



IMPROVING ENERGY GOVERNANCE IN THE CEMENT INDUSTRY: OPTUNA-BILSTM FRAMEWORK FOR LONG-RANGE ELECTRICAL LOAD FORECASTING

Mokhtaria Bekkaoui^{*1}, Hamed Benahmed², Meryem Kada Kloucha³ and Imane Nedjar⁴

^{1, 3, 4}First cycle department, École Supérieure en Sciences Appliquées de Tlemcen, Tlemcen, Algeria.

¹Manufacturing Engineering Laboratory of Tlemcen, University of Tlemcen, Algeria.

²Department of Computer Science, Faculty of Sciences, University of Tlemcen, Algeria.

³Laboratoire de Statistiques et Modélisations Aléatoires, Université de Tlemcen, Algeria.

⁴Biomedical Engineering Laboratory, University of Tlemcen, Algeria.

¹<http://orcid.org/0009-0007-7625-9106>, ²<http://orcid.org/0009-0007-0089-877X>,

³<http://orcid.org/0009-0003-1619-3046>, ⁴<http://orcid.org/0000-0002-5222-1193>

Email: *mokhtaria.bekkaoui@univ-tlemcen.dz, hamed.benahmed@univ-tlemcen.dz, meryem.kadakloucha@univ-tlemcen.dz, imane.nedjar@univ-tlemcen.dz

ARTICLE INFO

Article History

Received: March 9, 2026

Reviewed: April 9, 2026

Accepted: July 9, 2026

Published: August 31, 2026

Keywords:

Electricity consumption,
Forecasting,
Industrial energy modeling,
LSTM,
BiLSTM,
Optimization.

ABSTRACT

Precise long-range forecasting of industrial electrical demand constitutes a strategic imperative for energy governance, particularly within energy-intensive, continuous-process industries such as cement manufacturing. This investigation conduct a systematic evaluation of two advanced recurrent neural networks, Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM), fine-tuned via the Optuna optimization. Using realistic data from an operational cement plant, we quantify the trade-offs between predictive accuracy, computational overhead, and operational viability. Empirical results demonstrate that the Optuna-optimized BiLSTM framework achieves superior predictive performance, yielding an R2 score of 0.890 and a Mean Absolute Error (MAE) of 0.220. These findings delineate a modeling framework that reconciles high-fidelity forecasting with practical implementation, providing foundational guidelines for designing energy management systems that optimize maintenance, reduce costs, and advance sustainable operations.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

Prospective The global energy landscape is undergoing significant transformation due to rapid technological advancement, population growth, and expanding industrialization. Some alarming projections suggest that, unless the consumption of all devices is reduced, energy demand by 2040 will exceed the available energy supply [1]. This pressing challenge underscores the urgent need for accurate electricity consumption forecasting models, which are instrumental in enabling better energy resource management, cost reduction, and infrastructure planning [2]. Electricity demand forecasting has gained critical importance amid digital transformation, where Artificial Intelligence (AI) must address growing complexity in modern power systems [3], [4]. Traditional rule-based or statistical models, while historically useful, often struggle to capture the dynamic, non-linear patterns that characterize modern energy systems. As a result, the focus has shifted toward more advanced techniques capable of uncovering hidden patterns in sequential data. Time series forecasting plays a crucial role in anticipating future energy consumption by analyzing historical usage patterns. It is widely applied in modern energy systems for tasks such as electricity demand prediction, load management, and renewable energy integration. These forecasting techniques enable utilities and industries to optimize grid operation, reduce energy waste, and support data-driven decision-making in energy planning and maintenance. With the increasing complexity of smart grids and variable renewable energy sources, advanced artificial intelligence and deep learning models such as LSTM, BiLSTM, and hybrid neural networks have shown remarkable performance in capturing nonlinear temporal dependencies and improving forecast accuracy [5], [6].

Among these advancements, deep learning models, particularly Recurrent Neural Networks (RNNs) and their derivatives like LSTM and BiLSTM have gained attention for their ability to model complex temporal dependencies [7]. These architectures have demonstrated strong predictive performance in energy forecasting due to their capacity to capture long-term and nonlinear relationships. Nevertheless, they typically require substantial data volumes, high computational power, and fine-tuning of hyperparameters. In contrast, classical statistical models such as Autoregressive Integrated Moving Average (ARIMA) and Trend and Seasonal Components (TBATS) continue to be employed for their interpretability and efficiency, especially in low-data scenarios or when model transparency is critical [8]. While much of the existing literature, [9], [10] has focuses on residential energy forecasting, long-term electricity consumption prediction in industrial environments presents unique challenges and opportunities. Industrial facilities, unlike residential users, follow more consistent and structured consumption patterns that are closely aligned with production processes. In fact, advanced electricity forecasting provides a dual-function capability; beyond improving energy procurement and operational efficiency, it enables predictive maintenance strategies that constitute the initial safeguard against equipment failures capable of interrupting production. This dual capability is especially critical in industrial environments, where unplanned outages can escalate to costly production halts and irreversible equipment damage [4]. Therefore, robust and context-aware predictive models are essential to support energy efficiency, operational continuity, and sustainable industrial growth.

This study offers an in-depth investigation of electricity consumption forecasting methods in an industrial context, focusing particularly on a cement manufacturing plant, a sector recognized for its high energy requirements and repetitive production processes. The analysis used the predictive capabilities of advanced deep learning models, specifically LSTM and BiLSTM, which were optimized through the Optuna framework to enable automated and efficient hyperparameter tuning. This research aims to rigorously evaluate these models in terms of predictive accuracy, computational efficiency, and practical applicability when confronted with dataset industrial electricity consumption. In addition, the study highlights the potential of accurate forecasting to support preventive maintenance strategies, improve energy management practices, and strengthen decision-making in energy-intensive industries. By bridging advanced artificial intelligence methods with traditional statistical approaches, this work contributes to the development of intelligent forecasting systems that enhance operational reliability, reduce downtime risks, and foster sustainable energy practices in large-scale manufacturing environments. The rest of the paper is structured as follows. Section 2 provides an overview of state-of-the-art approaches to electricity consumption forecasting. Section 3 describes the adopted methodology. Section 4 presents the dataset, experimental results, and comparative discussion. Section 5 concludes the paper and highlights future research directions.

II. RELATED WORK

This section presents a review of recent methods, published from 2022 onwards, for forecasting electricity consumption. Various machine learning (ML) and deep learning (DL) techniques have been employed to capture the temporal and nonlinear characteristics of energy consumption patterns. Recent advances in deep learning have enabled the development of models capable of addressing complex prediction challenges in the energy domain. These models are particularly effective when large datasets are available and the input-output relationships exhibit high levels of complexity conditions commonly found in decentralized power systems. In such contexts, neural network architectures can be tailored to tackle multiple forecasting tasks efficiently. Recurrent architectures, especially LSTM networks, have emerged as powerful tools for capturing temporal patterns in electricity consumption data. For example, the study in [11] introduced a deep LSTM-based framework for short-term residential load forecasting, demonstrating its superiority over classical statistical techniques such as ARIMA. Similarly, [12], designed a sequence-to-sequence LSTM model that successfully leverages the sequential structure inherent in energy consumption time series. More recently, BiLSTM models have emerged as an enhanced alternative, leveraging both past and future context [13]. In a comparative study given by [7], BiLSTM provided better generalization on volatile load profiles compared to unidirectional LSTM. However, the study also highlighted the sensitivity of BiLSTM models to hyperparameter configurations. To address this issue, researchers have turned to automated hyperparameter optimization tools, such as Optuna. Applied Optuna for fine-tuning the learning rate, batch size, and network depth of a BiLSTM model used for smart grid forecasting. The optimized model showed a significant reduction in Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE), underlining the value of meta-optimization strategies in deep learning applications [14], [15]. The quality assessment of the studies involved applying inclusion and exclusion criteria, ensuring the identification of experimental research that included variables directly related to the research question. Data extraction was carried out by analyzing the results of the selected approaches. To summarize and synthesize the results, experimental studies were identified using specific keywords relevant to the research question. A table was then prepared, which contained a synthesis of the selected articles, including the publication year, the models used, the periods covered, and the dataset.

Table 1: Present the summary of the main articles related to the present work.

Autors	Methods	Dataset
[16]	Deep-learning hybrid forecasting model, namely: EMD-SVR-PSO-AR-GARCH	New South Wales (NSW, Australia) market
[17]	Ensemble learning Compared: gradient boosted regression trees, random forests and an adaptation of Adaboost.	office building located in the campus of ISEP/GECAD(building is occupied in daily basis by about 30 people)
[9]	Combines three models: LSTM, GRU, and Temporal Convolutional Network (TCN)	2000 clients from Bejaia, Algeria
[18]	XGBoost classification model, Multi-Layer Perceptron (MLP) model that is used to make the final electricity consumption prediction.	500 residential users placed in a southeastern region of Spain.
[8]	Evaluated the ARIMA and TBATS statistical prediction methods and compared them with neural forecasting methods	Smart house that is equipped with solar panels and batteries for energy storage.

[19]	Hybrid multi-algorithm method. called ICMD-ANN-EDLSTM.	Descriptive statistic of daily electricity demand (MWh) at four substation South-east Queensland, Australia.
[20]	Improved SSA and multivariable input RNNs to achieve adaptive decomposition of raw time series data and fusion input of multiple features, ADMIF.	Moroccan buildings' electricity consumption dataset (MORED).
[10]	Time series dense encoder (TiDE) model.	City level, of residential customers and the technological consumption of electricity suppliers.
[21]	Modeling spatiotemporal series using Graph Neural Networks (GNNs), with a focus on the structure of decentralized electrical networks.	12 regions of French mainland– excluding Corsica.

Source: Authors, (2026).

III. METHODOLOGY

This section introduces the dataset utilized in our experiments and outlines the baseline models, specifically LSTM and BiLSTM architectures. Additionally, we detail the performance evaluation metrics applied, which include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative MAE, and the Coefficient of Determination (R^2).

III.1 DATASET DESCRIPTION

The empirical analysis utilizes a proprietary dataset derived from a cement manufacturing plant located in the western region of Algeria, classified as a high-voltage consumer. The facility receives its electrical supply at 60 kV from the national transmission operator via a dedicated 220/60 kV substation. The dataset comprises detailed electricity consumption records collected over a five-month period at 10-minute intervals. It reflects real operational conditions, including variations in production cycles, equipment startup and shutdown phases, and environmental influences. The temporal resolution and industrial context of the data make it highly valuable for evaluating the performance of forecasting models in scenarios where energy demand is both high and tightly coupled to continuous manufacturing processes.

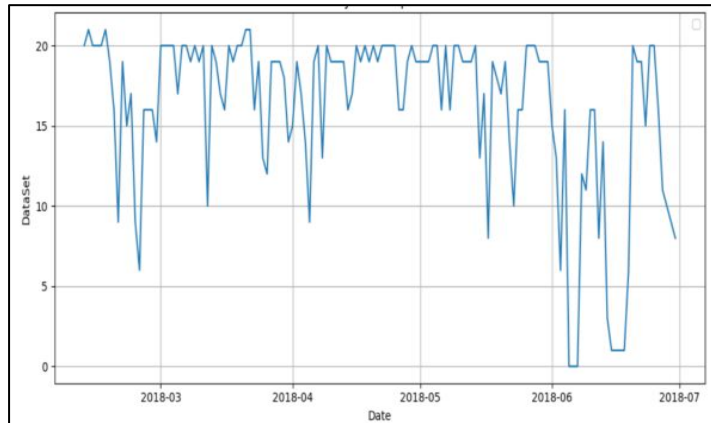


Figure 1: Electricity consumption curve.

Source: Authors, (2026).

III.2 MODELS BASELINE

In this study, both Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM) models are evaluated as baseline approaches, with and without Optuna-based optimization. The following paragraph describes the differences between the LSTM and BiLSTM architectures. LSTM: The Long Short-Term Memory (LSTM) network, is a type of Recurrent Neural Network (RNN) designed to capture long-term dependencies in sequential data by incorporating specialized memory units [22]. These memory units consist of a cell state and three gates (input, forget, and output) that regulate the flow of information, enabling the network to retain or discard information over time. This architecture is particularly effective in modeling non-linear and temporal relationships in time series data, such as electricity consumption patterns. BiLSTM: the Bidirectional LSTM (BiLSTM) extends the LSTM architecture by processing the input sequence in both forward and backward directions [23]. This dual-path structure enables the model to capture dependencies not only from past to future but also from future to past within the sequence, thereby improving the model's ability to learn complex temporal dynamics. Both architectures were implemented with a sequence of layers, including input normalization, one or more LSTM/BiLSTM layers, dropout regularization to prevent overfitting, and a fully connected dense layer for output generation. Hyperparameter tuning, such as the number of hidden units, learning rate, batch size, and sequence length, was performed using the Optuna optimization framework to maximize forecasting performance.

III.3 EVALUATION METRICS

The models were evaluated using the different metrics (see Table 2). These metrics were computed for each tested model LSTM, BiLSTM, LSTM (Optuna) and BiLSTM (Optuna) to allow a fair and multidimensional comparison of model accuracy, robustness, and generalization capability on the cement factory dataset.

Table 2: Evaluation Metric.

Metric	Equation	Description
Mean Absolute Error (MAE)	$MAE = \frac{1}{n} \sum_{i=1}^n X_i - \hat{X}_i $ (1)	Average absolute difference between predicted and actual values. Penalizes all errors equally.
Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \hat{X}_i)^2}$ (2)	Square root of the average squared errors. Heavily penalizes large deviations.
Relative MAE	$RMAE = \frac{\frac{1}{n} \sum_{i=1}^n X_i - \hat{X}_i }{\frac{1}{n} \sum_{i=1}^n X_i}$ (3)	Is a normalized version of the Mean Absolute Error (MAE).
Coefficient of Determination (R ²)	$R^2 = 1 - \frac{\sum_{i=1}^n (X_i - \hat{X}_i)^2}{\sum_{i=1}^n (X_i - \bar{X})^2}$ (4)	Proportion of variance in the dependent variable explained by the model. Ranges from 0 (no fit) to 1 (perfect fit).
Training Time	TrainT (measured in seconds/minutes)	Time required to train the model on historical data. Key for scalability.

Source: Authors, (2026).

IV. RESULTS AND DISCUSSION

This section presents the results of our research alongside an in-depth discussion of the experimental findings, highlighting the implementation choices, methodological steps, and comparative insights. The entire experiment was conducted on the Google Colab environment to leverage cloud-based computational resources. The architectures were implemented using the TensorFlow framework and the Keras library in Python to ensure a scalable and efficient training environment. For both the LSTM and BiLSTM models, a temporal window size of 72 was utilized, representing a 48-hour historical sequence, to capture the intricate short-term dependencies inherent in industrial electricity consumption. While initial training employed the Adam optimizer [24] with a MAE loss function, the performance of both models was systematically enhanced through the integration of Optuna, a Bayesian hyperparameter optimization framework. This optimization phase was pivotal in identifying the ideal configuration for critical parameters, including learning rates, hidden units, and dropout rates, thereby maximizing predictive accuracy and ensuring robust generalization across both neural architectures. The following section presents a detailed comparison of the results from the four methods, namely LSTM and BiLSTM, with and without Optuna optimization, in terms of predictive accuracy, computational complexity, and their applicability to industrial electricity forecasting scenarios.

IV.1 TRAINING AND VALIDATION RESULT

The training and validation loss patterns of the LSTM and BiLSTM architectures, including both baseline and Optuna-optimized configurations, are evaluated. The observed loss trends provide insights into each model's convergence behavior and generalization capability. Figure 2 shows the training and validation loss evolution of the baseline model LSTM which exhibits a rapid decrease in training loss, stabilizing around 0.002, while the validation loss remains higher and fluctuates between approximately 0.0025 and 0.005, indicating overfitting and limited generalization. In contrast, in Figure 3, the BiLSTM model demonstrates more stable convergence, with reduced variance in validation loss and a smaller gap between training and validation curves, reflecting improved generalization performance. The improved behavior of the BiLSTM can be attributed to its bidirectional architecture, which captures temporal dependencies in both forward and backward directions and provides richer contextual representations. Figures 4 and 5 illustrate the performance of the LSTM and BiLSTM architectures following Optuna-driven hyperparameter optimization. The optimized models exhibit accelerated convergence and significantly smoother loss trajectories, with the final loss values stabilizing at approximately 0.017. Furthermore, the minimal divergence between the training and validation loss curves suggests a substantial improvement in model robustness and a heightened resistance to overfitting. Overall, these results confirm that while BiLSTM architectures outperform standard LSTM models, Optuna-based hyperparameter optimization further improves convergence stability and predictive accuracy, highlighting the importance of systematic hyperparameter tuning for sequence-based learning tasks.

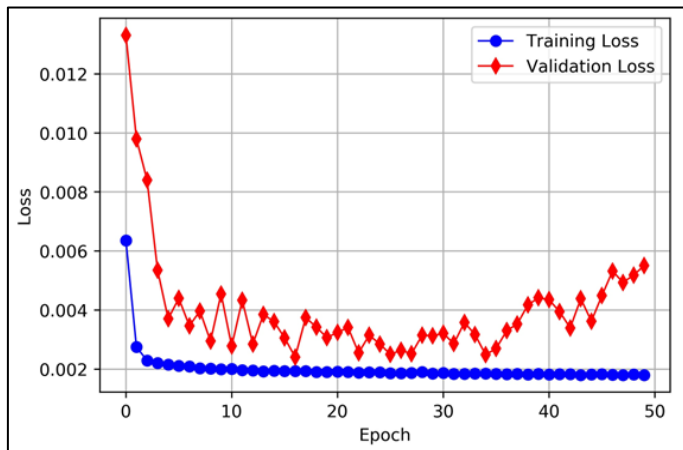


Figure 2: Loss of LSTM-based models vs epochs.
Source: Authors, (2026).

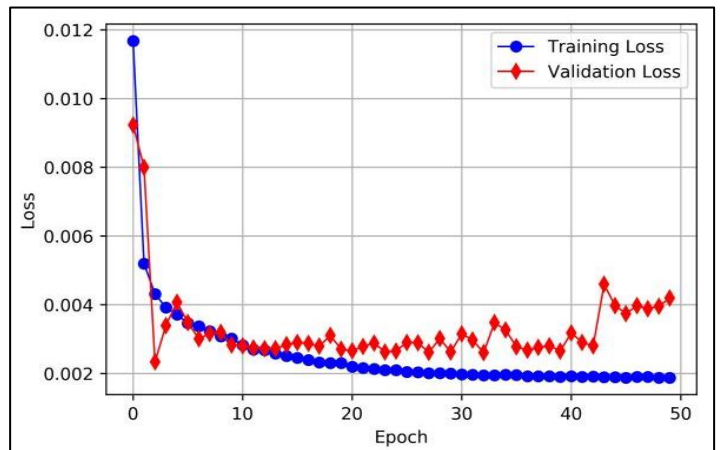


Figure 3: BiLSTM-based models vs epochs.
Source: Authors, (2026).

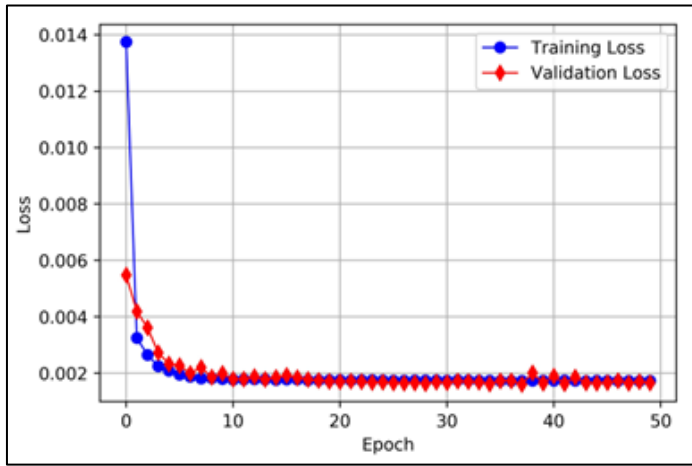


Figure 4: Loss of Optuna-Optimized LSTM .
Source: Authors, (2026).

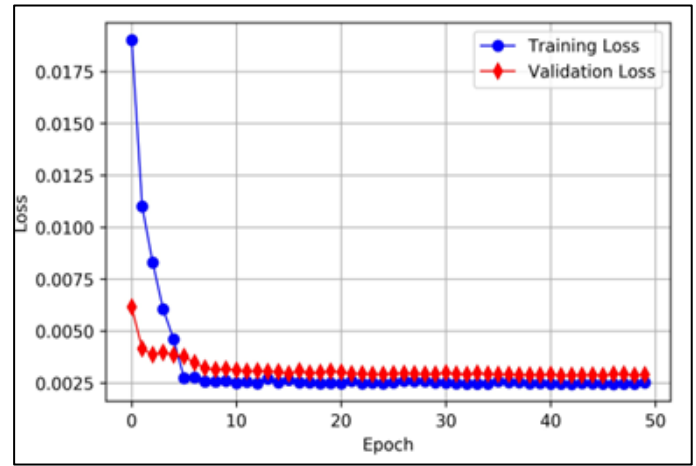


Figure 5: Loss of Optuna-Optimized BiLSTM-based models vs epochs.
Source: Authors, (2026).

IV.2 EVALUATION TEST AND COMPARATIVE STUDY

This section presents the experimental evaluation and comparative analysis of the four models: the LSTM, BiLSTM, Optuna-optimized LSTM and Optuna-optimized BiLSTM. All models were evaluated on the test dataset using standard metrics, including Root Mean Square Error (RMSE), Relative Mean Absolute Error (RMAE), R^2 score, and training time, providing a comprehensive assessment of predictive accuracy and computational efficiency. Table 3 summarizes the performance of LSTM and BiLSTM models with and without Optuna-based hyperparameter optimization. The results demonstrate that Optuna optimization significantly improves predictive accuracy, as reflected in lower RMSE and RMAE values and higher R^2 scores. Among all configurations, the Optuna-optimized BiLSTM achieves the best performance, balancing high predictive accuracy with moderate training time. Figure 6 compares the real and predicted energy consumption values for the two best-performing models. Both the Optuna-optimized LSTM and BiLSTM closely follow the real consumption trends, with the BiLSTM predictions showing slightly higher alignment and smoother tracking, highlighting its superior ability to capture temporal dependencies and improve forecasting accuracy.

Table 3: Comparative Performance of Baseline and Optuna-Optimized LSTM/BiLSTM.

Prediction Model	RMSE	RMAE (%)	R^2 Score	Training Time (TT)
LSTM	0.225	56	-0.002	4 min 57 s
BiLSTM	0.230	42	-0.001	8 min 19 s
Optuna-Optimized LSTM	12.08	24	0.665	5 min 15 s
Optuna-Optimized BiLSTM	0.220	20	0.890	6 min 05 s

Source: Authors, (2026).

For the LSTM and BiLSTM models, the window size and number of epochs were manually chosen. Several combinations were tested, and the best-performing configuration was retained for each model. Specifically, the LSTM model achieved its best results, yielding an RMSE of 0.225, an RMAE of 56%, and an R^2 score of -0.002, with a training time of 4 minutes 57 seconds. The BiLSTM model, achieved similar results (RMSE = 0.230, RMAE = 42%, R^2 = -0.001) with a significantly lower training time of 8 minutes 19 seconds. The Optimized-Optuna BiLSTM model achieved the best performance overall, with an RMSE of 0.220, an RMAE of 20%, and a training time of just 6 minutes 05 seconds, clearly demonstrating the advantage of automated tuning in improving both accuracy and efficiency.

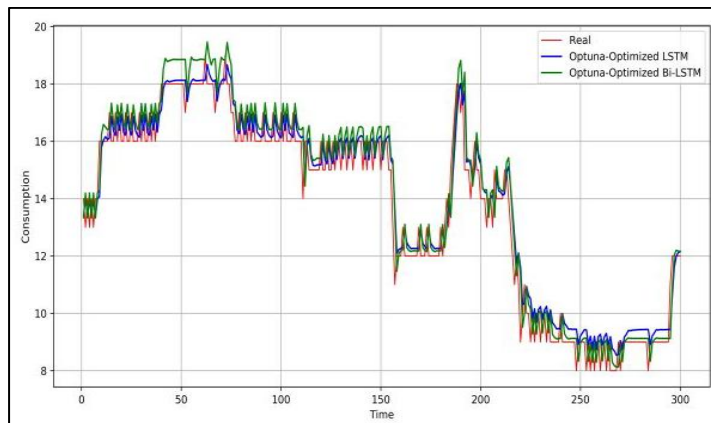


Figure 6: Real Vs Predicted Consumption of Optuna-Optimized LSTM/BiLSTM.
Source: Authors, (2026).

The visualization in Figure 6 enables a direct comparison of how the optimized models handle the test set data. Both models generally capture the overall trend, successfully tracking major peaks and the general movement of real consumption, particularly during periods of higher usage (time steps 0–200). However, the Optuna-Optimized BiLSTM consistently demonstrates superior performance across the entire time series. Its predictions remain closer to the actual consumption, indicating lower overall error. This advantage is most pronounced during periods of high volatility (time steps 200–270), where real consumption dips sharply from approximately 15 to 8. The BiLSTM accurately follows this decline and subsequent low-consumption plateau, reflecting its ability to leverage bidirectional temporal dependencies. In contrast, the Optuna-Optimized LSTM exhibits clear limitations in the low-consumption segment. It significantly underestimates the actual values during the dip, highlighting the unidirectional LSTM reliance solely on past trends, which reduces its effectiveness in capturing abrupt changes and maintaining accuracy during volatile periods. These results indicate that for this consumption forecasting task, access to both past and future temporal context is highly informative. The Optuna-Optimized BiLSTM emerges as the preferred model, as its bidirectional memory structure enables more robust and accurate predictions across varying consumption levels and volatility.

V. CONCLUSIONS

This study presented a comprehensive analysis of advanced RNN architectures, LSTM, BiLSTM, enhanced with Optuna-based hyperparameter optimization, the context of long-term electricity consumption forecasting for a cement manufacturing facility. Through rigorous experimentation on real-world industrial data, with Optuna-Optimized BiLSTM achieving the best overall results across all evaluation metrics. The automatic parameter selection provided by Optuna not only improved prediction quality but also reduced training time, making the model more suitable for practical deployment in energy-intensive environments. These findings reinforce the potential of optimized deep learning methods to serve as reliable tools in smart energy management systems, enabling predictive maintenance, cost savings, and more sustainable production planning in industrial settings.

VI. AUTHOR'S CONTRIBUTION

Conceptualization: Mokhtaria Bekkaoui.

Methodology: Mokhtaria Bekkaoui and Hamed Benahmed.

Investigation: Meryem Kada Kloucha and Imane Nedjar.

Discussion of results: Mokhtaria Bekkaoui, Hamed Benahmed and Imane Nedjar.

Writing – Original Draft: Mokhtaria Bekkaoui.

Writing – Review and Editing: Mokhtaria Bekkaoui, Hamed Benahmed and Imane Nedjar.

Resources: Meryem Kada Kloucha.

Supervision: Meryem Kada Kloucha and Imane Nedjar.

Approval of the final text: Hamed Benahmed and Imane Nedjar.

VII. REFERENCES

- [1] S. Ghasemian, A. Faridzad, P. Abbaszadeh, A. Taklif, A. Ghasemi, and R. Hafezi, "An overview of global energy scenarios by 2040: Identifying the driving forces using cross-impact analysis method", *Int. J. Environ. Sci. Technol.*, vol. 21, no. 11, pp. 7749-7772, 2024.
- [2] International Energy Agency, "World Energy Outlook 2025", Paris, France: IEA Publications, 2025. [Online]. Available: <https://www.iea.org/events/world-energy-outlook-2025>
- [3] A. Román-Portabales, M. López-Nores, and J. J. Pazos-Arias, "Systematic review of electricity demand forecast using ANN-based machine learning algorithms", *Sensors*, vol. 21, no. 13, p. 4544, 2021.
- [4] International Energy Agency, "World Energy Outlook 2022", Paris, France: IEA Publications, 2022. [Online]. Available: <https://iea.blob.core.windows.net/assets/830fe099-5530-48f2-a7c1-11f35d510983/WorldEnergyOutlook2022.pdf>
- [5] 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", *Renew. Sustain. Energy Rev.*, vol. 144, p. 110992, 2021.
- [6] C. Cai, Y. Tao, T. Zhu, and Z. Deng, "Short-term load forecasting based on deep learning bidirectional LSTM neural network", *Appl. Sci.*, vol. 11, no. 17, 2021.
- [7] D. Guimarães Da Silva, A. Alvarenga, M. Meneses, and D. R. Vera Paz, "Comparing LSTM and BLSTM deep neural networks for power consumption prediction: Preliminary studies", *Energy Rep.*, vol. 10, pp. 3315-3334, 2023.
- [8] A. Gellert, U. Fiore, A. Florea, R. Chis, and F. Palmieri, "Forecasting electricity consumption and production in smart homes through statistical methods", *Sustain. Cities Soc.*, vol. 76, p. 103426, 2022.
- [9] D. Hadjout, J. F. Torres, A. Troncoso, A. Sebaa, and F. Martínez-Álvarez, "Electricity consumption forecasting based on ensemble deep learning with application to the Algerian market", *Energy*, vol. 243, 2022.
- [10] D. Pătelează et al., "Electricity consumption forecasting for sustainable smart cities using machine learning methods", *Internet Things*, vol. 27, p. 101322, 2024.
- [11] A. Muneer, R. F. Ali, A. Almaghthawi, S. M. Taib, A. Alghamdi, and E. A. A. Ghaleb, "Short term residential load forecasting using long short-term memory recurrent neural network", *Int. J. Electr. Comput. Eng.*, vol. 12, no. 5, pp. 5589-5599, Oct. 2022.
- [12] Y. Mu, M. Wang, X. Zheng, and H. Gao, "An improved LSTM-Seq2Seq-based forecasting method for electricity load", *Front. Energy Res.*, vol. 10, 2023.
- [13] N. Mughees, S. A. Mohsin, A. Mughees, and A. Mughees, "Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting", *Expert Syst. Appl.*, vol. 175, p. 114844, 2021.

- [14] M. Feurer and F. Hutter, “Hyperparameter optimization”, in *Automated Machine Learning*, pp. 3-33, 2019.
- [15] N. Burkart and M. F. Huber, “A survey on the explainability of supervised machine learning”, *J. Artif. Intell. Res.*, vol. 70, pp. 245-317, 2021.
- [16] G. F. Fan, X. Wei, Y. T. Li, and W. C. Hong, “Forecasting electricity consumption using a novel hybrid model”, *Sustain. Cities Soc.*, vol. 61, p. 102320, 2020.
- [17] T. Pinto, I. Praça, Z. Vale, and J. Silva, “Ensemble learning for electricity consumption forecasting in office buildings”, *Neurocomputing*, vol. 423, pp. 747-755, 2021.
- [18] F. Lazzari et al., “User behaviour models to forecast electricity consumption of residential customers based on smart metering data”, *Energy Rep.*, vol. 8, pp. 3680-3691, 2022.
- [19] S. Ghimire, R. C. Deo, D. Casillas-Pérez, and S. Salcedo-Sanz, “Efficient daily electricity demand prediction with hybrid deep-learning multi-algorithm approach”, *Energy Convers. Manag.*, vol. 297, p. 117707, 2023.
- [20] L. Fang and B. He, “A deep learning framework using multi-feature fusion recurrent neural networks for energy consumption forecasting”, *Appl. Energy*, vol. 348, p. 121563, 2023.
- [21] E. Campagne, Y. Amara-Ouali, Y. Goude, and A. Kalogeratos, “Leveraging graph neural networks to forecast electricity consumption”, *arXiv preprint*, 2024. [Online]. Available: <http://arxiv.org/abs/2408.17366>
- [22] S. Hochreiter and J. Schmidhuber, “Long short-term memory”, *Neural Comput.*, vol. 9, no. 8, pp. 1735-1780, Nov. 1997.
- [23] S. Siami-Namini, N. Tavakoli, and A. S. Namin, “The performance of LSTM and BiLSTM in forecasting time series”, in *Proc. IEEE Int. Conf. Big Data*, 2019, pp. 3285-3292.
- [24] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization”, in *Proc. Int. Conf. Learn. Representations (ICLR)*, 2017. [Online]. Available: <https://hdl.handle.net/11245/1.505367>