Exploring Load Forecasting: Bridging Statistical Methods to Deep Learning Techniques in Real-World Environment

Ioannis Makris[†], Nikolaos Moschos[†], Panagiotis Radoglou-Grammatikis[¶‖], Nikolaos Andriopoulos[‡], Nikolaos Tzanis[‡], Kyriaki-Nefeli Malamaki[‡], Maria Fotopoulou[§], Konstantinos Kaousias[¶], George Lampsidis Tompros[¶], and Panagiotis Sarigiannidis[¶†]

***Abstract*—Load forecasting has a significant impact on energy management and planning, facilitating efficient allocation of resources and grid operations. In this study, a comparative analysis of traditional statistical methods and deep learning techniques is conducted utilizing a real-world dataset from the Ikaria islanded grid. This paper focuses on four different forecasting approaches: Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), Long Short-Term Memory (LSTM) networks, and Deep Neural Networks (DNN). Through the appropriate processing of the data, extensive experimentation took place, aiming to capture the complex and nonlinear patterns of the dataset. The results indicated that LSTM and DNN outperformed both ARIMA and SARIMAX in all three evaluation metrics, achieving 0.13, 0.09, and 2.11%, RMSE, MAE, and MAPE, respectively. As a result, this study validates the superiority of deep learning techniques in real-world islanded grid environments being capable of accurately predicting future load values based on historical data.**

***Index Terms*—Deep Neural Networks, Electricity Demand Prediction, Load Forecasting, Long Short-term Memory, Time Series Analysis**

I. Introduction

Electrical load forecasting plays a crucial role in the operation of power systems. The ability to predict system load accurately enables utility companies to make well-informed decisions regarding generation, distribution, and maintenance, thus ensuring the reliable, cost-efficient, and sustainable supply of electricity [1]. Traditionally, utility companies employed statistical methods and time series analysis to perform load forecasting. Nevertheless, these methods, even effective, do not capture the complex, non-linear relationships and intrinsic variabilities of electricity demand patterns, present especially at smaller scale electricity grids [2], [3].

The emergence of smart grids and the proliferation of high-resolution and high-volume data stemming from the plethora of smart meters across the grid led to the wide adoption of Machine Learning (ML) techniques for load forecasting [4], [5]. These techniques offer the possibility to model electricity demand patterns more accurately by learning from historical data, considering multiple factors, and adapting to changing patterns over time [6]. However, despite significant advancements in these models, the utilization of ML in load forecasting raises many challenges. These include handling the high dimensionality of input data, accounting for temporal and spatial dependencies, and ensuring the replicability of forecasting models across different occasions and scales.

This work presents a thorough comparative analysis between traditional statistical methods such as Autoregressive Integrated Moving Average (ARIMA), Seasonal ARIMA with Exogenous Variables (SARIMAX) against modern deep learning techniques, namely Long Short-Term Memory (LSTM) networks, and Deep Neural Networks (DNNs). This comparison offers insights into their competencies and limitations in terms of forecasting accuracy, computational efficiency, and applicability across different forecasting horizons. Moreover, it showcases the advantages and trade-offs associated with each method, providing a deep understanding of how these techniques can be applied to real-world scenarios, such as the one tested in this work, focusing on a challenging real-world dataset of an Islanded grid. The results of this study benchmark the predictive performance of these models and highlight their operational implications in the energy sector.

The contributions of this work can be summarised as follows:

1) Deep Learning techniques can be utilized in the context of real-world environmental settings

This research was funded by the European Union's Horizon Europe Innovation Action, SINNOGENES (STORAGE INNOVATIONS FOR GREEN ENERGY SYSTEMS), under grant agreement No. 101096992. Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

[†]I. Makris, Nikolaos Moschos, and P. Sarigiannidis are with MetaMind Innovations, Kozani, Greece - E-Mail: {imakris, nmoschos, psarigiannidis}@metamind.gr

[‡] N. Andriopoulos, N. Tzanis, and K. N. Malamaki are with the Independent Power Transmission Operator, Athens, Greece - E-Mail: {k.andriopoulos, n.tzanis, k.malamaki}@admie.gr

[‡] M. Fotopoulou is with the Centre for Research and Technology Hellas (CERTH), Athens, Greece - E-Mail: fotopoulou@certh.gr

[¶] K. Kaousias, and G. Lampsidis Tompros are with the Hellenic Electricity Distribution Network Operator (H.E.D.N.O.), Athens, Greece - E-Mail: {k.kaousias, g.lampsidis}@deddie.gr

[¶] P. Radoglou-Grammatikis and P. Sarigiannidis are with the Department of Electrical and Computer Engineering, University of Western Macedonia, Kozani, Greece - E-Mail: {pradoglou, psarigiannidis@uowm}.gr

[‖] P. Radoglou-Grammatikis is with K3Y Ltd, Sofia, Bulgaria - E-Mail: pradoglou@k3y.bg

2) Complex forecasting models can be integrated as part of a digital twin of a real system, as they are capable of predicting future values based on limited data, and, thus, they accurately represent actual forecasts
3) Nonlinear patterns that are present in current energy consumption can be captured by deep learning techniques

The remainder of the paper is organized as follows: A relevant literature review is presented in II so as to highlight the contributions of this work. In Section III, the methodological framework and the corresponding mathematical analysis of the forecasting methods are presented, the experimental evaluation, describing the setup, the dataset, and the experimental results in Section IV. Finally, Section V concludes the paper with the main findings and proposes directions for future research.

II. Relevant Literature

Various kinds of forecasting techniques for energy and load forecasting in power networks are analyzed in [7], highlighting the benefits of Artificial Neural Networks (ANN), Machine Learning (ML), time series analysis, such as Autoregressive Integrated Moving Average (ARIMA) and probabilistic forecasting or even the combination of the aforementioned techniques. In this context, it was noted that the forecast of small-scale systems, e.g., a building or a low-voltage microgrid, was less accurate than the forecast of large systems, where the load/generation was aggregated. Furthermore, the accuracy of a forecasting technique was reduced when the time horizon was expanded, meaning that forecasts of the demand or generation within the next hour of the day tend to be more accurate than those of the next/last time steps.

The authors of [8] perform demand forecasting in the large-scale power system of Ukraine, exploiting the advantages of a hybrid classical statistical and ML algorithm. The proposed solution is particularly effective in long-term demand forecasts, on an hourly resolution, showing up to 96.83% accuracy. The authors of [9] aim to forecast a single household's daily electricity consumption profile. For this purpose the authors compare ANN and ARIMA and comment on the seasonality and the difficulties of forecasting such small scale load accurately, highlighting that the average error of forecasting for both the models was smaller in winter than in summer and also suggesting that future research focuses on a hybrid model based on both ARIMA and ANN models.

The authors of [10] assess the performance of physics-informed ML models for wind turbine and photovoltaic power forecasting. According to the results, the performance of hybrid CNN-LSTM and Random Forest (RF) is considered to be better than others when it comes to five-minute and hourly resolution. Particularly for the wind turbine forecast with an hourly resolution, the hybrid CNN-LSTM model had the lowest RMSE, equal to 24.77%, and the lowest MAPE, equal to 28.01%. The same applies to the photovoltaic forecast with hourly resolution, where the respective values are equal to 6.11% and 10.94%, resulting in a lower error in photovoltaic-related forecasts is expected since irradiation is more predictable than wind speed.

Following another approach, in [11] the Theta method is proposed. This method is less complex than the aforementioned ones and is considered to be similar to Simple Exponential Smoothing (SES), which is the simplest of the exponentially smoothing methods as it only incorporates one parameter. Despite its simplicity, the Theta method became popular for exceeding the rest of its competitors in the M3 competition and is, therefore, suitable for demand, generation, or other time series forecasting.

In [12] a novel hybrid forecasting model is introduced combining Grey Wolf Optimization (GWO), Convolutional Neural Networks (CNN), and Bidirectional Long Short-Term Memory (BiLSTM) for predicting building electricity consumption with high accuracy and stability. Utilizing one-dimensional CNN with BiLSTM for nonlinear feature extraction from time series data, the performance was tested on datasets from buildings with diverse characteristics, outperforming conventional forecasting models. In [13] a hybrid forecasting model is proposed that integrates Residual Neural Network (ResNet) with LSTM to enhance short-term load forecasting accuracy by leveraging load time series characteristics. By reconstructing data with multiple features for ResNet-based feature extraction followed by LSTM for forecasting and considering weather variables prediction, the model demonstrates superior prediction accuracy over conventional methods.

In [14] the authors propose a method based on statistical analysis to pre-process the time series data before solving the problem of short-term load prediction. The main feature of this method is to optimize the super-parameters of a neural network by using the statistical attributes of each time series data set and transform the given data set into a form that allows maximum advantage of the CNN algorithm. In [15], the short-term load forecasting is investigated by comparing various stochastic and deterministic methods. Utilizing historical load data over two years, the study identifies the CNN-LSTM hybrid model as the most accurate.

A combined probabilistic forecasting method was proposed in [16]. Authors introduce a Combined Probabilistic Forecasting Model (CPFM) that improves traditional statistical and quantile regression machine learning models. By employing an improved multi-objective optimizer, the CPFM forecast accuracy is superior. The effectiveness of the CPFM is tested upon a case study using ISO New England data, outperforming 13 other models in comparative analysis. In[17], an innovative short-term load forecasting is suggested leveraging wild horse optimization method for feature extraction, combined with deep learning (WHODL-STLFS). Experimental results indicate that the proposed technique achieves high accuracy in load forecasting. In [18], the authors propose a tuned LSTM model for multivariate time-series forecasting of electricity load, utilizing an open European dataset for benchmarking. The study demonstrates a significant improvement in forecasting accuracy through parameter tuning, comparing the performance of traditional grid search methods.

Finally in [19] authors present a unified machine learning framework designed for simultaneous real-time electrical load forecasting and unsupervised anomaly detection. The proposed approach optimizes forecasting by adapting based on past performance and identifies anomalies by comparing current and historical load fluctuations, addressing class imbalance issues common in anomaly detection. The method was evaluated on a complex dataset, and demonstrated superior performance in both forecasting and anomaly detection, showcasing its effectiveness and potential for practical application in smart grid management.

III. Methodology

This paper employs a methodology that is designed to thoroughly investigate and provide a comparison of the performance of several forecasting techniques in the domain of load forecasting. Based on the criticality of load forecasting in overall energy management and planning, accurate prediction of future energy consumption is vital in ensuring optimal resource allocation and operational efficiency. This section outlines and describes the methodological framework for the evaluation of four (4) forecasting techniques, namely Autoregressive Integrated Moving Average (ARIMA), Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX), Long Short-Term Memory (LSTM) Networks, and Deep Neural Networks (DNNs). Each of the aforementioned methods represents a different approach to modelling and forecasting load demand, ranging from traditional statistical analysis techniques to Machine Learning/Deep Learning (ML/DL) methods. Through this comparative analysis, this paper aims to reveal the strengths, limitations, and overall performance of these methods in identifying temporal patterns present in load data.

A. ARIMA

ARIMA is a widely known statistical method for time series forecasting and analysis. In particular, this technique consists of three (3) key components:

1) Autoregression (AR) component, which is responsible for modelling the relationship between an observation and a number of lagged observations (previous time stamps). AR assumes that the value of a variable at a specific given time linearly depends on its own previous values.
2) Differencing (I) component, which is utilized to point out the differencing of raw observations to make the time series stationary. Stationarity ensures that statistical characteristics such as mean, variance, and autocorrelation remain constant through time.
3) Moving Average (MA) component, which models the relationship between an observation and an error from a moving average model which is applied to lagged observations, and captures short-term, and potentially random variations in the data.

An ARIMA model is typically expressed as *ARIMA(p, d, q)*, where:

- p, represents the order of the AR component,
- d, represents the degree of Differencing component to achieve stationarity to the time series data,
- q, represents the order of the MA component.

The mathematical expression of ARIMA is:

$$\begin{aligned} Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} \\ + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t \end{aligned}$$

Where:

- Y_t is the value of the time series at time t.
- μ is the constant (mean).
- $\phi_1, \phi_2, \ldots, \phi_p$ are the parameters of the autoregressive (AR) part of the model, where p represents the order of the autoregressive part.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the parameters of the moving average (MA) part of the model, where q represents the order of the moving average part.
- ε_t is the error term at time t.

ARIMA technique gained popularity due to its ability to capture short-term and long-term trends in time series data, making it suitable for a wide range of forecasting tasks. However, its strong dependency on linear relationships and its sensitivity to outliers, significantly reduce its overall performance.

B. SARIMAX

SARIMAX extends the ARIMA model by incorporating seasonal components and exogenous variables into the forecasting processes. This technique is particularly useful in cases where seasonal patterns and external factors influence the corresponding time series data. SARIMAX consists of the following components:

1) Seasonal (S) component, which is responsible for capturing seasonal patterns in time series data, and, simultaneously, introducing additional parameters to model seasonal variations that may exist in fixed periods,
2) ARIMA model, which included the autoregressive, differencing, and moving average components, as described in the previous subsection,
3) Exogenous Variables (X), also known as external regressors, are factors that possibly influence the time series, and their inclusion can potentially improve the accuracy of forecasting.

A SARIMAX model is typically expressed as *SARIMAX(p, d, q)(P, D, Q, s)*, where:

- p, d, q are the non-seasonal ARIMA parameters,
- P, D, Q are the seasonal ARIMA parameters that SARIMAX introduces,
- s is the seasonal period, providing the number of time steps in each seasonal cycle.

The SARIMAX equation can be expressed as follows:

$$
\begin{aligned}
&(1-\phi_1 L-\phi_2 L^2-\ldots-\phi_p L^p)\\
&\quad\times(1-\Phi_1 L^s-\Phi_2 L^{2s}-\ldots-\Phi_P L^{Ps})\\
&\quad\times(1-L)^d(1-L^s)^D y_t\\
&=(1+\theta_1 L+\theta_2 L^2+\ldots+\theta_q L^q)\\
&\quad\times(1+\Theta_1 L^s+\Theta_2 L^{2s}+\ldots+\Theta_Q L^{Qs})x_t+\epsilon_t
\end{aligned}
$$

Where:

- L is the lag operator.
- $\phi_1,\ldots,\phi_p$ and $\Phi_1,\ldots,\Phi_P$ are the autoregressive parameters for the non-seasonal and seasonal components, respectively.
- $\theta_1,\ldots,\theta_q$ and $\Theta_1,\ldots,\Theta_Q$ are the moving average parameters for the non-seasonal and seasonal components, respectively.
- d and D are the orders of non-seasonal and seasonal differencing, respectively.
- x_t represents the exogenous variables.
- ϵ_t is the error term.

SARIMAX is applied in various fields, such as economics, finance, and energy management, which are characterized by seasonal patterns and external influences. Its utilization offers a flexible method for modelling and forecasting complex data, providing insights for decision-making processes.

C. Long Short-Term Memory (LSTM)

LSTM networks are a specialization of Recurrent Neural Networks (RNNs) architecture, aiming to handle and tackle the issue of vanishing gradients in traditional RNNs, making them suitable to effectively model and forecast time series data. In particular, LSTMs offer a way to capture long-term dependencies that are present in sequential time-stamped data. LSTMs consist of several features:

1) Memory Cell, which enables LSTMs to keep track of captured information over long periods of time. This memory cell serves as storage and is responsible for retaining, or discarding information based on the input and internal gates.
2) Gates, which focus on controlling the flow of information within the network. There exist several different types of gates, such as forget, input, and output gates, each serving a different purpose.
3) Non-linear Activation Function, such as sigmoid, or hyperbolic tangent (tanh) to regulate the flow of information and compute the output of the aforementioned gates.

The internal architecture of an LSTM network consists of several LSTM cells in a layered approach, in which each LSTM cell contains the memory cell, along with the forget, input, and output gates, while the input to each cell is the combination of the current input and output from previous cells.

The mathematical expression of LSTM is:

$$
\begin{aligned}
f_t &= \sigma(W_f\cdot[h_{t-1},x_t]+b_f)\\
i_t &= \sigma(W_i\cdot[h_{t-1},x_t]+b_i)\\
\tilde{C}_t &= \tanh(W_C\cdot[h_{t-1},x_t]+b_C)\\
C_t &= f_t * C_{t-1}+i_t*\tilde{C}_t\\
o_t &= \sigma(W_o\cdot[h_{t-1},x_t]+b_o)\\
h_t &= o_t*\tanh(C_t)
\end{aligned}
$$

Where:

- f_t, i_t, o_t are the forget, input, and output gates respectively.
- $\tilde{C}_t$ is the candidate cell state.
- C_t is the cell state.
- h_t is the output of the LSTM cell.
- W_f, W_i, W_C, W_o are weight matrices, and b_f, b_i, b_C, b_o are bias vectors.

LSTM is a powerful and flexible approach in time series forecasting, providing promising results as it is capable of capturing long-term dependencies that traditional statistical methods may struggle with, and capturing complex and non-linear relationships in time series data.

D. Deep Neural Networks (DNNs)

DNNs are another machine-learning technique, capable of identifying and capturing complex patterns and relationships toward time series forecasting. While it is widely known that DNNs are utilized in the context of image recognition and natural language processing, they can also be effectively applied to time series forecasting tasks. In an abstract manner, DNNs consist of:

1) Input Layer, which receives the sequential data, represented as a time series, corresponding to a feature, or a lagged value of the time data,
2) Hidden Layers which perform feature extraction and nonlinear transformations on the input data. It is important to note that multiple hidden layers exist, aiming to learn hierarchical representations of the time series data,
3) Output Layer, which produces the forecasted values of the data. More precisely, in the context of time series forecasting, this layer consists of a single neuron predicting the next value in the sequence.

The mathematical expression of DNN is:

$$
\begin{aligned}
y = f\Big(&W^{(L)}\cdot f\Big(W^{(L-1)}\cdot f(\ldots f(W^{(1)}\cdot x+b^{(1)})\ldots)\\
&+b^{(L-1)}\Big)+b^{(L)}\Big)
\end{aligned}
$$

Where:

- y is the output of the neural network.
- x is the input.
- $f(\cdot)$ represents the activation function, such as sigmoid, ReLU, etc.

- $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias vector, respectively, for the l-th layer of the network.

Overall, by utilizing deep learning techniques to capture complex patterns and dependencies, DNNs offer a powerful approach to time series forecasting. With careful architecture design, pre-processing steps, and appropriate training, DNNs can provide accurate and robust forecasts for multiple time series applications.

IV. Evaluation Analysis

A. Experimental Setup

As already described, this paper examines different forecasting techniques, namely, ARIMA, SARIMAX, LSTM, and DNNs. Regarding the experimental setup, the main programming language used was Python. Pandas framework was utilized for the pre-processing of the dataset, while statsmodels and Keras framework were utilized for the development and testing of the forecasting techniques. It's worth noting that these experiments were conducted in a Macbook with the M1 chip.

B. Dataset Description

For the purposes of training the aforementioned models, historical data concerning Ikaria's grid operation and electrical production had to be gathered. The dataset is publicly available on HEDNO's (Hellenic Energy Distribution Network Operator – Greek DNO) website and is a combination of raw and processed data. The features that are reported are summarised in Table I.

TABLE I
FEATURES DESCRIPTION IN IKARIA'S DATASET

Feature	Description
Timestamp	The time stamp that the measure was reported
Mean Hourly Load	The actual mean load of the island for the specified hourly frame in MW
Thermal Energy Production	Thermal energy production through conventional production units in MW
Active Thermal Units	Specification of which conventional production units are actively producing at the specified time frame.
Wind Energy Production	Production from wind energy of the islands Wind Parks in MW
Wind Energy Restriction Commands	Maximum allowed total production from wind energy specified by HEDNO
Maximum Potential Wind Energy Absorption	Maximum potential energy production in MW from the Wind Parks in a theoretical scenario without curtailments
Max Wind	Actual wind energy production (4.) to maximum potential energy absorption (6.) ratio
Hydroelectric Production	Total energy production that comes from Ikaria's Hydroelectric Hybrid Station in MW

This dataset is monitored and gathered live by HEDNO's metering points located at the start of the Medium Voltage transport lines, the thermal production station, and the dispersed production feeders respectively. The live data is displayed in HEDNO's SCADA systems and is also saved in a structured form in HEDNO's database. The active thermal units and wind energy restriction commands are compiled from HEDNO's daily operational controls, and the maximum potential wind energy absorption is calculated by extrapolating the recorded wind speed of each park on a wind speed-production curve. Lastly, the max speed is derived by dividing the wind energy production by the maximum potential wind energy absorption. This data-gathering process is necessary for the DNO in order to surveil the grid's operation and intervene in real-time, create a historical archive for further analysis and optimization, and publish the non-sensitive data to be utilized by any interested third parties. Furthermore, certain indicators showcase the excess wind production that was not utilized and are useful to HEDNO's decarbonization goals.

As far as experimentation is concerned, the first six months of the dataset were fed as input to train the forecasting models, while the last three months were used for evaluating the performance.

C. Experimental Results

Aiming to measure the performance of each forecasting method, and provide a comparative study, appropriate metrics need to be selected. As stated in similar comparative studies, this paper will measure the following metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). The mathematical formulas of these metrics are presented in the equations (1), (2), and (3):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{1}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{2}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{3}$$

In Table II, the aforementioned metrics per forecasting method are reported and presented.

TABLE II
PERFORMANCE PER FORECASTING METHOD

	RMSE	MAE	MAPE
ARIMA	1.20	0.95	25.05%
SARIMAX	4.02	3.77	92.78%
LSTM	0.14	0.11	2.45%
DNN	**0.13**	**0.09**	**2.11%**

Based on Table II, the superiority of deep learning techniques is obvious compared to traditional statistical methods. In particular, LSTM and DNN achieved an extremely low RMSE, MAE, and MAPE with the DNN providing slightly better performance with 0.14, 0.11, and only 2.45% RMSE, MAE, and MAPE respectively. On the other hand, SARIMAX had the poorest performance with 4.02, 3.77, and 92.78% RMSE, MAE, and MAPE respectively. Figure 1 illustrates and validates the high performance of DNN in predicting the values of the last three months.

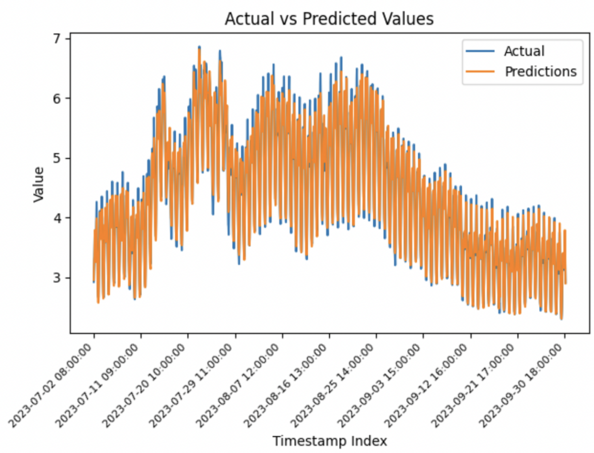

Fig. 1. Load Forecasting with DNN

V. Conclusions & Future Work

In conclusion, this comparative analysis of load forecasting techniques provided insights into the fast-evolving landscape of predictive modelling in energy management. Through multiple experiments and evaluations, this study proved that deep learning techniques, particularly LSTM and DNN, outperform traditional statistical forecasting methods such as ARIMA and SARIMAX in real-world load forecasting settings. The ability of deep learning techniques to capture complex patterns and non-linear dependencies in the time series data serves as the foundation of their superiority and their performance in successfully analyzing modern dynamic environmental settings. However, while deep learning approaches demonstrate high performance, their implementation, and deployment need careful consideration in the context of computational resources, model complexity, data requirements, response time, and the interpretability and accountability of these methods. Toward this, future works can focus on enhancing the transparency, and explainability of deep learning approaches in load forecasting. Additionally, the investigation of ensemble techniques to take advantage of the benefits of different forecasting methods for load forecasting combined with the exogenous factors and knowledge can significantly enrich the overall predictive capabilities.

References

[1] A. Lytos, D. Pliatsios, P.-A. Karypidis, P. Radoglou-Grammatikis, I. M. Bothos, I. Pragidis, A. Sarigiannidis, D. Margounakis, E. Grigoriou, T. Saoulidis *et al.*, "The evident platform: Leveraging crowdsourcing approaches for assessing consumers behavior," in *2022 2nd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED)*. IEEE, 2022, pp. 1–6.

[2] H. Habbak, M. Mahmoud, K. Metwally, M. M. Fouda, and M. I. Ibrahem, "Load forecasting techniques and their applications in smart grids," *Energies*, vol. 16, no. 3, 2023. [Online]. Available: https://www.mdpi.com/1996-1073/16/3/1480

[3] L. Zhang, J. Wen, Y. Li, J. Chen, Y. Ye, Y. Fu, and W. Livingood, "A review of machine learning in building load prediction," *Applied Energy*, vol. 285, p. 116452, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261921000209

[4] F. Dewangan, A. Y. Abdelaziz, and M. Biswal, "Load forecasting models in smart grid using smart meter information: A review," *Energies*, vol. 16, no. 3, 2023. [Online]. Available: https://www.mdpi.com/1996-1073/16/3/1404

[5] M. Cordeiro-Costas, D. Villanueva, P. Eguía-Oller, M. Martínez-Comesaña, and S. Ramos, "Load forecasting with machine learning and deep learning methods," *Applied Sciences*, vol. 13, no. 13, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/13/7933

[6] S. Desai, T. Dalal, S. Kadam, and S. Mishra, "Electrical load forecasting using machine learning," in *2021 International Conference on System, Computation, Automation and Networking (ICSCAN)*, 2021, pp. 1–6.

[7] R. Wazirali, E. Yaghoubi, M. S. S. Abujazar, R. Ahmad, and A. H. Vakili, "State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques," *Electric Power Systems Research*, vol. 225, p. 109792, 2023.

[8] T. G. Grandón, J. Schwenzer, T. Steens, and J. Breuing, "Electricity demand forecasting with hybrid classical statistical and machine learning algorithms: Case study of ukraine," *Applied Energy*, vol. 355, p. 122249, 2024.

[9] C. Tarmanini, N. Sarma, C. Gezegin, and O. Ozgonenel, "Short term load forecasting based on arima and ann approaches," *Energy Reports*, vol. 9, pp. 550–557, 2023.

[10] D. V. Pombo, M. J. Rincón, P. Bacher, H. W. Bindner, S. V. Spataru, and P. E. Sørensen, "Assessing stacked physics-informed machine learning models for co-located wind–solar power forecasting," *Sustainable Energy, Grids and Networks*, vol. 32, p. 100943, 2022.

[11] E. Spiliotis, V. Assimakopoulos, and S. Makridakis, "Generalizing the theta method for automatic forecasting," *European Journal of Operational Research*, vol. 284, no. 2, pp. 550–558, 2020.

[12] C. Sekhar and R. Dahiya, "Robust framework based on hybrid deep learning approach for short term load forecasting of building electricity demand," *Energy*, vol. 268, p. 126660, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544223000543

[13] X. Chen, W. Chen, V. Dinavahi, Y. Liu, and J. Feng, "Short-term load forecasting and associated weather variables prediction using resnet-lstm based deep learning," *IEEE Access*, vol. 11, pp. 5393–5405, 2023.

[14] N. Andriopoulos, A. Magklaras, A. Birbas, A. Papalexopoulos, C. Valouxis, S. Daskalaki, M. Birbas, E. Housos, and G. P. Papaioannou, "Short term electric load forecasting based on data transformation and statistical machine learning," *Applied Sciences*, vol. 11, no. 1, p. 158, 2020.

[15] K. Shahare, A. Mitra, D. Naware, R. Keshri, and H. Suryawanshi, "Performance analysis and comparison of various techniques for short-term load forecasting," *Energy Reports*, vol. 9, pp. 799–808, 2023, 2022 9th International Conference on Power and Energy Systems Engineering. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2352484722024726

[16] D. Zhang, S. Wang, Y. Liang, and Z. Du, "A novel combined model for probabilistic load forecasting based on deep learning and improved optimizer," *Energy*, vol. 264, p. 126172, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0360544222030584

[17] A. Motwakel, E. Alabdulkreem, A. Gaddah, R. Marzouk, N. M. Salem, A. S. Zamani, A. A. Abdelmageed, and M. I. Eldesouki, "Wild horse optimization with deep learning-driven short-term load forecasting scheme for smart grids," *Sustainability*, vol. 15, no. 2, 2023. [Online]. Available: https://www.mdpi.com/2071-1050/15/2/1524

[18] N. Bacanin, C. Stoean, M. Zivkovic, M. Rakic, R. Strulak-Wójcikiewicz, and R. Stoean, "On the benefits of using metaheuristics in the hyperparameter tuning of deep learning models for energy load forecasting," *Energies*, vol. 16, no. 3, 2023. [Online]. Available: https://www.mdpi.com/1996-1073/16/3/1434

[19] X. Wang, Z. Yao, and M. Papaefthymiou, "A real-time electrical load forecasting and unsupervised anomaly detection framework," *Applied Energy*, vol. 330, p. 120279, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0306261922015367