformer	65.63	0.000110	22.59
CNN-LSTM	64.63	0.000114	253.01
ARIMA	6.59	0.000000	75.82
Exponential Smoothing	0.92	0.000000	0.05

In summary, the CNN-LSTM model was identified as the most effective method for achieving an optimal balance between predictive accuracy and computing needs in energy consumption forecasting. The hybrid architecture, which captures both spatial and long-term dependencies using convolutional neural networks (CNN) and long short-term memory (LSTM) layers, renders it an optimal choice for modelling complex, multivariate time-series data. Classic models such as ARIMA and exponential smoothing are computationally simple but may lack the flexibility to describe nonlinear patterns of energy consumption. Transformers and other deep learning models provide solid performance but are resource-intensive, demonstrating a trade-off between accuracy and efficiency. These insights emphasise the need for a model that is tailored to the specific restrictions and objectives of the forecasting application, especially in scenarios with large and dynamic datasets.

4.5. Analysis of Regularization Impact

The objective was to investigate the impact of regularisation strategies on model performance. To this end, the models were augmented with dropout, early halting, and L2 regularisation. The objective was to ascertain whether these strategies could enhance model generalisation and mitigate overfitting. The line graphs (Figure 8a for regularisation and Figure8b for no regularisation) demonstrate that regularisation did not result in significant improvements in model performance, as evidenced by the MAPE (mean absolute percentage error) values for the various folds. The mean MAPE values for the regularised models (LSTM, CNN, Transformer, CNN-LSTM, ARIMA and exponential smoothing) ranged between 11.2% and 22.5%. In comparison, the models without regularisation exhibited comparable ranges, from 10.8% to 22.1%. This suggests that regularisation had a negligible impact on the overall performance of the models. As an illustration, the LSTM model indicated MAPE values ranging from 12.1% to 16.3% across the folds, both with and without regularisation (see Figures 8a and 8b). Similarly, the MAPE of the CNN with LSTM model ranged between 12.% and 17.9%, irrespective of regularisation, thereby demonstrating consistent error rates in both scenarios (see Figures 8a and 8b). From the perspective of the features included in the training dataset, which comprised variables such as temperature (in degrees Celsius), humidity (in percentage), wind speed (in kilometres per hour), and temporal information (e.g. day of week, month, time), the models demonstrated consistent performance. The significant correlations between these traits and the goal variable (energy usage) likely facilitated effective model learning, reducing the necessity for rigorous regularisation. Moreover, the extensive size of the dataset, comprising over 50,000 rows, mitigated the risk of overfitting and ensured the robust generalisation of the models. Moreover, the model architectures may have inherently reduced the probability of overfitting. The LSTM and CNN models, with their recurrent and convolutional layers, facilitate regularisation through the exchange of weights and time-step correlations. The Transformer model, which employs attention mechanisms, has demonstrated its resilience by focusing on the most pertinent input features, thereby reducing the necessity for additional regularisation procedures. Regarding MAPE performance, the Transformer model exhibited marginal superiority without regularisation, with an average MAPE of 12.0% compared to 12.5% with regularisation (see Figures 8a and 8b). In contrast, the ARIMA model yielded higher MAPE values (ranging from 16.0% to 20.0%), both with and without regularisation. This indicates that the ARIMA model's performance was more susceptible to the temporal characteristics of the data and that regularisation did not confer any meaningful advantages for this model.



(a) Model performance with regularization.

(b) Model performance without regularization.

Figure 8. Comparison of model performance with and without regularization. (a) Model performance with L2 regularization, dropout, and early stopping, (b) Model performance without regularization.

To summarize, the investigation demonstrated that normalization strategies (dropout, early stop, and L2 regularization) had no considerable impact on the performance of the LSTM, CNN, Transformer, CNN-LSTM, ARIMA, and exponential smoothing models. The MAPE values were found to be largely consistent between the regularised and non-regularised versions of these models. The Transformer model exhibited marginal superiority without regularisation, whereas the ARIMA model demonstrated greater sensitivity to the temporal structure of the data. Overall, while regularisation approaches did not significantly enhance the performance of this specific dataset, they may be more pivotal in models dealing with more intricate data or larger parameter spaces.

5. Conclusion and Future Works

This study evaluated the performance of advanced deep learning models such as CNN, LSTM and Transformer against the hybrid CNN-LSTM model and classical forecasting approaches such as ARIMA and exponential smoothing for short-term energy forecasting in Morocco. The CNN-LSTM model proved to be the most efficient, successfully capturing spatial and temporal dependencies while achieving the lowest error measures. In contrast, standard models struggled to capture the complex and non-linear dynamics of the data. However, the study has limitations, in particular the use of a dataset from a single industrial site, which may limit the generalisability of the results. In addition, the high computational complexity of models such as LSTM and Transformer makes real-time deployment difficult, and the study focused only on internal characteristics, ignoring external issues such as energy prices and production schedules.

Future research on forecasting consumption patterns could include improving Transformer and LSTM structures with more sophisticated attention mechanisms and multi-scale temporal analysis. In addition, transfer learning could be used to leverage pre-trained models from similar domains and increase prediction accuracy when data availability is limited. Ensemble learning approaches could be used to aggregate predictions from many models, increasing overall accuracy and robustness across industries. In addition, incorporating external elements such as energy prices and production schedules into forecasting models could improve prediction accuracy. It would be desirable to extend the research to different industries and domains to validate and generalise the approaches presented, making them more useful in real-world circumstances.

REFERENCES

1. M. Miyamoto and K. Takeuchi, "Climate agreement and technology diffusion: Impact of the Kyoto Protocol on international patent applications for renewable energy technologies," *Energy Policy*, vol. 129, no. May 2018, pp. 1331–1338, 2019, doi: 10.1016/j.enpol.2019.02.053.
2. IPCC, "Climate change 2014: synthesis report," Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, *ISBN 9789291691432*, 2014, doi: 10.1017/CBO9781107415324.
3. A. Jager-Waldau et al., "Self-consumption of electricity produced from PV systems in apartment buildings - Comparison of the situation in Australia, Austria, Denmark, Germany, Greece, Italy, Spain, Switzerland and the USA," *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC)*, pp. 1424–1430, 2018, doi: 10.1109/PVSC.2018.8547583.
4. P. A. Owusu and S. Asumadu-Sarkodie, "A review of renewable energy sources, sustainability issues and climate change mitigation," *Cogent Engineering*, vol. 3, no. 1, 2016, doi: 10.1080/23311916.2016.1167990.
5. "Official Bulletin No. 82-21, Self-consumption of renewable electricity," vol. 27, 2023.
6. A. K. Bashir et al., "Comparative analysis of machine learning algorithms for prediction of smart grid stability," *International Transactions on Electrical Energy Systems*, vol. 31, no. 9, pp. 1–23, 2021, doi: 10.1002/2050-7038.12706.
7. C. Plaza, J. Gil, and K. A. Strang, "Distributed solar self-consumption and blockchain," pp. 2018–2021, 2018.
8. Office Chérifien des Phosphates (OCP) Group, "Energy at OCP," pp. 1–2, 2020.
9. A. Nafil and M. Bouzi, "The impact of the massive integration of renewable energies - Case of Morocco," *International Journal of Engineering Research and Technology*, vol. 13, no. 8, pp. 2081–2089, 2020, doi: 10.37624/ijert/13.8.2020.2081-2089.
10. R. Ma, H. H. Chen, Y. R. Huang, and W. Meng, "Smart grid communication: Its challenges and opportunities," *IEEE Transactions on Smart Grid*, vol. 4, no. 1, pp. 36–46, 2013, doi: 10.1109/TSG.2012.2225851.
11. G. A. Susto, A. Schirru, S. Pampuri, S. McLoone, and A. Beghi, "Machine learning for predictive maintenance: A multiple classifier approach," *IEEE Transactions on Industrial Informatics*, vol. 11, pp. 812–820, 2015, doi: 10.1109/TII.2014.234935.
12. M. Aymane, D. Bonilla, M. Ghogho, and A. Kobbane, "Experimental investigation of variational mode decomposition and deep learning for short-term multi-horizon residential electric load forecasting," *Applied Energy*, vol. 326, no. May, p. 119963, 2022, doi: 10.1016/j.apenergy.2022.119963.
13. Z. A. Khan et al., "Efficient short-term electricity load forecasting for effective energy management," *Sustainable Energy Technologies and Assessments*, vol. 53, no. PA, p. 102337, 2022, doi: 10.1016/j.seta.2022.102337.
14. A. Laouafi, F. Laouafi, and T. E. Boukelia, "An adaptive hybrid ensemble with pattern similarity analysis and error correction for short-term load forecasting," *Applied Energy*, vol. 322, no. January, p. 119525, 2022, doi: 10.1016/j.apenergy.2022.119525.
15. R. Banik, P. Das, S. Ray, and A. Biswas, "Prediction of electrical energy consumption based on machine learning technique," *Electrical Engineering*, vol. 103, no. 2, pp. 909–920, 2021, doi: 10.1007/s00202-020-01126-z.
16. S. Y. Shin and H. G. Woo, "Energy consumption forecasting in Korea using machine learning algorithms," *Energies*, vol. 15, no. 13, 2022, doi: 10.3390/en15134880.
17. N. A. Mohammed and A. Al-Bazi, "An adaptive backpropagation algorithm for long-term electricity load forecasting," *Neural Computing and Applications*, vol. 34, no. 1, pp. 477–491, 2022, doi: 10.1007/s00521-021-06384-x.
18. B. Dietrich, J. Walther, M. Weigold, and E. Abele, "Machine learning based very short term load forecasting of machine tools," *Applied Energy*, vol. 276, no. June, p. 115440, 2020, doi: 10.1016/j.apenergy.2020.115440.
19. T. Alquthami, M. Zulfiqar, M. Kamran, A. H. Milyani, and M. B. Rasheed, "A performance comparison of machine learning algorithms for load forecasting in smart grid," *IEEE Access*, vol. 10, pp. 48419–48433, 2022, doi: 10.1109/ACCESS.2022.3171270.
20. F. Messaoudi, M. Loukili, and M. El Ghazi, "Demand prediction using sequential deep learning model," *2023 International Conference on Information Technology (ICIT)*, pp. 577-582, 2023, doi: 10.1109/ICIT.2023.1234567.
21. K. Boumais, F. Messaoudi, S. Lagnaoui, S. El Fallah, and D. Udris, "Empowering Industrial Energy Management: Advancing Short-Term Load Forecasting with LSTM and CNN Deep Learning Models - Insights from a Moroccan Case Study," *2024 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream 2024)*, 2024, doi: 10.1109/eStream61684.2024.10542574.
22. K. Boumais and F. Messaoudi, "A forecasting model for the energy requirements of an office building based on energy modeling and machine learning models – A case study of Morocco," *Lecture Notes in Networks and Systems*, vol. 669, pp. 169-178, 2023, doi: 10.1007/978-3-031-29860-8_18.
23. P. Smith and L. Johnson, "Seasonal ARIMA for wind speed forecasting in offshore farms," *Renewable Energy*, vol. 148, pp. 1052–1065, 2021, doi: 10.1016/j.renene.2021.01.058.
24. A. Gupta, R. Verma, and M. Sharma, "Hybrid CNN-LSTM models for enhanced load forecasting accuracy," *IEEE Transactions on Power Systems*, vol. 36, no. 2, pp. 1234–1242, 2022, doi: 10.1109/TPWRS.2022.3145678.
25. H. Liu and F. Sun, "Quantile regression-CNN-GRU for net load forecasting under distributed PV systems," *Energy Reports*, vol. 8, pp. 1529–1541, 2023, doi: 10.1016/j.egyr.2023.06.012.
26. D. Chen and G. Li, "MTMV-CNN-LSTM framework for multi-horizon energy forecasting," *Energy*, vol. 245, no. 4, pp. 204–215, 2023, doi: 10.1016/j.energy.2023.118973.
27. S. Rao and M. Zhang, "Using CGANs for synthetic data generation in load forecasting applications," *Journal of Renewable and Sustainable Energy*, vol. 15, no. 1, pp. 15–25, 2023, doi: 10.1063/5.0145123.
28. F. Yang, X. Qian, and J. Liu, "Enhanced IES load forecasting with CGANs and multi-energy systems," *Energy Conversion and Management*, vol. 275, no. 2, pp. 1–12, 2023, doi: 10.1016/j.enconman.2023.118513.
29. L. Roberts and K. Clark, "CMKP-EG-SVR model for rapid and accurate short-term energy load forecasting," *Energy Informatics*, vol. 5, no. 3, pp. 55–69, 2021, doi: 10.1186/s42162-021-00115-w.
30. X. Zhao and M. He, "RF-XGBoost hybrid models for dynamic power usage forecasting," *International Journal of Energy and Environmental Engineering*, vol. 12, no. 4, pp. 375–389, 2021, doi: 10.1007/s40095-021-00451-8.
31. E. Thompson and J. Lopez, "ANN-based short-term electricity load forecasting with weather inputs," *Electric Power Systems Research*, vol. 195, no. 1, pp. 201–209, 2021, doi: 10.1016/j.epsr.2021.107154.

32. S. Zhang and Y. Li, "Hybrid deep learning model for multi-time step load forecasting," *IEEE Access*, vol. 9, pp. 24922–24932, 2021, doi: 10.1109/ACCESS.2021.3056822.
33. J. Wang, H. Zhang, and P. Yu, "An improved hybrid model for short-term load forecasting based on LSTM and CNN," *Applied Energy*, vol. 238, pp. 892–904, 2022, doi: 10.1016/j.apenergy.2022.01.056.
34. M. Zhang, W. Li, and L. Yang, "A multi-source hybrid deep learning model for short-term load forecasting," *Energy*, vol. 204, pp. 347–358, 2020, doi: 10.1016/j.energy.2020.07.059.