Heliyon 10 (2024) e25821

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Addressing energy challenges in Iraq: Forecasting power supply and demand using artificial intelligence models

Morteza Aldarraji [a], Belén Vega-Márquez [a,*], Beatriz Pontes [a], Basim Mahmood [b], José C. Riquelme [a]

[a] Dept. Computer Languages & Systems, University of Seville, Seville, 41012, Spain
[b] Dept. Computer Science, University of Mosul, Mosul, 41002, Iraq

ARTICLE INFO

Keywords:
Time series forecasting
Electricity supply/demand forecasting
Deep learning
Machine learning

ABSTRACT

The global surge in energy demand, driven by technological advances and population growth, underscores the critical need for effective management of electricity supply and demand. In certain developing nations, a significant challenge arises because the energy demand of their population exceeds their capacity to generate, as is the case in Iraq. This study focuses on energy forecasting in Iraq, using a previously unstudied dataset from 2019 to 2021, sourced from the Iraqi Ministry of Electricity. The study employs a diverse set of advanced forecasting models, including Linear Regression, XGBoost, Random Forest, Long Short-Term Memory, Temporal Convolutional Networks, and Multi-Layer Perceptron, evaluating their performance across four distinct forecast horizons (24, 48, 72, and 168 hours ahead). Key findings reveal that Linear Regression is a consistent top performer in demand forecasting, while XGBoost excels in supply forecasting. Statistical analysis detects differences in models performances for both datasets, although no significant differences are found in pairwise comparisons for the supply dataset. This study emphasizes the importance of accurate energy forecasting for energy security, resource allocation, and policy-making in Iraq. It provides tools for decision-makers to address energy challenges, mitigate power shortages, and stimulate economic growth. It also encourages innovative forecasting methods, the use of external variables like weather and economic data, and region-specific models tailored to Iraq's energy landscape. The research contributes valuable insights into the dynamics of electricity supply and demand in Iraq and offers performance evaluations for better energy planning and management, ultimately promoting sustainable development and improving the quality of life for the Iraqi population.

1. Introduction

Introducing the context of this research needs to present the global issue of energy demand and the issue in the country of the case study. Moreover, describing the commonly used forecasting models is also important before exploring the energy forecasting literature, which enables in accurately stating the problem of this research and the approaches for overcoming it. These aspects are explored and covered in this section.

* Corresponding author.
E-mail address: bvega@us.es (B. Vega-Márquez).

https://doi.org/10.1016/j.heliyon.2024.e25821
Received 9 June 2023; Received in revised form 9 January 2024; Accepted 2 February 2024

Available online 6 February 2024
2405-8440/

1.1. Global energy demand

The demand for energy has become a fundamental requirement for the development of nations due to the continuous growth of technological devices and the significant increase in the global population. Consequently, there has been a significant increase in the demand for energy worldwide. The production of large electrical appliances, the proliferation of factories in urban areas, and the rising population have all contributed to this trend.

In addition to the continued growth of technological devices and the rising global population, the negative impact of non-renewable energy sources on the climate has become increasingly apparent. As a result, there is a growing demand for renewable energy sources such as hydro, geothermal, wind, and solar. Many countries aim to transition to using only renewable energy by 2050 [1]. However, the generation of energy from renewable sources is only part of the solution. There is also a need for effective utilization of this energy through proper planning and distribution. Grid systems seek to supply energy based on demand to avoid storage costs or oversupply of energy in certain regions, while other regions experience a shortage [2]. One of the reasons for the energy shortage is that traditional grid systems cannot accurately estimate energy demand. Moreover, fluctuations in energy demand cause traditional grid systems to store large amounts of energy at certain times of the year and run out of energy supply at other times [3]. To solve this problem, it is crucial to accurately estimate the energy demand at all times. Forecasting energy demand would help with accurate planning and the proper distribution of energy to endpoints. Given the significant investment required for network reinforcements and expansions, it is appropriate to forecast future load and demand to ensure proper planning. Economic conditions, time of day, weather patterns, and other random factors all have an impact on the system load. On the other hand, energy demand typically follows general consumption patterns in the economy and is subject to fluctuations based on changes in demographics, industry activity, and weather conditions [4].

Smart grid systems come as solutions to these problems [5]. Energy distribution and utilization can be monitored and controlled. The advent of modern systems, such as smart meters and other advanced metering frameworks, allows data on the bidirectional flow of energy to be obtained [6] [7]. Such data can be analyzed and utilized for future prediction and forecasting.

1.2. Energy demand issue in Iraq

The unstable security situation in Iraq has had a negative impact on electric power generation, which results in a shortage of supply. Additionally, the newly introduced technologies, the lack of strategic planning, mismanagement, and infrastructure together increase the energy demand in Iraq. Other reasons, such as low gas supply rates, the use of traditional grid systems, the exposure of power stations and transmission lines to terrorist attacks, the failure to use the smart meter, and the control of violators on distribution lines and traditional grid systems, have also had a great impact on the stability of the power grid in Iraq. Currently, the demand for electricity exceeds the supply and capacity to produce electricity in Iraq, as shown in Fig. 1.

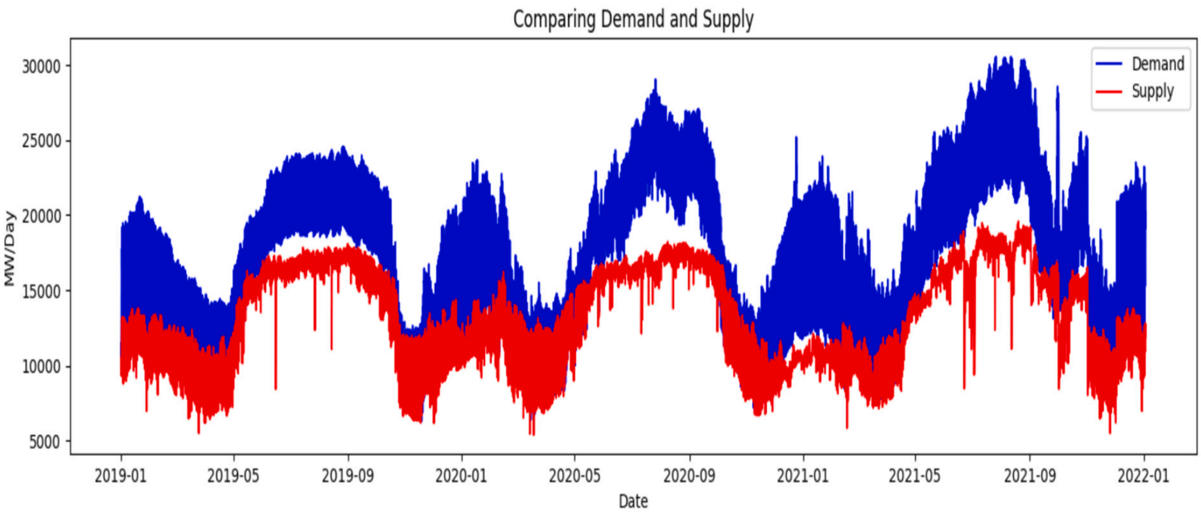

Fig. 1. Comparing Demand and Supply.

Iraq has recently made efforts to upgrade and develop its infrastructure to keep pace with the latest technological developments in the electricity sector. After 2003, Iraq was opened to the global energy market. The 2014 report from the United Nations Development Program (UNDP) shows that 35% of Iraqis demand the provision of electricity and consider it a top priority [8]. The electricity grid in Iraq has been severely damaged by wars, successive conflicts, and economic sanctions in the 1990s. To date, there are no studies that address the issue of electrical energy in Iraq in terms of forecasting demand and prices.

Many power plants were built in Iraq between the mid-1970s and 1980s, with a few small gas-fired plants operating in 2003. Most current power plants are thermal, which use crude oil supported by gas and hydropower plants. The unserved demand is currently served by distributed diesel generators, which are privately owned. On 9 January 2021, according to a statement by the Iraqi Ministry of Electricity, some estimations indicated that Iraq produces and imports 19 to 21 thousand megawatts of electricity, while the actual need exceeds 30 thousand megawatts. Therefore, Iraq needs to increase its production capacity by nearly double to secure stable levels of electrical energy, while its population may double by 2050. This means that its energy consumption will increase by a higher percentage than the increase in electricity production.

Despite the increase in electricity production during 2021 in Iraq, which amounted to about 20 thousand megawatts, the scene of electricity cuts continues, especially during the peak (in summer), when the temperature exceeds fifty degrees Celsius, and the size of the shortage in the supply of electric power in Iraq exceeds 10,000 megawatts due to several factors as follows:

- The increasing targeting of electric power systems and towers by sabotage.
- The decline in gas emissions supplied by Iran to operate the stations.
- The governorates lack commitment to the quotas approved for them in terms of the amount of energy supplied.
- The continuation of the emergence of informal agricultural and squatter areas adds new burdens to the system.
- The rise in temperature and the technical symptoms that accompany it.
- Obsolescence of transmission and distribution networks.

The energy system in Iraq is currently hierarchical, with the Ministry of Electricity exercising control over every aspect of the process, including providing electricity and equipment to consumers as well as billing and accounting services. This approach to control causes confusion and internal conflicts within the ministry, resulting in substandard service. Furthermore, the ministry functions as a policymaker, operator, regulator, and supplier, creating a potential conflict of interest. In addition, the electricity sector lacks a formal regulatory framework, and despite the issuance of invoices, there is no interaction with consumers regarding electricity services.

1.3. Energy forecasting

Energy forecasting is a crucial factor for any energy utility company. It helps guide their decision about whether there is a need for infrastructural development, the energy supply per time, load switching decisions, or the cost of energy, to mention a few. Accurate forecasting of energy demand is essential in preparing for the future and ensures that consumers do not experience energy shortages. And could use people's opinions to add to the model and improve performance [9]. Load forecasting can be classified into four categories according to the time horizon over which the forecast is made [10]:

- *Long-Term Load Forecasting (LTLF)*: This is a class of load forecasting whose time horizon is measured in months or perhaps years. They are mostly used when price or risk management assessments are done.
- *Midterm Load Forecasting (MTLF)*: This is a class of load forecasting where the time horizon is in a couple of days to a few months. They are useful when stakeholders want to evaluate the financial implications of their systems. They are used when the energy price needs to be fixed or a risk management assessment is necessary.
- *Short-Term Load Forecasting (STLF)*: In this type of load forecasting, the time horizon is between a few minutes and a few days as well. This is critical when the utility company needs to have a robust understanding of the energy consumption behavior of its end-users.
- *Very Short-Term Load Forecasting (VSTLF)*: This is a type of forecasting whose time zone is in minutes or a few hours. They are typically not more than 3 hours.

In recent years, researchers have been using machine learning techniques and deep learning models to predict energy demand [11][12]. Deep learning models consist of layers of interconnected units called "perceptions" that are trained on data to make accurate predictions [13]. However, traditional artificial neural networks are designed to work with static data, which is not typically found in smart grid systems. Instead, smart grid data usually takes the form of "time series" data, which change over time and follow a pattern based on past events [14] [15]. To effectively learn from this type of data, a neural network would need to forget unimportant information and retain important information for future use. This is where recurrent neural networks (RNNs) come in; they are designed to handle sequential data by selectively retaining information. A specific type of RNN, known as a Long-Short-Term

Memory (LSTM) model, uses input gates, forget gates, and output gates to selectively store and retrieve information, making it well suited for processing time series data.

1.4. Problem statement and contribution

Predicting power supply and demand in an unstable country like Iraq is a challenging task for two main reasons. *First*, it's difficult to collect time-series data from traditional power grid systems. *Second*, the inconsistent operation of power plants and the varying amounts of imported power from neighboring nations throughout the year have a significant impact on the completeness and accuracy of the data collected. These factors make it even more challenging for Iraqi officials to design plans to improve the current power grid and make decisions about how to handle the growing population and increasing demand for power due to advances in technology.

The existing literature on the Iraqi power grid network indicates a severe lack of comprehensive studies or datasets related to time-series-based supply and demand. Furthermore, most studies that exist rely on traditional approaches for analyzing supply and demand, which are insufficient considering the rapid growth in population and technology. Therefore, this work addresses a critical issue in Iraq by focusing on forecasting electricity supply and demand. The novelty of the paper lies in its empirical analysis of various machine learning and deep learning models for predicting electricity supply and demand. It also distinguishes itself from the existing literature by addressing the limitations of previous works and offering insights into the unique challenges faced in Iraq's energy sector. Therefore, this study aims to make the following contributions:

- Collecting a time-series-based dataset for the years 2019 to 2021, encompassing a range of electricity demand and supply values. The dataset is novel and was officially collected with the support of the Operation and Control Office, Ministry of Electricity, Baghdad, Iraq. The minimum electricity demand value is 6336 MW/day, and the maximum value is 29059 MW/day. The minimum supply value is 5399 MW/day, and the maximum value is 18233 MW/day.
- Using the collected dataset, the supply and demand of electricity in 15 provinces in Iraq were predicted. The structural time-series modeling approach was applied to annual data for the period between 2019 and 2021, using estimated equations and value assumptions. To make the predictions, the study used a range of machine learning and deep learning models, including Linear Regression (LR), XGBoost (XGB), Random Forest (RF), Long Short-Term Memory (LSTM), Temporal Convolutional Networks (TCN), and Multilayer Perceptron (MLP) Models. The study also used various metrics for benchmarking and statistical analysis for verifying the differences between the involved models.

The remainder of this article is organized as follows: the next section outlines the research methodology, including details of the data collection process, the models used, the settings of the experiments in terms of the parameters of the models, and the metrics involved in the process of performance evaluation. Section 3 presents and discusses the results obtained and assess the differences of models' performance using statistical testing approaches. Finally, Section 5 concludes the article and presents future directions as well as the limitations of this research.

2. Literature review

This section explores the related energy forecasting literature and presents the state-of-the-art.

2.1. Threat to validity

Before delving into the literature, it was essential to acknowledge potential threats to the validity of the literature search. The search for related literature was conducted using specific search strings and databases to ensure comprehensive coverage. However, it was important to recognize that despite the efforts spent in this work, some relevant sources may not have been included in this work. The search strings employed included variations of terms such as "*Energy Forecasting*", "*Electricity Demand Prediction*", "*Machine Learning*", "*Deep Learning*", and "*Iraq Energy Sector*". These strings were designed to capture a wide range of relevant articles and studies. The publishers explored in this research encompassed many academic publishers, including but not limited to *Elseveir*, *IEEE*, *MDPI*, and *Springer*. While the aim was to be as exhaustive as possible, the vast and dynamic nature of the energy forecasting literature may still lead to some omissions. To mitigate potential biases and ensure the relevance of the selected works, the focus was on the peer-reviewed articles published in high reputable journals and conferences [16].

Additionally, the state-of-the-art publications were considered, prioritizing those from the last decade to ensure the applicability of the findings to contemporary energy forecasting challenges. Despite these measures, it was important to recognize that the landscape of energy forecasting is continually evolving. New methods, data sources, and insights emerge regularly. Therefore, this work represents a snapshot of the literature available up to the authors' knowledge cutoff date in September 2022. While making diligent efforts to provide a comprehensive review of relevant literature, the limitations inherent in any literature search may have influenced the selection of sources. Readers are encouraged to consider this context when interpreting the findings and conclusions presented in the following sections. Table 1 presents a summary of the threat to validity components of this work.

Table 1
Summary of the threat to validate of this research.

Strings and Keywords	Publishers	Databases
Energy Forecasting",	**Elsevier**,	**Web of Science**,
"Power Forecasting",	**IEEE**,	**Scopus**,
"Electricity Forecasting",	**MDPI**,	**PubMed**,
"Electricity Prediction",	**Springer**,	**IEEE Xplore**,
"Electricity Supply Forecasting",	**Wiley& Sons**,	**ScienceDirect**,
"Electricity Supply Prediction",	**IntechOpen**,	**DOAJ**,
"Electricity Demand Prediction",	**Intellect**,	and **JSTOR**
"Electricity Demand Forecasting",	**Citeseer**,	
"Machine Learning",	**UDA**,	
"Deep Learning",	**PeerJ,**	
"Iraq Energy Sector",	**MCB UP**	
"Power Demand",	and **Hindawi**	
and *"Power Supply"*		

2.2. Related work

Over the last few decades, researchers have relied on the use of statistical methods to forecast energy demand, energy balance, or energy supply. A commonly used method is the autoregressive integrated moving average (ARIMA), which has been successful in predicting energy demand in stable load situations. However, this method is not always effective in real-world scenarios, where extreme peak loads can occur intermittently [17].

Moreover, with the advent of neural networks, better solutions are now available to researchers. Neural networks can learn hidden patterns from data, which is a significant improvement over purely statistical methods. In fact, neural networks operate similarly to how humans learn: by making a prediction, receiving feedback, and then adjusting their prediction accordingly. A deep neural network that accounts for sequential data is useful for applications that involve time.

The literature includes a large number of studies on predicting electricity supply and demand. Researchers have employed a variety of methods, ranging from statistically based approaches to machine learning and deep learning methods. The choice of a particular method depends on the specific characteristics of the dataset, such as whether it is a time series, has seasonality, or is stationary. Consequently, researchers typically test their dataset before selecting an appropriate algorithm. For example, Dittmer et al. [18] forecast the demand for electricity in a rural region in Germany. They first examined their data for seasonality and trends and subsequently used ARIMA and other statistical models to forecast 48 hours ahead. The results showed that the models they employed were suitable for performing the forecasting task and allowed predictions to be made up to 14 days in advance. In another study, Kim et al. [19] developed a hybrid deep learning model using LSTM and CNN for the prediction of power demand. The study used a real-world dataset, and the results indicated that the hybrid model was more accurate in predicting power demand compared to using each model individually.

[20] investigated the issue of forecasting short-term electricity demand in Uruguay over the period 2010 to 2019. They employed a variety of models, including linear regression, ridge, KNN, random forest, gradient boosting, MLP, and ExtraTrees, and used benchmarking metrics such as MAE, MAPE, and RMSE. The results indicated that the models mentioned above were suitable for forecasting the hourly power demand. Similarly, Velasquez et al. [21] analyzed the time series of per demand in Brazil for the period 2014 to 2019, using various forecasting approaches. They found that incorporating regression and seasonality with mixing time-series approaches can help reduce forecasting errors.

Other researchers aim to test different approaches and determine the most appropriate method for their dataset. Pallonetto et al. [22] recently compared deep neural networks and the Support Vector Machine (SVM) approach. Their results indicated that LSTM provided more accurate forecasting when the load data used in training was sufficient, while SVM performed better when the load was insufficient. The two approaches were applied to one-hour-ahead and one-day-ahead load forecasting. Similarly, Banga et al. [23] compared the power demand forecasting performance of ten models: ARIMA, Prophet, LR, SVM, XGBoost, RF, KNN, RNN, LSTM, and GRU. They evaluated the performance of these models using metrics such as RMSE, MAE, MAPE, and R2. Their findings suggested that at the hourly and daily levels, the Prophet model provided more accurate forecasting compared to the other models.

Additionally, several studies have investigated the energy demand and supply of different countries through prediction processes. For example, Raza et al. [24] focused on Pakistan and aimed to create a balance between power demand and supply for economic purposes. They used the Long-Range Energy Alternatives Planning System (LEAP) to perform forecasts, and the results suggested that Pakistan can generate more power to meet future needs. Similarly, Jaramillo et al. [25] studied the case of Ecuador and used the SARIMA modeling approach for the monthly forecast of the power demand, which proved to be efficient. These studies demonstrate the importance of forecasting in the achievement of sustainable energy systems worldwide.

Table 2 summarizes the aforementioned studies in terms of the model used, datasets, limitations, and advantages.

Table 2
A summary of the literature.

Study	Model	Dataset	Limitations	Advantages
[17]	ARIMA	the station's monthly discharge period from water year 1960-1961water year 2006-2007	Not appropriate for real-world data	Stable in predicting energy demand load situation
[18]	ARIMA	Real world data of 48 h horizon	No accurate when predicting more than 14 days in advance	performing the forecasting task and enabled predictions to be made up to 14 days in advance.
[19]	Hybrid (LSTM-CNN)	USA District Public consumption for 2016	Not appropriate for long terms prediction	Appropriate for real-world datasets
[20]	Linear Regression, Ridge, KNN, Random Forest, Gradient Boosting, MLP, and ExtraTrees	The studied dataset contains residential energy consumption measurements collected between January 2017 and December 2018	Not appropriate for uncertainty in electricity demand	These models are appropriated for hourly prediction
[22]	SVM	hourly load data from 2013 to 2018, with a total of 52,584 data.And daily load from 2013 to 2018, with 2191 data.	Difficult to deal with anomalies in data	These models are appropriated for hourly and daily prediction
[23]	ARIMA, Prophet, LR, SVM, XGBoost, RF, KNN, RNN, LSTM, and GRU	Electricity consumption dataset of house from 11 Jan, 2016, to 27 May 2016 (around 4.5 Months duration) per 10-minute observation	Difficult to deal with many features	Prophet outperformed the other models in terms of hourly and daily prediction
[24]	LEAP	electricity consumption data and growth rate of electricity consumption for the period 2008 to 2018.	Accuracy needs to be improved	Able to predict the power demand for the period between 2018-2030
[25]	SARIMA	Ecuadorian annual maximum demand from 1990 to 2019	Accuracy needs to be improved	Stable in predicting energy demand load situation

3. Research methodology

This section describes about the data collection process as well as the forecasting models involved in this research. Also, the setup of the experiments in terms of optimizing the models' parameters and the evaluation metrics are also explained.

3.1. Dataset collection

The data used in this work was officially collected from the Department of Operations and Control of the Ministry of Electricity in Baghdad, Iraq, for the period 2019 to 2021. The data consisted of hourly time-series data on the supply and demand of 15 provinces in Iraq. The collection process was strictly regulated due to governmental procedures in Iraq, which took approximately 4 months to complete. Then it was preprocessed and cleaned to address missing values and outliers. In total, the dataset consisted of 26,352 rows and 15 columns, each corresponding to a different province.

3.2. Time series forecasting models

One of the main considerations in the analysis of time series data is the examination of their inherent seasonality [26]. Upon conducting a rigorous Dickey-Fuller analysis on the dataset (see Figs. 2 and 3), it was evident that both demand and production exhibited non-stationary behaviors. Consequently, in light of this assessment, the decision was made to employ forecasting models that are robust to the absence of stationarity in the series. This judicious choice not only enhances the robustness of the analysis, but also confers adaptability by bypassing the strict stipulation that the data must conform to stationarity, a conventional prerequisite in many established statistical models.

In this study, an extensive analysis was conducted using six distinct prediction models. Three of them are deep learning-based models: TCN (Temporal Convolutional Network), MLP (Multi-layer Perceptron), and LSTM (Long Short-Term Memory). The remaining three correspond to machine learning models, namely linear regression, XGBoost, and random forest. In the following, a brief description of each model is provided:

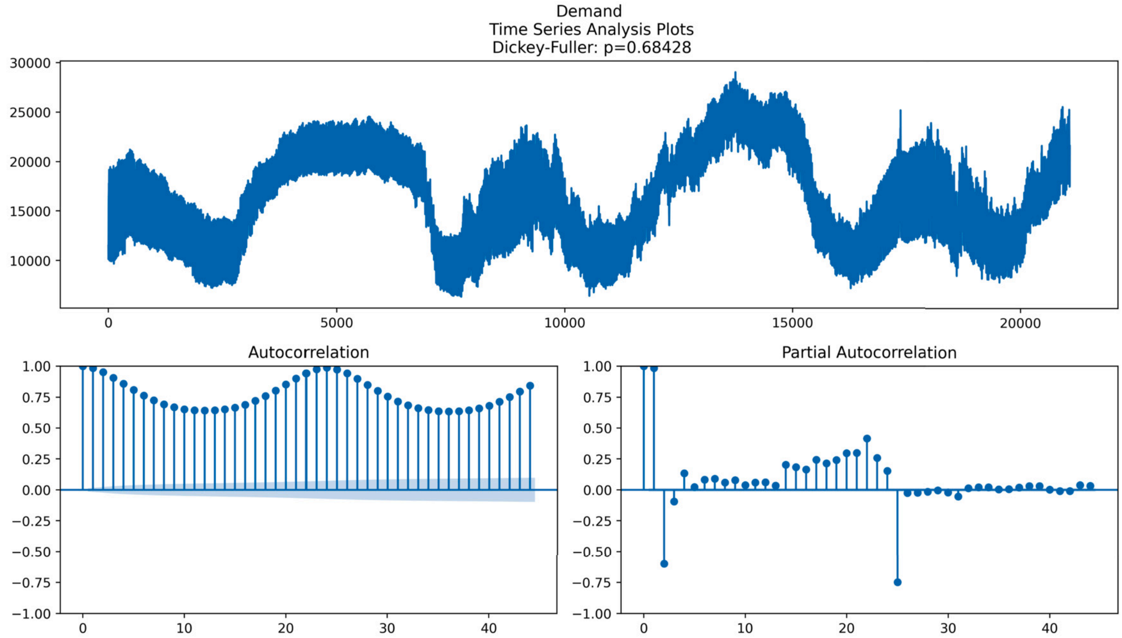

Fig. 2. Dickey-Fuller Test for Demand.

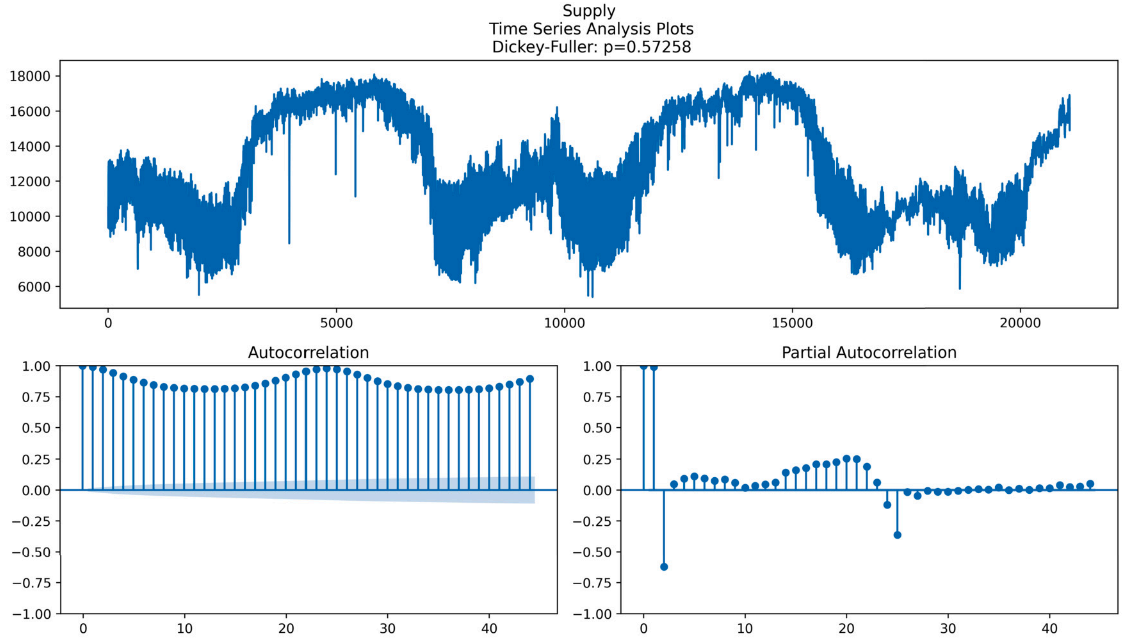

Fig. 3. Dickey-Fuller Test for Supply.

3.2.1. Deep learning models

- **Long Short-Term Memory Network.** The Long Short-Term Memory (LSTM) algorithm is a deep learning method used for prediction purposes that can handle individual data points or a sequence of data points. It has proven to be an effective algorithm that provides accurate predictions based on recent information in the data. LSTM is capable of retaining information for a long period of time to predict, process, and classify time-series datasets. It primarily utilizes four neural networks and memory blocks (cells), which store information and control the information flow using three gates produced by a sigmoid function. The *Input Gate* is used to include useful information; the *Forget Gate* is used to discard data that are no longer useful; and Output Gate is used to extract relevant information from a cell state [27].

- **Temporal Convolutional Network.** The Temporal Convolutional Network (TCN) places more emphasis on temporal series. The TCN is a time-series processing algorithm developed by Bai et al. in 2018 to address the challenge of extracting long-term time-series information. It combines causal convolution, dilated convolution, and residual blocks. The TCN requires low memory for training due to its shared convolutional filters and can process long input sequences through parallel convolutions, making it a more stable training scheme [28] [29].
- **Multi-layer Perceptron.** The Multilayer Perceptron (MLP) is one of the most popular neural networks used to train deep learning models. In this network, the input is presented to the network with the desired output, and the weights are adjusted so that the network attempts to produce the desired output. The MLP consists of three layers: the input layer, which contains the input neurons that feed information to the hidden layer; the hidden layer, which performs calculations based on the input data and forwards the output to the output layer; and the output layer, which represents the model results. The number of hidden layers determines the depth of the network, and it is the reason why an MLP network with more hidden layers is considered a deep learning model [30].

3.2.2. Machine learning models

- **Linear Regression.** Regression analysis is a statistical approach that allows us to determine the strength of the relationship between one or more variables. It can therefore help us predict unknown values based on these relationships. *Simple linear regression* involves using one independent variable to model a linear relationship with a dependent variable. *Multiple linear regression*, on the other hand, involves using multiple independent variables to predict the dependent variable [31]. Although Linear Regression is not inherently suited for modeling non-linear time series data, in this context, some justifications can be found for using it as a prediction model. Linear regression models are simple and easy to interpret and can be a good starting point for modeling time-series data, providing a baseline for understanding the data's structure. After evaluating the performance of the model, linear regression is considered an appropriate choice for the data under study, as shown by the experimental results in Section 4.
- **XGBoost.** XGBoost is a popular gradient boosting algorithm used for machine learning tasks. It was developed by Tianqi Chen [32] as an improvement on the GBM algorithm, using a more regularized model to prevent overfitting. XGBoost is known for its efficiency, flexibility, and portability and has been shown to outperform other algorithms in tasks such as classification, regression, and ranking. The algorithm combines multiple weak learners to create a strong learner, with each weak learner trained on a subset of the data. The algorithm works by training decision trees and combining their predictions to make a final prediction [33].
- **Random Forest.** Random forest models are a popular type of nonparametric machine learning model used for both classification and regression tasks. They belong to the ensemble method category, specifically bagging methods. Ensemble methods use a group of weak learners to create a stronger and more accurate model. Random forests are a collection of many decision trees that are known to be prone to overfitting. By combining multiple trees, random forest models are able to mitigate this issue and provide a more flexible and powerful model with lower variance. This allows larger and more predictive trees to grow, resulting in better performance in both training and unseen data. Additionally, random forest models retain the simplicity and interpretability of decision trees [34].

3.3. Experimental setup

The sliding or rolling window approach is used in time series forecasting to handle the sequential and temporal nature of the data. It involves training the forecasting model on a fixed-length window of past observations and then using the model to make predictions for a specific forecast horizon. The window slides forward in time and is repeated at regular intervals. In this context, the concept of forecast horizon refers to the time in the future for which predictions are to be made. Therefore, it determines how far ahead the prediction of future values is intended based on historical data. On the other hand, the past history horizon, also known as the historical data window, refers to the length of the time series data that are used to make predictions in the forecast horizon.

Both values are essential because they influence the complexity and accuracy of the forecasting model. Furthermore, in the case of past history, it is important to strike a balance between using enough historical data to capture relevant patterns and trends and not including too much data that may introduce noise or outdated information.

The purpose of the study is to analyze the behavior of six different prediction models when forecasting the demand and supply of energy in Iraq for four different horizons ahead: 24, 48, 72, and 168 hours. To do that, the experiments in this study were carried out in two phases, as can be seen in Fig. 4: the first phase was designed to optimize the hyperparameters of each model, while in the second phase, the optimized values were used to train the models on different scenarios, according to the past history and forecast horizons, on both datasets.

The datasets have been split into two parts: training and validation, with a proportion of 80% and 20%, respectively. The first subdivision was used to train the models, while the second one was used to evaluate and compare their performance in both phases. This approach allows us to assess the model's performance on data entirely independent of those used for training, minimizing the risk of overfitting. The use of a test set not seen during the training process ensures that the model not only fits the training data but can also effectively generalize to new observations. This validation approach provides a solid foundation to have confidence in

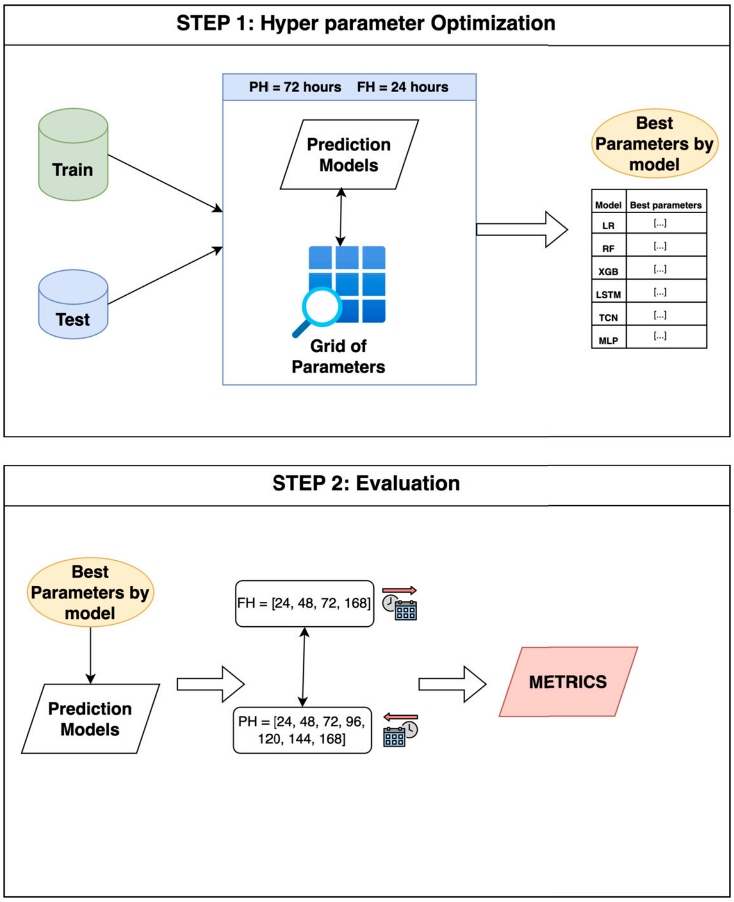

Fig. 4. Hyperparameter optimization process.

Table 3
Training parameters used for deep learning models.

Parameters	Values
Batch size	32, 64
No. of epochs	200
Max steps per epoch	10000
Optimizer	Adam
Learning rate	0.001, 0.01
Normalization	minmax, zscore

the models' ability to make accurate predictions in real-world scenarios, as its performance has been comprehensively evaluated on previously unseen data.

The first phase of the experimentation consists of establishing the best parameter combination for each model by performing an experiment with the entire parameter grid and setting the prediction horizon to 24 hours and the past history to 72 hours. Tables 3 and 4 show the parametrization used in the first phase, including the parameter names together with all the values tested for the three deep learning models. Similarly, Table 5 shows the parametrization for the three machine learning models. In both cases, the parameters that are not specified have been left with their default values.

Various configurations for batch size, number of epochs, maximum steps per epoch, optimizer, and learning rate are discussed in the context of deep learning models. For the batch size, the commonly used values (32 and 64) were selected, along with the number of epochs (200), maximum steps per epoch (10000), and learning rates (0.001 and 0.01). The Adam optimizer was chosen due to its suitability for a broad range of machine learning problems, as reported in the literature [35]. Furthermore, for data preprocessing, two widely used normalization techniques were employed: mean normalization and min-max scaling, also known as z-score normalization, as shown in Table 3. These normalization methods have been used in both deep learning and machine learning models. Moreover, Using different normalization methods in the study serves several purposes: 1) it enhances the robustness of the

Table 4
The parameters used for LSTM, TCN, and MLP deep learning models.

Deep Learning Models	Parameters	Values
LSTM	Recurrent Layers	1, 2, 4
	Units	32, 64, 128
	Return sequence	True, False
	Recurrent dropout	0, 0.2
	Dense dropout	0, 0.2
	Hidden Layers (2)	[8, 16],[16, 8],[32, 64],[64, 32]
	Hidden Layers (3)	[8, 16, 32],[32, 16, 8], [32, 64, 128],[128, 64, 32]
	Hidden Layers (5)	[8, 16, 32, 16, 8],[32, 64, 128, 64, 32]
TCN	Number of stacks	1, 3
	Number of filters	32, 64
	Dilations	[1, 2, 4, 8], [1, 2, 4, 8, 16]
	Kernel size	3, 6
	Return sequence	True, False
	Recurrent dropout	0, 0.2
	Dense dropout	0, 0.2
	Hidden Layers (2)	[8, 16],[16, 8],[32, 64],[64, 32]
	Hidden Layers (3)	[8, 16, 32],[32, 16, 8] [32, 64, 128],[128, 64, 32]
	Hidden Layers (5)	[8, 16, 32, 16, 8],[32, 64, 128, 64, 32]
MLP	Hidden Layers (1)	8 and 32 neurons
	Hidden Layers (2)	[8, 16], [16, 8], [32, 64], [64, 32]
	Hidden Layers (3)	[8, 16, 32], [32, 16, 8], [128, 64, 32], [32, 64, 128]
	Hidden Layers (5)	[8, 16, 32, 16, 8], [32, 64, 128, 64, 32]

Table 5
The parameters used for Random Forest, Linear Regression, and XGBoost machine learning models.

Machine Learning Models	Parameters	Values
Random Forest	Number of estimators	100, 300,600
	Max depth	2, 4, 6, 8, 10
	Min samples split	2, 4, 6, 8
	Min samples leaf	1, 3, 5, 7
Linear Regression	Fit intercept	[true, false]
	Normalize	[true, false]
	Positive	[true, false]
XGBoost	Booster	gbtree
	Number of estimators	[100, 300, 600]
	Min child weight	[1, 5, 10]
	Subsample	[0.5, 0.6, 0.8, 1.0]
	Colsample bytree	[0.5, 0.6, 0.8, 1.0]
	Max depth	[3, 4, 5, 6]

results by accounting for variations in data characteristics, 2) enables comparisons to identify the most effective normalization technique, and 3) assesses the generalization of forecasting models to different data representations and aids in data exploration by revealing specific data patterns.

Table 6 contains the description of each model for demand and supply for the best parameters found. The best parameter combination was then used in the second phase of the experimentation, where experiments have been conducted for each of the remaining forecast and past history horizons.

Table 6
The best parametrization for each model and dataset.

MODELS	DEMAND MODEL DESCRIPTION	SUPPLY MODEL DESCRIPTION
LSTM	Recurrent Layers = 1 Units = 128 Return sequence = False Recurrent dropout = 0 Dense dropout = 0.2 Dense layers = [32, 64, 128]	Recurrent Layers = 1 Units = 64 Return sequence = False Recurrent dropout = 0 Dense dropout = 0.2 Dense layers = [32, 64, 128, 64, 32]
TCN	Number of stacks = 3 Number of filters = 64 Dilations = [1, 2, 4, 8] Kernel size = 3 Return sequences = False TCN dropout = 0.2 Dense dropout = 0.2 Dense layers = [64, 32]	Number of stacks = 3 Number of filters = 64 Dilations = [1, 2, 4, 8] Kernel size = 6 Return sequences = True TCN dropout = 0 Dense dropout = 0 Dense layers = [32, 64]
MLP	Hidden layers = [32]	Hidden layers = [32, 64]
LR	Fit intercept = False Normalize = True Positive = False	Fit intercept = False Normalize = True Positive = False
RF	Number of estimators = 300 Max depth = 10 Min samples split = 4 Min samples leaf = 1	Number of estimators = 300 Max depth = 10 Min samples split = 2 Min samples leaf = 3
XGB	Booster = gbtree Number of estimators = 300 Min child weight = 10 Subsample = 0.8 Colsample bytree = 0.8 Max depth = 6	Booster = gbtree Number of estimators = 300 Min child weight = 10 Subsample = 0.8 Colsample bytree = 0.5 Max depth = 6

3.4. Evaluation metrics

The purpose of this study is to evaluate and compare the performance of previous models under different scenarios. To achieve this, five widely recognized metrics from the forecasting literature [36,37] were selected: mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE) and weighted average percentage error (WAPE). The chosen metrics were used as a basis for evaluating and comparing the effectiveness of the algorithms in each scenario. The respective formulas for these metrics are defined in Equations (1), (2), (3), (4) and (5), respectively.

- **Mean Absolute Error (MAE).** MAE reflects the average of the absolute differences between the actual and predicted observations in the test sample. It is calculated as follows:

$$MAE = \frac{1}{n}\sum_{j=1}^{n} |y_i - \hat{y}_j|, \tag{1}$$

  where i and j are indexes, n is the number of observations, y_j and $\hat{y}_j$ are the actual and the predicted values, respectively.
- **Mean Squared Error (MSE).** It measures the average of squared differences between actual and predicted observations. Using the same symbols as in MAE, the formulation is as follows:

$$MSE = \frac{1}{n}\sum_{j=1}^{n} (y_i - \hat{y}_j)^2 \tag{2}$$

- **Root Mean Squared Error (RMSE).** It measures the square root of the average of squared differences between the actual and predicted observation and, using the previous symbols, RMSE can be formulated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_i - \hat{y}_j)^2} \tag{3}$$

- **Mean Absolute Percentage Error (MAPE).** This metric evaluates the accuracy of the prediction of the forecasting model. It is formulated using the following equation:

$$MAPE = \frac{1}{n}\sum_{j=1}^{n}\frac{|y_i - \hat{y}_j|}{\max(\epsilon, |y_i|)} \tag{4}$$

- **Weighted Average Percentage Error (WAPE).** This metric evaluates the accuracy of the prediction of the forecasting model, taking into account the weights of the observations. It is formulated using the following equation:

$$WAPE = \frac{1}{n}\sum_{j=1}^{n}\frac{|y_i - \hat{y}_j| \times w_j}{\max(\epsilon, |y_i|)}, \tag{5}$$

where w_j represents the weight assigned to each observation.

4. Results and discussions

This section discusses the results achieved through the use of the chosen forecasting models and the collected dataset. The objective was to evaluate and compare the performance of the different models for demand and supply datasets in various scenarios. Specifically, the objective is to select the best prediction model for each forecast horizon (24 h, 48 h, 72 h, and 168 h). For this purpose, the models were tested with seven different past history periods. Reproducibility of experiments can be performed with the code from the public repository located at [38].

Consistent with previous studies, five widely recognized metrics from the forecasting literature were selected to evaluate the models; all of them have been explained in the previous section. All implementations were carried out using the Python programming language.

Before the experiments were carried out, it was necessary to test the trends and seasonality of the data. Fig. 5 depicts the trend of demand and seasonality of the data. According to Fig. 5, it could be observed that the data has trended at specific time series and seasonality features. The demand trend showed the general direction in which demand for power is moving over time. Clearly, some periods showed an increase or decrease in demand based on the period of time. However, seasonality in the figure refers to patterns in demand that repeat over time, such as higher demand for certain times during a season. The residuals represent the differences between the actual demand for power and the predicted demand. They represent the variability in demand that cannot be explained by trends or seasonality. The same behavior is observed in the supply data, as shown in Fig. 6. Visualizing the demand and supply shows the fluctuation of power in Iraq.

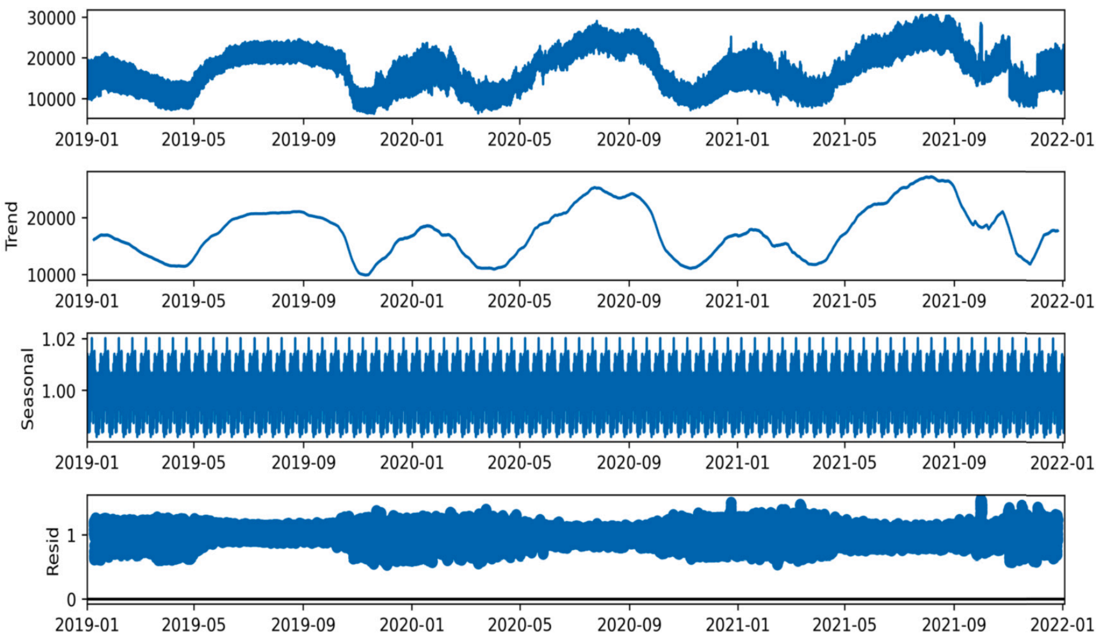

Fig. 5. Demand trend and seasonality analysis.

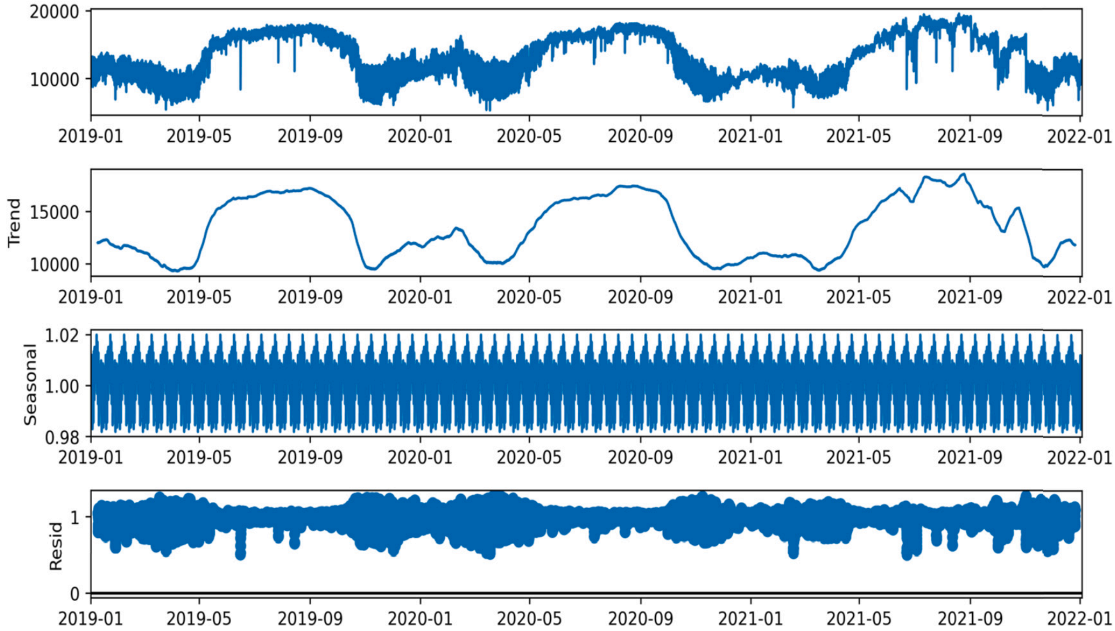

Fig. 6. Supply trend and seasonality analysis.

As announced in Section 3 methodology is divided into two parts. The first step was designed to determine which hyperparameters are best for each model with a specific past history of 72 hours and a forecast horizon of 24 hours. Once these parameters and thus the best model have been identified, the model is trained on different scenarios in which the prediction horizon and past history vary.

To make a more extensive comparison and to see the behavior of these data with different prediction models, it has been decided to use a total of six models, three of which are considered machine learning and the other three deep learning models. The models chosen were as follows: linear regression (LR), random forest (RF), XGBoost (XGB), long-short-term memory network (LSTM), temporal convolutional network (TCN), and multi-layer perceptron (MLP). The performance of these models is evaluated using the metrics MSE, RMSE, MAE, WAPE, and MAPE.

The proposed methodology aims to establish an experimental framework for the prediction of energy supply and demand in Iraq. The following sections will present the results separately for the two-time series analyzed in this study, beginning with the demand and then the supply.

4.1. Energy demand forecasting

Table 7 provides results for forecast horizon 24 h and demand dataset. The results are divided into two main categories: more classical machine learning algorithms and deep learning models. Models were benchmarked and compared with respect to past history and five performance metrics. The two best-performing models are linear regression for machine learning models and TCN for deep learning. The first obtains MAE value of 450 and MAPE of 0.025% for linear regression, while the second gets MAE value of 493 and MAPE of 0.026%. Regarding the models with the worst results, it can be seen that the worst in the first category was obtained with random forest with MAE value of 870 and MAPE of 0.040%. In the second category, MLP was the worst, with a MAE value of 521 and MAPE of 0.028%.

Table 8 provides results for forecast horizon 48 and demand dataset. The two best-performing models are linear regression for machine learning and LSTM for deep learning. The first obtains MAE value of 643 and MAPE of 0.035% while the second gets MAE value of 678 and MAPE of 0.036%. As for the models with the worst results, it can be seen that the worst in the first category was obtained by random forest with MAE value of 1002 and MAPE of 0.048%. In the second category, TCN was the worst, with MAE value of 685 and MAPE of 0.037%.

Table 9 provides the results for the forecast horizon 72 h and demand dataset. The two best-performing models are linear regression for machine learning models and LSTM for deep learning. The first one obtains MAE value of 784 and MAPE of 0.042%, while the second gets MAE value of 815 and MAPE of 0.043%. As for the models with the worst results, it can be seen that the worst in the first category was obtained with random forest with MAE value of 1096 and MAPE of 0.053%. In the second category, MLP was the worst, with MAE value of 849 and MAPE of 0.045%.

Table 7
Forecasting demand with a forecast horizon of 24 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 24 H DEMAND							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
MACHINE LEARNING	LR	24	1083220	581.6717	502.9881	0.026775	0.027288
		48	955023.4	544.9984	463.4439	0.024786	0.025514
		72	902146.3	535.9016	454.1736	0.024341	0.025129
		96	901062.8	534.4753	452.8799	0.024281	0.025125
		120	**893241.5**	**532.5168**	**450.8758**	**0.024165**	**0.025012**
		144	894859.1	535.2261	453.0205	0.02428	0.025122
		168	897856.9	537.4003	455.4497	0.02441	0.025241
	RF	24	**1777895**	**1004.792**	**870.3464**	**0.040312**	**0.040828**
		48	1783794	1008.703	874.9128	0.040556	0.041089
		72	1786500	1012.165	879.3987	0.040759	0.041282
		96	1787945	1014.781	882.0955	0.040885	0.041402
		120	1802335	1020.769	887.2238	0.041098	0.041599
		144	1816157	1026.992	892.7668	0.041359	0.04186
		168	1824805	1028.919	893.9624	0.041408	0.041909
	XGB	24	1405847	797.0603	662.6271	0.030384	0.031097
		48	1420858	**781.875**	**655.0541**	**0.030094**	**0.030448**
		72	**1381567**	783.4709	661.1899	0.030483	0.030734
		96	1465919	816.2758	690.202	0.031836	0.032145
		120	1627843	872.1356	737.7352	0.033911	0.034171
		144	1691088	898.1677	761.9106	0.035062	0.035242
		168	1768986	928.5309	794.6021	0.036452	0.036505
DEEP LEARNING	LSTM	24	1144054	638.9068	545.846	0.028696	0.029133
		48	984984.4	627.0448	541.006	0.027617	0.028323
		72	925205.1	595.9861	507.4458	0.026359	0.027138
		96	**910309.9**	**586.0333**	**501.2808**	**0.025987**	**0.026749**
		120	974763.3	628.0135	530.1516	0.027475	0.028304
		144	953322.1	634.6448	545.7355	0.02779	0.028519
		168	1030154	689.2143	591.5543	0.02997	0.030618
	TCN	24	1152463	651.2944	560.162	0.029504	0.029947
		48	1016793	607.09	512.1052	0.026991	0.027858
		72	**945475.2**	**583.6522**	**493.6367**	**0.026075**	**0.026877**
		96	1007753	611.8488	521.1543	0.027649	0.028704
		120	1280490	811.7644	702.6569	0.035907	0.03724
		144	1015919	649.027	553.7207	0.02878	0.029635
		168	1122628	714.8256	604.2888	0.031649	0.032668
	MLP	24	1210091	678.1158	589.9583	0.030914	0.031547
		48	1027575	621.4995	**521.9852**	**0.027398**	**0.028208**
		72	1016364	642.0646	548.8354	0.028629	0.029414
		96	1089717	710.9572	596.4573	0.030895	0.03187
		120	1011618	625.9169	529.1662	0.027787	0.028696
		144	1046958	665.4122	562.4341	0.029188	0.030082
		168	**1003728**	**620.1257**	522.4953	0.027718	0.028654

Table 10 provides the results for the 168 h forecast horizon and the demand data set. The two best-performing models are linear regression for machine learning models and MLP for deep learning. The first obtains the MAE value of 1123 and MAPE of 0.060%, while the second gets MAE value of 1162 and MAPE of 0.062%. As for the models with the worst results, it can be seen that the worst in the first category was obtained with random forest with MAE value of 1368 and MAPE of 0.068%. In the second category, LSTM was the worst, with MAE value of 1187 and MAPE of 0.063%.

Fig. 7 provides an overview of the previous tables, taking into account the MAE metric. This figure groups the results according to the forecast horizon, showing a comparison between each of the models used in the experimentation. It is interesting to note at first glance how, as the prediction horizons increase, the results get worse for all models. This decrease may be mainly due to the fact that higher prediction horizons imply greater difficulty in prediction. On the basis of this figure, it seems that the linear regression model outperformed the other models, with the lowest MAE across all four forecast horizons. However, random forest seems to almost always obtain the worst results, although if not only the individual results are taken into account, but also the median of the distribution, the worst model would be RF. Finally, it is worth noting that there is a certain difference in results between the DL and ML models used, the latter being the ones that obtain greater variability between models, which may be due to the fact that DL models in this case are more suitable for capturing the temporal characteristics of the data set.

Table 8
Forecasting demand with a forecast horizon of 48 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 48 H DEMAND							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
MACHINE LEARNING	LR	24	1802788	810.2798	685.2839	0.036443	0.037186
		48	1630950	773.3018	648.7147	0.034592	0.035509
		72	**1592371**	**766.11**	**643.1999**	**0.034326**	**0.035321**
		96	1594239	766.6057	643.9121	0.034368	0.035413
		120	1593819	767.8873	644.3848	0.034365	0.035409
		144	1594905	771.1242	647.1179	0.034486	0.035515
		168	1594976	773.5811	649.7903	0.034629	0.035647
	RF	24	2391115	1167.108	**1002.596**	**0.047561**	**0.048195**
		48	2360015	**1165.339**	1003.585	0.047664	0.04828
		72	**2347167**	1170.524	1009.928	0.047902	0.048457
		96	2360748	1177.735	1016.408	0.048168	0.048714
		120	2368809	1183.015	1020.23	0.048316	0.048844
		144	2385238	1185.532	1022.206	0.048425	0.048967
		168	2392329	1189.519	1025.292	0.048561	0.049098
	XGB	24	2205185	1020.789	848.3964	0.039442	0.04015
		48	**2145587**	**994.6335**	**833.9088**	**0.038913**	**0.03917**
		72	2172596	1030.319	869.4493	0.040367	0.040629
		96	2333627	1078.039	912.2416	0.042222	0.042534
		120	2558740	1140.924	962.5131	0.04452	0.044707
		144	2740464	1192.588	1011.527	0.046635	0.046754
		168	2763566	1211.867	1035.271	0.047752	0.04777
DEEP LEARNING	LSTM	24	1880940	871.5793	738.2946	0.038666	0.03924
		48	**1628318**	**805.7887**	**678.4556**	**0.035587**	**0.036495**
		72	1667014	862.0683	728.2714	0.037642	0.038521
		96	1654046	848.2781	719.0411	0.038009	0.039209
		120	1733785	902.0217	780.9808	0.039825	0.040674
		144	1692101	883.8551	751.2145	0.038614	0.039472
		168	1890780	996.885	838.3699	0.043165	0.044227
	TCN	24	1889494	856.2178	723.4069	0.038303	0.039126
		48	1710081	826.531	697.5408	0.036804	0.037784
		72	**1682362**	**817.2671**	**685.4659**	**0.036445**	**0.037561**
		96	1823857	934.5217	775.2336	0.040206	0.041349
		120	1895499	949.2139	823.3499	0.04243	0.043457
		144	1799922	878.4809	740.2786	0.039094	0.040355
		168	2143434	1110.038	936.7238	0.047366	0.048361
	MLP	24	1929685	891.4968	756.4653	0.039866	0.040713
		48	**1689021**	**816.0728**	**680.9173**	**0.036003**	**0.036978**
		72	1725579	858.2358	716.1583	0.037721	0.038835
		96	1761828	841.7797	703.5016	0.037501	0.038748
		120	1766339	870.6981	735.7779	0.038496	0.039554
		144	1807041	909.212	772.0895	0.040092	0.041122
		168	1819894	912.3422	760.1011	0.039903	0.041027

Fig. 8 shows the comparison between the predicted and actual values for the last month of the study. The best model among all the predictions of the experiments served as the basis for the comparison. In this case, the best result was obtained with a prediction horizon of 24 hours with the linear regression model.

The Friedman test has been applied to assess the overall significance of differences among the performances of the six models. The Friedman test is a non-parametric test used to compare three or more matched groups (in this case, models) without assuming that the data follow a specific distribution, and the objective is to determine whether there are significant differences between the groups. In this context, each model provides 28 evaluation values (7 different past values for 4 different forecast windows) from Tables 7 - 10. The Friedman test determined a significant difference in the models' performances, both for MAE and RMSE values, with a p-value equal to $6.49e-24$ and $7.53e-24$, respectively.

Afterwards, the test used for two-to-two comparisons is the Mann-Whitney U test (also known as the Wilcoxon rank-sum test). The p-value is then corrected using the Bonferroni-Dunn method. After analyzing in the comparison tests between the six models, significant differences were found in some but not all comparisons. The presence of significant differences in at least some comparisons indicates that the models are not equivalent in terms of their performance in the circumstances evaluated. This allows us to establish the following ranking in the model's performance: The best-performing model is LR, followed by

Table 9
Forecasting demand with a forecast horizon of 72 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 72 H DEMAND							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
MACHINE LEARNING	LR	24	2343361	969.2559	815.0429	0.04338	0.044146
		48	2194789	939.0349	787.3046	0.041973	0.042938
		72	**2163190**	**934.0767**	**784.3679**	**0.041838**	**0.042889**
		96	2179585	938.2074	787.8978	0.042021	0.043112
		120	2188277	940.9846	789.8374	0.042087	0.043168
		144	2183029	942.9451	791.3409	0.042136	0.043194
		168	2179172	945.0924	793.0812	0.042228	0.043274
	RF	24	2843364	**1286.215**	**1096.141**	**0.052777**	**0.05327**
		48	2784749	1286.537	1099.299	0.052916	0.05339
		72	**2754846**	1286.674	1100.026	0.052894	0.053328
		96	2759947	1290.505	1102.892	0.053005	0.053428
		120	2782343	1297.687	1109.155	0.053253	0.05366
		144	2791838	1300.271	1110.829	0.053346	0.053752
		168	2802719	1303.531	1113.451	0.053483	0.053874
	XGB	24	2719869	1171.211	974.6917	0.045809	0.046326
		48	**2649420**	**1154.008**	**964.7845**	**0.045354**	**0.045535**
		72	2714473	1192.106	1001.543	0.046989	0.047144
		96	2931581	1252.13	1052.446	0.049163	0.049332
		120	3278267	1330.852	1123.34	0.052317	0.052419
		144	3328298	1359.648	1154.938	0.053922	0.053924
		168	3349189	1374.13	1168.922	0.054669	0.054564
DEEP LEARNING	LSTM	24	2384366	1012.584	852.4807	0.044892	0.045506
		48	2346997	1086.683	934.6706	0.047782	0.048505
		72	**2135899**	**968.358**	**815.0538**	**0.042992**	**0.043995**
		96	2191907	1000.249	839.9378	0.044161	0.045158
		120	2452033	1114.831	961.8425	0.049228	0.049978
		144	2525372	1128.269	941.3932	0.049361	0.050675
		168	2314688	1050.92	890.9157	0.046523	0.04742
	TCN	24	2412693	**1007.948**	**843.7587**	**0.044764**	**0.045463**
		48	2359442	1087.277	939.75	0.048301	0.049127
		72	**2326596**	1065.875	916.8954	0.047433	0.048398
		96	2354439	1050.395	891.3047	0.046665	0.047749
		120	2388654	1072.106	920.5687	0.047755	0.048674
		144	2405361	1082.162	908.5797	0.047351	0.048368
		168	2385619	1055.99	901.3585	0.047314	0.048281
	MLP	24	2406789	1018.402	852.6924	0.045053	**0.045732**
		48	2530987	1082.555	924.6539	0.048367	0.049807
		72	**2277146**	1029.653	873.125	0.045628	0.046588
		96	2340006	**1015.634**	**849.3662**	**0.044998**	0.046293
		120	2382516	1052.136	892.1342	0.04671	0.047733
		144	2510326	1123.572	956.0905	0.049549	0.050525
		168	2306154	1020.625	858.7795	0.045353	0.046381

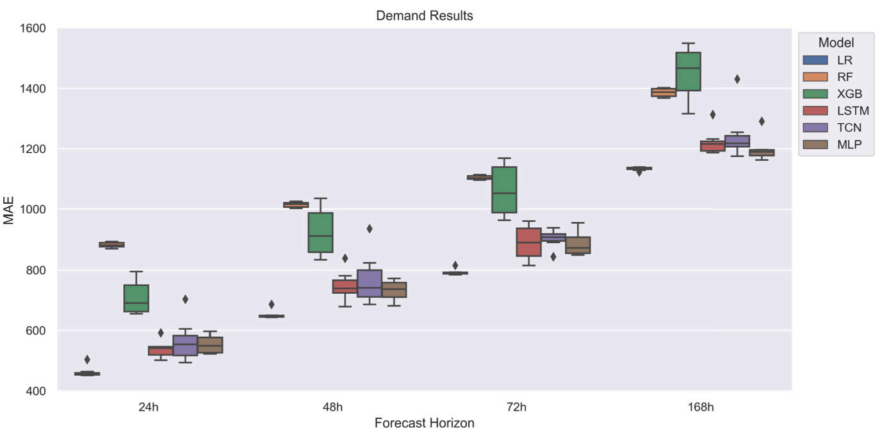

Fig. 7. Distribution of results in terms of MAE grouped by forecast horizon.

Table 10
Forecasting demand with a forecast horizon of 168 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 168 H DEMAND							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
		24	3800349	1383.785	1136.715	0.060547	0.0614108
		48	**3729226**	**1368.163**	**1123.062**	**0.059875**	**0.060936**
		72	3751450	1372.508	1128.569	0.060209	0.061336
	LR	96	3801594	1382.574	1137.932	0.060724	0.061882
		120	3806957	1385.469	1139.332	0.060763	0.061900
		144	3786811	1384.241	1136.500	0.060546	0.061643
		168	3779494	1384.868	1136.100	0.060512	0.061589
		24	4132224	**1645.646**	**1368.324**	**0.067707**	**0.068097**
		48	4132224	1645.646	1368.324	0.067707	0.068097
MACHINE LEARNING		72	**4124540**	1658.615	1379.626	0.068147	0.068412
	RF	96	4148383	1666.749	1387.417	0.068480	0.068708
		120	4185005	1676.267	1395.212	0.068866	0.069091
		144	4228470	1685.671	1401.884	0.069175	0.069382
		168	4228470	1685.671	1401.884	0.069175	0.069382
		24	**4238792**	**1608.761**	**1316.703**	**0.063256**	**0.063539**
		48	4455385	1664.261	1363.103	0.065114	0.065111
		72	4673260	1724.463	1422.586	0.067710	0.067614
	XGB	96	4894699	1770.473267	1466.999	0.069817	0.069655
		120	5111436	1819.981	1507.251	0.072056	0.071793
		144	5177563	1846.604	1530.110	0.073505	0.073266
		168	5288407	1872.017	1549.172	0.074731	0.074559
		24	3844556	1444.743	1192.793	0.062690	0.063439
		48	3793894	1443.222	**1187.218**	**0.062465**	**0.063381**
		72	**3744980**	1473.913	1231.977	0.063936	0.064635
	LSTM	96	3834768	1467.208	1215.460	0.063787	0.064650
		120	3866136	**1437.989**	1192.848	0.063318	0.064419
		144	4082703	1586.656	1313.151	0.068273	0.069206
		168	3809797	1476.846	1216.722	0.064437	0.065349
		24	**3832966**	1462.073	1203.622	0.063298	0.063888
		48	4227610	1670.172	1430.471	0.072372	0.072844
DEEP LEARNING		72	3868559	**1421.634**	**1175.009**	**0.062504**	**0.063839**
	TCN	96	4098126	1470.115	1217.771	0.064670	0.066255
		120	3998622	1456.357	1209.385	0.064342	0.065665
		144	4040086	1494.787	1231.038	0.065116	0.066609
		168	4233132	1496.094	1254.320	0.066543	0.068311
		24	3984785	1424.518	1184.110	0.062992	0.064197
		48	**3785355**	1412.873	**1162.771**	**0.061658**	**0.062676**
		72	3917988	**1408.398**	1169.994	0.062392	0.063881
	MLP	96	3961381	1440.025	1189.306	0.063348	0.064654924
		120	3982736	1438.945	1195.718	0.063720	0.065118
		144	4174662	1573.977	1290.176	0.067576	0.068745
		168	3938162	1445.909	1197.005	0.063666	0.064913

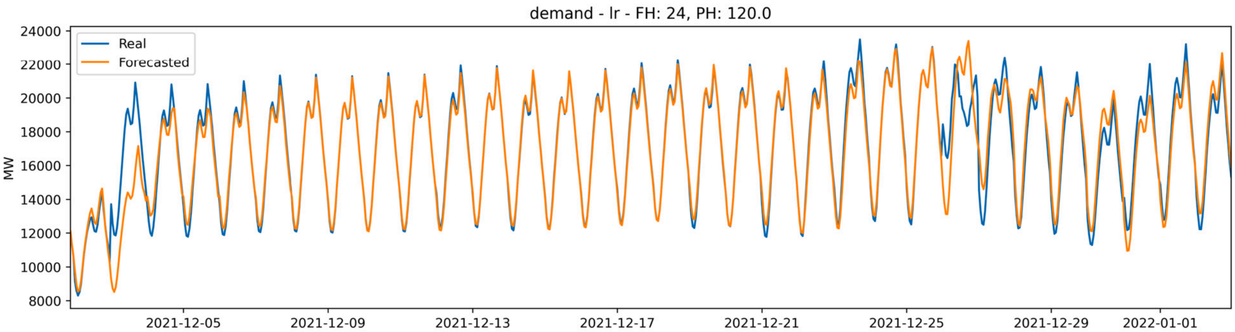

Fig. 8. Comparison between actual and predicted values for demand in the last month of the data set. The observed results are those obtained with the best model.

LSTM, TCN, and MLP, which are considered equivalent. The third place is for XGB and last place (worst performing model) for RF.

An overfitting analysis has also been carried out using learning curves, for the best model in the ranking. Learning curves provide insights into how the classifier's performance evolves as the training set size increases. By plotting the model's training and validation metrics against the number of training instances, it is possible to identify if overfitting or underfitting is occurring.

In Fig. 9, the learning curves for the best experiment have been represented, corresponding to LR model, for 24 hours forecast prediction and 120 hours for past history. X-axis represents the iteration number, where more than 20000 increasing size subsets have been used. Y-axis represents the normalized MAE metric values for both train and validation subsets. As it can be derived from this figure, the convergence of both curves, as well as the level in which it occurs indicate a situation of a well-generalized model, indicating that it generalizes well to unseen data.

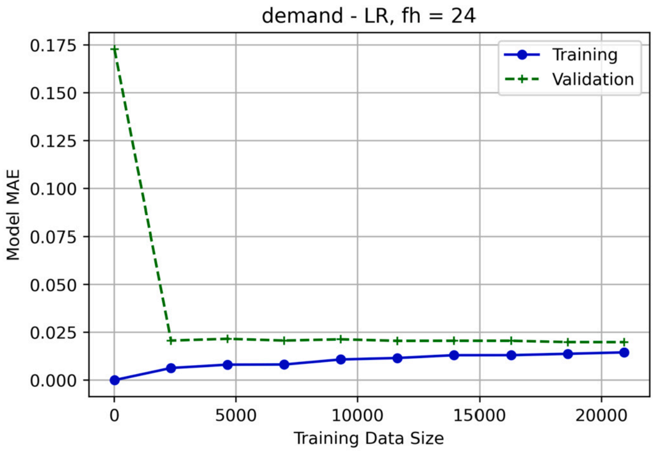

Fig. 9. Learning curves for Linear Regression (24 h Forecast and 120 h Past history).

4.2. Energy supply forecasting

Table 11 provides results for forecast horizon 24 and the supply dataset. The two best-performing models are XGBoost for machine learning models and LSTM for deep learning. The first obtains MAE value of 509 and MAPE of 0.035% while the second gets MAE value of 541 and MAPE of 0.040%. As for the models with the worst results, it can be seen that the worst in the first category was obtained with random forest with MAE value of 604 and MAPE of 0.042%. In the second category, TCN was the worst, with MAE value of 556 and MAPE of 0.041%.

Table 12 provides results for forecast horizon 48 and the supply dataset. The two best-performing models are XGBoost for machine learning models and LSTM for deep learning. The first one obtains MAE value of 588 and MAPE of 0.041% while the second gets MAE value of 644 and MAPE of 0.048%. As for the models with the worst results, it can be seen that the worst in the first category was obtained with random forest with MAE value of 690 and MAPE of 0.049%. In the second category, MLP was the worst, with MAE value of 658 and MAPE of 0.049%.

Table 13 provides results for forecast horizon 72 and the supply dataset. The two best-performing models are XGBoost for machine learning models and LSTM for deep learning. The first obtains MAE value of 637 and MAPE of 0.045% while the second gets MAE value of 704 and MAPE of 0.053%. As for the models with the worst results, it can be seen that the worst in the first category was obtained with random forest with MAE value of 747 and MAPE of 0.053%. In the second category, TCN was the worst, with MAE value of 736 and MAPE of 0.054%.

Table 14 provides results for forecast horizon 168 and the supply dataset. The two best-performing models are XGBoost for machine learning models and TCN for deep learning. The first obtains MAE value of 875 and MAPE of 0.061% while the second gets MAE value of 917 and MAPE of 0.069%. As for the models with the worst results, it can be seen that the worst in the first category was obtained with random forest with MAE value of 914 and MAPE of 0.066%. In the second category, LSTM was the worst, with MAE value of 923 and MAPE of 0.069%.

From Fig. 10, it can be noted that the behavior of the results compared to the forecast horizons follows the same pattern as the results for demand: the longer the forecasting horizon, the worse the results obtained. Overall, the XGBoost model is apparently the one that obtains the best results in the four horizons; however, it is also the one that obtains the worst results, so it would not be the most appropriate model to be considered. If the behavior of LR is observed, it could be determined that it is indeed one of the best since it is the second one, and also that its variability of results is very low. As for the deep learning models, it should be noted that the LSTM model shows the best performance from the TCN and MLP models on four forecast horizons, while

Table 11
Forecasting supply with a forecast horizon of 24 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 24 H SUPPLY							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
MACHINE LEARNING	LR	24	943822.6	682.9484	548.6101	0.039596	0.041359
		48	854748.4	656.4009	529.6794	0.038292	0.040064
		72	831635.4	647.6702	523.407	0.037798	0.03954
		96	820456.9	638.5702	516.3807	0.037298	0.039062
		120	**816745.1**	**638.2175**	**515.9971**	**0.037275**	**0.039056**
		144	819814.3	639.7692	517.4634	0.037359	0.03916
		168	823071.6	640.5274	518.431	0.037424	0.039225
	RF	24	936778.7	720.2302	609.3076	0.041641	0.043396
		48	911437.6	**714.058**	**604.6331**	**0.041255**	**0.04298**
		72	909258.2	714.9388	606.2348	0.04132	0.043046
		96	912878.9	717.2787	608.4985	0.041448	0.043175
		120	**906365.4**	716.8146	608.2583	0.041437	0.04316
		144	926628.9	726.4942	617.9252	0.041998	0.043731
		168	947477.4	736.452	627.9886	0.042581	0.044317
	XGB	24	920911	680.0305	572.2517	0.03745	0.038856
		48	853411.6	623.476	514.1702	0.034285	0.035721
		72	**848551.3**	**619.2674**	**509.2676**	**0.033901**	**0.035286**
		96	889692.3	644.1775	533.3848	0.035185	0.036549
		120	930461.8	673.465	561.4321	0.036742	0.038075
		144	1126261	762.759	650.1283	0.041758	0.043061
		168	1330560	832.6375	718.9174	0.045652	0.046923
DEEP LEARNING	LSTM	24	961336.3	711.7715	585.0637	0.041599	0.043224
		48	869003.5	681.9467	558.2548	0.039909	0.041645
		72	845333	668.4855	547.3058	0.039181	0.040903
		96	889901.9	721.1263	606.2318	0.042553	0.044121
		120	**845149.4**	**663.6271**	**541.9016**	**0.038831**	**0.040648**
		144	849592.6	670.0827	550.0086	0.039417	0.041312
		168	859016.4	678.8035	557.0241	0.039803	0.041664
	TCN	24	976933.6	715.4326	578.8704	0.041497	0.043098
		48	890542.8	690.6279	563.1333	0.040469	0.042169
		72	**868605.1**	**680.0923**	**556.3324**	**0.03991**	**0.041587**
		96	992228.4	800.5181	671.4391	0.047006	0.048472
		120	894916.8	694.1565	568.8765	0.040908	0.042804
		144	962084.5	722.8193	597.1552	0.042994	0.045177
		168	882630.6	700.7717	572.8522	0.04101	0.042772
	MLP	24	999371.6	709.0023	571.7159	0.0412	0.043071
		48	916344.3	715.2784	586.7975	0.041743	0.043344
		72	881194.1	686.6265	559.9553	0.040167	0.041863
		96	1019028	816.0222	696.2365	0.048315	0.049684
		120	**860946.4**	**678.2794**	552.659	**0.039685**	**0.041476**
		144	876239.9	680.7979	**552.2707**	0.03978	0.041671
		168	991120.9	800.4314	682.1794	0.047459	0.048965

the performance of MLP is generally weaker than the other models. Finally, and as was the case for demand, when examining the median of each of the models, the worst of them is again RF, so it is definitely not a model to be taken into account in this type of data.

Fig. 11 shows the comparison between predicted and actual values for the last month of the study. The best model among all the experiments' predictions served as the basis for the comparison. In this case, the best result was obtained with a prediction horizon of 24 hours with the XGBoost model.

As in the case of the demand dataset, the Friedman test has been applied to assess the overall significance of differences among the performance of the six models, using both MAE and RMSE values in Tables 11 - 14. The Friedman test determined a significant difference in the models' performances, both for MAE and RMSE values, with a p-value equal to $4.76e-10$ and $1.21e-8$, respectively. Nevertheless, when performing a two-by-two comparison analysis, no significant differences were found in any case. Therefore, it is not possible to establish any ranking for the models in this dataset. This may be due to a correction for multiple comparisons: When applying the Bonferroni correction for two-to-two comparisons, the significance threshold becomes more stringent. This means that differences must be more pronounced to reach significance in individual comparisons. In addition, variability in data can influence the ability to detect significant differences. If the data are highly variable, it is more difficult to detect differences with confidence.

Table 12
Forecasting supply with a forecast horizon of 48 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 48 H SUPPLY							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
MACHINE LEARNING	LR	24	1244022	819.7863	653.5809	0.046983	0.049411
		48	1145044	794.0604	635.2987	0.045684	0.048108
		72	1115437	781.5335	625.5649	0.044939	0.047342
		96	**1104333**	**775.0649**	**620.2901**	**0.044556**	**0.046991**
		120	1105343	776.5651	621.3151	0.04462	0.047089
		144	1110431	778.8851	623.2781	0.044738	0.047224
		168	1117514	782.0579	626.3575	0.044982	0.047461
	RF	24	1181664	837.5823	693.6875	0.047613	0.049963
		48	1153721	832.6733	691.3552	0.047338	0.049669
		72	1145203	**830.2626**	**690.2974**	**0.047243**	**0.049563**
		96	**1141658**	830.7579	691.4028	0.047296	0.049604
		120	1162454	843.7208	704.774	0.048065	0.050373
		144	1185969	853.3672	713.952	0.048603	0.050931
		168	1197823	858.2774	718.537	0.048904	0.051243
	XGB	24	1113984	763.8137	623.4591	0.041514	0.043404
		48	1096336	**728.0057**	**588.7542**	**0.039579**	0.041497
		72	**1087248**	732.5203	593.9985	0.039607	**0.041431**
		96	1119461	750.7307	611.9357	0.040554	0.042316
		120	1277231	828.8156	682.5908	0.044447	0.046136
		144	1581587	933.7476	785.7593	0.050297	0.051947
		168	1710972	977.5866	831.7676	0.053021	0.054621
DEEP LEARNING	LSTM	24	1239675	832.9388	670.3175	0.047879	0.050213
		48	1147729	815.9227	663.9738	0.047168	0.049428
		72	1114197	**798.9838**	**644.5757**	**0.045971**	**0.048328**
		96	1116197	812.0108	663.9557	0.046809	0.049095
		120	**1111901**	804.6887	654.5242	0.046503	0.048903
		144	1119775	802.4375	647.2431	0.046248	0.048832
		168	1172338	849.6089	698.6963	0.049284	0.051699
	TCN	24	1351519	867.2089	694.9586	0.049885	0.052619
		48	1199993	820.1698	656.6393	0.047286	0.04985
		72	1199555	846.6304	682.5513	0.048615	0.050991
		96	**1143465**	815.1444	661.3401	0.047118	0.049416
		120	1157356	**808.7868**	**650.0604**	**0.046682**	**0.04928**
		144	1223935	878.3085	721.7542	0.050853	0.053133
		168	1216965	834.9221	671.432	0.04827	0.051077
	MLP	24	1284083	852.4871	684.5681	0.04891	0.051253
		48	1183657	820.8609	658.5483	**0.047181**	**0.049554**
		72	1200489	836.1209	673.3013	0.048158	0.050697
		96	1177896	**819.4739**	**658.1074**	0.047239	0.049872
		120	**1152818**	835.4841	687.072	0.048575	0.050837
		144	1164851	826.4666	669.9477	0.047813	0.050314
		168	1165627	836.4939	687.7874	0.048701	0.050978

In this sense, higher variability can be observed for both measures than in the case of the demand dataset, which may justify this result.

As well as in the previous section, an overfitting analysis has also been carried out to check overfitting for the best performing model. In this case, as it has not been possible to obtain a ranking on the models, XGB has been selected based on the best results reported in Tables 11 to 14.

The learning curves in Fig. 12 correspond to the XGB model, for 24 hours forecast prediction and 72 hours of past history. X-axis represents the iteration number, where more than 20000 increasing size subsets have been used. Y-axis represents the normalized MAE metric values for both train and validation subsets. Similar to what occurred in the LR model for demand, it is possible to appreciate that the convergence of both curves, as well as the level in which it occurs indicate a situation of a well-generalized model, indicating that it generalizes well to unseen data.

In both demand and supply forecasting results, the TCN models stand out as the most complex and time-intensive deep learning models due to their specific architecture and extended training times. The LSTM models exhibit moderate complexity with shorter training durations, while the MLP models are the simplest with the briefest training periods. In the machine learning models, Random Forest and XGBoost are more complex than Linear Regression due to their ensemble nature. XGBoost, in particular, requires the most extensive training time, while Linear Regression remains the simplest and least time-intensive model [39].

Table 13
Forecasting supply with a forecast horizon of 72 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 72 H SUPPLY							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
MACHINE LEARNING	LR	24	1459621	917.8376	728.2289	0.052129	0.055003
		48	1349595	889.7269	707.3645	0.050631	0.053513
		72	1318611	877.6987	697.0462	0.049844	0.052717
		96	**1311857**	**874.2694**	**693.4075**	**0.049578**	**0.052495**
		120	1314467	876.2826	694.6138	0.04965	0.052601
		144	1322183	879.8804	697.8358	0.049887	0.052848
		168	1330983	884.3657	702.0588	0.050221	0.053182
	RF	24	1352087	920.0064	757.0996	0.051925	0.054655
		48	1315155	911.452	750.8071	0.051425	0.054149
		72	**1292454**	**905.7166**	**747.1924**	**0.051165**	**0.053857**
		96	1301596	913.251	755.503	0.051634	0.054319
		120	1308330	919.3112	761.9553	0.052007	0.0547
		144	1331951	928.689	769.8703	0.052526	0.055255
		168	1351993	937.0027	778.0192	0.053071	0.055811
	XGB	24	1251767	828.8964	668.8187	0.044822	0.047038
		48	**1229959**	**796.9199**	**637.3623**	**0.042985**	**0.0452**
		72	1238112	809.4034	648.0222	0.043332	0.045434
		96	1351185	864.8413	700.6358	0.046205	0.048209
		120	1561201	946.273	778.882	0.050585	0.052517
		144	1819359	1028.916	863.5504	0.055476	0.057337
		168	1924770	1066.952	901.6436	0.057941	0.059786
DEEP LEARNING	LSTM	24	1440086	925.863	741.9356	0.052712	0.055528
		48	1349133	920.3989	743.861	0.052567	0.055219
		72	1313311	**884.6562**	**704.6068**	0.050401	0.053441
		96	**1307586**	886.2141	705.702	**0.050347**	**0.053389**
		120	1328841	919.308	743.2232	0.052325	0.055121
		144	1389613	914.9723	726.4176	0.052202	0.055626
		168	1380168	911.5681	723.7208	0.051795	0.055152
	TCN	24	1496213	937.1817	744.113	0.053266	0.056215
		48	1395739	936.4147	759.4592	0.053741	0.05637
		72	1370337	922.4636	744.8928	0.052847	0.055539
		96	**1338450**	**911.5703**	**736.5898**	**0.05221**	**0.054901**
		120	1358225	918.0253	737.0704	0.052376	0.055321
		144	1508831	1030.438	862.2694	0.059845	0.062209
		168	1551358	991.3788	793.2364	0.056511	0.060043
	MLP	24	1503186	944.8389	752.1997	0.053706	0.056611
		48	1703893	1056.063	865.3154	0.061152	0.064686
		72	**1354636**	913.4916	737.1891	0.052306	0.055025
		96	1390850	912.2988	725.584	0.051913	0.0551
		120	1368896	**900.2811**	**714.2753**	**0.051192**	**0.054365**
		144	1491865	966.5842	769.5187	0.054835	0.058233
		168	1363234	922.2897	742.4427	0.052692	0.055504

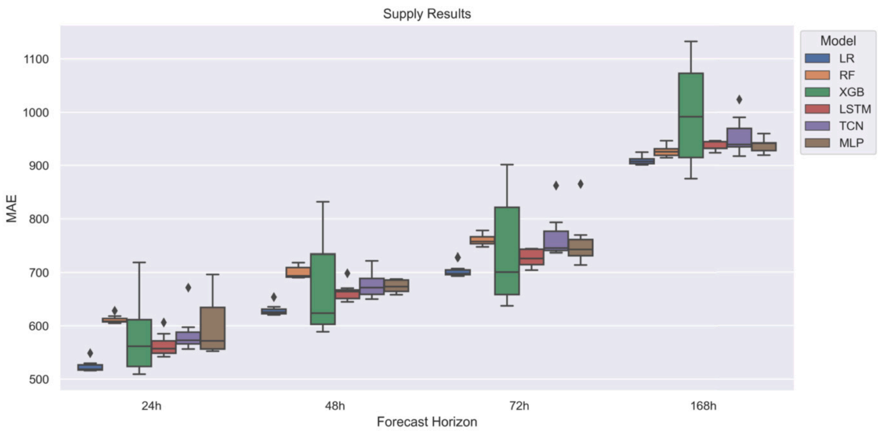

Fig. 10. Distribution of results in terms of MAE grouped by forecast horizon.

Table 14
Forecasting supply with a forecast horizon of 168 h and different past history scenarios. The optimal outcomes of each model have been emphasized in bold.

FORECAST 168 H SUPPLY							
	Model	Past History	MSE	RMSE	MAE	WAPE	MAPE
MACHINE LEARNING	LR	24	2111994	1178.135	924.9979	0.065506	0.069584
		48	2004668	1156.202	907.3498	0.064259	0.068353
		72	**1983148**	**1151.655**	901.7935	0.063850	0.067953
		96	1983140	1151.924	**901.0131**	**0.063807**	**0.067949**
		120	1990370	1155.811	904.9072	0.064097	0.068264
		144	2001700	1160.808	910.0048	0.064492	0.068663
		168	2014099	1165.357	914.4302	0.064825	0.068973
	RF	24	1878323	1140.670	925.6115	0.063599	0.067281
		48	**1820957**	**1127.581**	**914.5916**	**0.062863**	**0.066571**
		72	1822555	1129.542	916.7100	0.063006	0.066728
		96	1834655	1135.455	921.1788	0.063313	0.067062
		120	1850609	1141.128	926.1666	0.063661	0.067452
		144	1887854	1153.377	936.6010	0.064398	0.068240
		168	1922642	1164.195	946.4780	0.065113	0.068980
	XGB	24	**1898801**	**1107.076**	**875.2318**	**0.058658**	**0.061613**
		48	2020675	1131.595	894.0466	0.059645	0.062560
		72	2172103	1173.007	935.8090	0.061912	0.064725
		96	2364242	1225.931	991.7233	0.065135	0.067864
		120	2537484	1276.717	1048.054	0.068511	0.071154
		144	2694297	1323.663	1097.999	0.071640	0.074248
		168	2802252	1359.890	1132.752	0.073961	0.076543
DEEP LEARNING	LSTM	24	2078655	1172.615	**923.7048**	**0.065356**	0.069400
		48	1978771	1176.086	946.2766	0.066153	0.069894
		72	**1947589**	**1163.194**	932.0106	0.065374	**0.069191**
		96	1952236	1174.203	944.0125	0.065921	0.069751
		120	1970424	1176.957	945.3444	0.066316	0.070216
		144	1971948	1171.079	932.0645	0.065363	0.069491
		168	1974833	1171.558	932.2214	0.065447	0.069655
	TCN	24	2124284	1189.333	939.2398	0.066419	0.070406
		48	2351414	1282.785	1023.559	0.072261	0.077167
		72	2037952	1184.804	932.4838	0.066044	0.070354
		96	**2032557**	**1171.729**	**917.6052**	**0.065124**	**0.069556**
		120	2093140	1195.882	937.6273	0.066585	0.071220
		144	2248696	1254.329	990.2022	0.070215	0.075211
		168	2106936	1206.322	948.6727	0.067430	0.072046
	MLP	24	2169926	1201.310	942.1428	0.066817	0.071120
		48	2117168	1199.680	942.4788	0.066770	0.071246
		72	**1980106**	**1164.668**	**919.4992**	**0.065004**	**0.069072**
		96	2030036	1175.449	924.0086	0.065447	0.069705
		120	2000592	1177.180	931.6234	0.065815	0.069955
		144	2113420	1215.333	959.9425	0.067916	0.072312
		168	2079056	1197.226	941.4677	0.066823	0.071343

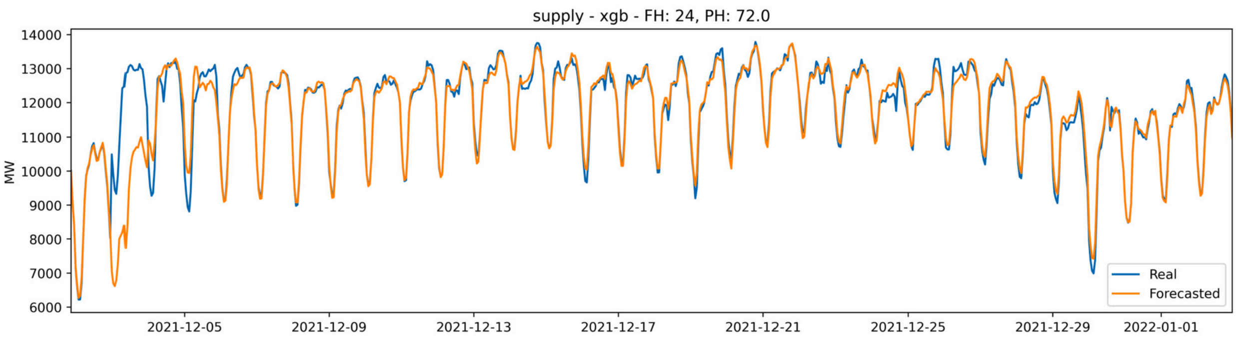

Fig. 11. Comparison between actual and predicted values for supply in the last month of the data set. The observed results are those obtained with the best model.

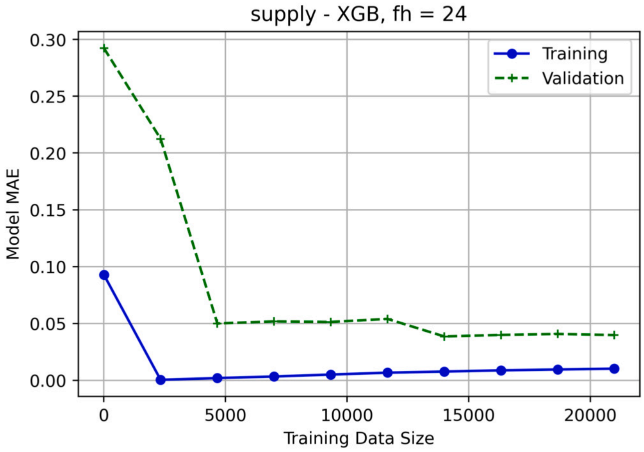

Fig. 12. Learning curves for XGB (24 h Forecast and 72 h Past history).

4.3. Discussions

In the realm of energy forecasting, benchmarking the performance of methodologies against state-of-the-art techniques and ground truth data is essential to evaluate their effectiveness. In this section, a discussion about the performance of the models is illustrated for both the existing state-of-the-art methods and ground truth data.

Comparison with State-of-the-Art: An extensive evaluation is performed on the six forecasting models in comparison to well-established state-of-the-art methods that are commonly utilized in the field of energy forecasting. These benchmark methods encompass a range of approaches, including deep learning and machine learning architectures. Comparison was performed across various forecasting horizons, including 24 hours, 48 hours, 72 hours, and 168 hours, to account for short-term and long-term prediction requirements. The metrics used for benchmarking encompass widely recognized error measures such as MAE, MSE, RMSE, WAPE, and MAPE. These metrics offer a holistic view of forecasting accuracy, accounting for errors of different magnitudes, and providing a balanced assessment of model performance.

Comparison with Ground Truth Data: To further validate the reliability of the forecasting models, a comparison was performed on their predictions with ground truth data obtained from authoritative sources. The ground truth data includes actual energy consumption and supply measurements observed during the evaluation period. This data has not been used before, therefore, it cannot be compared with other previous articles. Using the models involved in this work, the findings confirm that these models exhibit a high degree of accuracy in replicating actual energy demand and supply trends, indicating a strong alignment between predictions and ground truth data.

Finally, the favorable comparison results against state-of-the-art methods and the alignment with ground truth data validate the robustness and efficacy of the proposed forecasting. These findings underscore the potential of this research to significantly enhance the precision and reliability of energy forecasting in the context of the Iraqi energy sector.

5. Conclusions

In conclusion, this study has significantly advanced the understanding of energy demand and supply forecasting in the complex landscape of a liberalized energy market. The study addressed the critical challenge of real-time balance between energy demand and supply in a distributed environment, underscoring the need for continuous model maintenance to ensure reliable forecasts. 4.3 A novel time-series dataset was collected for the years 2019 to 2021, encompassing a range of Iraqi electricity demand and supply values. The study carefully compared different architectural models and how they could be used to predict Iraq's power supply and demand. The study also revealed approaches to improve the accuracy and usefulness of these predictions by optimizing the models' parameters. Using the collected dataset, six prominent models were used in performing the forecasting process, including LSTM, TCN, MLP, LR, XGB, and RF. The performance of these models was rigorously assessed using key metrics such as MAE, MSE, RMSE, WAPE, and MAPE.

Furthermore, the study findings unveiled crucial insights into the realm of demand and supply forecasting. For demand forecasting, LR emerged as the standout performer across multiple forecast horizons, demonstrating its prowess as a machine learning-based model. Also, LSTM showcased its excellence in deep learning-based forecasting for specific horizons, while TCN and MLP displayed their strengths in other contexts. In the realm of supply forecasting, XGB and LSTM led the way, representing the pinnacle of machine learning and deep learning approaches, respectively. Conversely, RF and MLP lagged behind, revealing their limitations in modeling intricate temporal relationships.

Additionally, the results underscored the pivotal role of data preprocessing techniques in shaping forecasting performance. Also, the study illuminated the concentration of the highest electricity demand in Baghdad, driven by factors such as population growth and industrial expansion. While this increase in demand contributed to environmental problems, various regions also saw improvements in energy efficiency. The study holds immense promise in assisting the Iraqi government in tackling energy issues, offering invaluable insights for the selection of the most effective forecasting models. Precise energy demand predictions, as the research highlights, are indispensable for ensuring a stable and dependable energy supply, thereby bolstering economic development and enhancing the well-being of Iraq's population.

As the study anticipates future demand reaching between 30,000 and 35,000 megawatts per day, the ability to predict and manage this demand becomes paramount. Therefore, this research not only addresses Iraq's pressing energy problems but also establishes a robust foundation for future investigations in this domain. These endeavors will undoubtedly contribute to more effective and sustainable energy resource management, propelling Iraq's economic growth and the overall welfare of its people.

As future work, the study aspires to delve into hybrid forecasting models that combine machine learning and deep learning models, aiming to further elevate the accuracy and reliability of energy demand and supply predictions in Iraq. The study also plans to introduce exogenous variables such as weather data and economic indicators into the models to enhance their predictive capabilities. Future research may also look into how different time frames affect model performance and make region-specific forecasting models that are tailored to the specific energy needs of Iraq's provinces. In addition, extending the dataset to include data for two more years will also contribute to a more accurate discussion of the limitations. Finally, these endeavors collectively hold the potential to shape more effective strategies for managing Iraq's energy resources, ultimately fostering the nation's economic prosperity and the well-being of its citizens.

Credits

The icons incorporated in this article were acquired from the Flaticon website designed by Freepik.

CRediT authorship contribution statement

Morteza Aldarraji: Writing – review & editing, Writing – original draft, Methodology, Investigation, Data curation, Conceptualization. **Belén Vega-Márquez:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Methodology, Investigation, Conceptualization. **Beatriz Pontes:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Conceptualization. **Basim Mahmood:** Writing – review & editing, Data curation. **José C. Riquelme:** Supervision, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data used in this study was obtained from the Iraqi government. The participants of this study did not give written consent for their data to be shared publicly, so due to the sensitive nature of the research supporting data is not available.

Acknowledgements

This work was supported by the Ministry of Electricity/Operation and Control Office/Baghdad/Iraq [Contract 22156]; the Andalusian Regional Government [PYC20 RE 078 US]; and by the Spanish Ministry of Science and Innovation [MICINN PID2020-117954RB-C22, MICIN TED2021-131311B].

References

[1] S. Singer, J.-P. Denruyter, D. Yener, The energy report: 100% renewable energy by 2050, in: Towards 100% Renewable Energy: Techniques, Costs and Regional Case-Studies, Springer, 2017, pp. 379–383.
[2] M. Liu, Y. He, H. Zhang, H. Su, Z. Zhang, The feasibility of solar thermal-air source heat pump water heaters in renewable energy shortage regions, Energy 197 (2020) 117189.
[3] Z. Shao, F. Chao, S.-L. Yang, K.-L. Zhou, A review of the decomposition methodology for extracting and identifying the fluctuation characteristics in electricity demand forecasting, Renew. Sustain. Energy Rev. 75 (2017) 123–136.
[4] M.E. El-Hawary, Advances in Electric Power and Energy Systems: Load and Price Forecasting, John Wiley & Sons, 2017.
[5] A. Chehri, R. Saadane, I. Fofana, G. Jeon, Smart grid for sustainable cities: strategies and pathways for energy efficiency solutions, in: Sustainability in Energy and Buildings 2021, Springer, 2022, pp. 317–327.
[6] T.N. Le, W.-L. Chin, D.K. Truong, T.H. Nguyen, M. Eissa, Advanced metering infrastructure based on smart meters in smart grid, in: Smart Metering Technology and Services-Inspirations for Energy Utilities, 2016.
[7] M. Popa, Data collecting from smart meters in an advanced metering infrastructure, in: 2011 15th IEEE International Conference on Intelligent Engineering Systems, IEEE, 2011, pp. 137–142.
[8] E. Herring, Variegated neo-liberalization, human development and resistance: Iraq in global context, Int. J. Contemp. Iraqi Stud. 5 (3) (2012) 337–355.
[9] K. Shaukat, T.M. Alam, M. Ahmed, S. Luo, I.A. Hameed, M.S. Iqbal, J. Li, M.A. Iqbal, A model to enhance governance issues through opinion extraction, in: 2020 11th IEEE Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), IEEE, 2020, pp. 0511–0516.

[10] J.S. Roungkvist, P. Enevoldsen, Timescale classification in wind forecasting: a review of the state-of-the-art, J. Forecast. 39 (5) (2020) 757–768.
[11] F. Lucas, P. Costa, R. Batalha, D. Leite, I. Škrjanc, Fault detection in smart grids with time-varying distributed generation using wavelet energy and evolving neural networks, Evolv. Syst. 11 (2020) 165–180.
[12] M. Ibrar, M.A. Hassan, K. Shaukat, T.M. Alam, K.S. Khurshid, I.A. Hameed, H. Aljuaid, S. Luo, A machine learning-based model for stability prediction of decentralized power grid linked with renewable energy resources, Wirel. Commun. Mob. Comput. 2022 (2022) 1–15.
[13] M.R. Kumar, S. Vekkot, S. Lalitha, D. Gupta, V.J. Govindraj, K. Shaukat, Y.A. Alotaibi, M. Zakariah, Dementia detection from speech using machine learning and deep learning architectures, Sensors 22 (23) (2022) 9311.
[14] E. Crisostomi, C. Gallicchio, A. Micheli, M. Raugi, M. Tucci, Prediction of the Italian electricity price for smart grid applications, Neurocomputing 170 (2015) 286–295.
[15] E. Mocanu, P.H. Nguyen, W.L. Kling, M. Gibescu, Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning, Energy Build. 116 (2016) 646–655.
[16] K. Shaukat, S. Luo, V. Varadharajan, I.A. Hameed, M. Xu, A survey on machine learning techniques for cyber security in the last decade, IEEE Access 8 (2020) 222310–222354.
[17] M. Valipour, M.E. Banihabib, S.M.R. Behbahani, Parameters estimate of autoregressive moving average and autoregressive integrated moving average models and compare their ability for inflow forecasting, J. Math. Stat. 8 (3) (2012) 330–338.
[18] C. Dittmer, J. Krümpel, A. Lemmer, Power demand forecasting for demand-driven energy production with biogas plants, Renew. Energy 163 (2021) 1871–1877.
[19] M. Kim, W. Choi, Y. Jeon, L. Liu, A hybrid neural network model for power demand forecasting, Energies 12 (5) (2019) 931.
[20] R. Porteiro, L. Hernández-Callejo, S. Nesmachnow, Electricity demand forecasting in industrial and residential facilities using ensemble machine learning, Rev. Fac. Ing. Univ. Antioq. 102 (2022) 9–25.
[21] C.E. Velasquez, M. Zocatelli, F.B. Estanislau, V.F. Castro, Analysis of time series models for Brazilian electricity demand forecasting, Energy 247 (2022) 123483.
[22] F. Pallonetto, C. Jin, E. Mangina, Forecast electricity demand in commercial building with machine learning models to enable demand response programs, Energy AI 7 (2022) 100121.
[23] A. Banga, S. Sharma, Electricity demand forecasting models at hourly and daily level: a comparative study, in: 2022 International Conference on Advanced Computing Technologies and Applications (ICACTA), IEEE, 2022, pp. 1–5.
[24] M.A. Raza, K.L. Khatri, A. Israr, M.I.U. Haque, M. Ahmed, K. Rafique, A.S. Saand, Energy demand and production forecasting in Pakistan, Energy Strat. Rev. 39 (2022) 100788.
[25] M. Jaramillo, S. Llamuca, A proposed model for electricity demand forecasting in Ecuador considering Akaike criterion, in: Communication, Smart Technologies and Innovation for Society: Proceedings of CITIS 2021, Springer, 2021, pp. 345–355.
[26] R.Y. Tse, An application of the arima model to real-estate prices in Hong Kong, J. Prop. Finance 8 (2) (1997) 152–163.
[27] Y. Liu, D. Li, S. Wan, F. Wang, W. Dou, X. Xu, S. Li, R. Ma, L. Qi, A long short-term memory-based model for greenhouse climate prediction, Int. J. Intell. Syst. 37 (1) (2022) 135–151.
[28] Y.A. Farha, J. Gall, Ms-tcn: multi-stage temporal convolutional network for action segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3575–3584.
[29] S. Bai, J.Z. Kolter, V. Koltun, An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, arXiv preprint arXiv:1803.01271, 2018.
[30] C. Voyant, G. Notton, S. Kalogirou, M.-L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: a review, Renew. Energy 105 (2017) 569–582.
[31] I. Ilic, B. Görgülü, M. Cevik, M.G. Baydoğan, Explainable boosted linear regression for time series forecasting, Pattern Recognit. 120 (2021) 108144.
[32] T. Chen, C. Guestrin, Xgboost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794.
[33] X. Liao, N. Cao, M. Li, X. Kang, Research on short-term load forecasting using xgboost based on similar days, in: 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS), IEEE, 2019, pp. 675–678.
[34] K. Manish, M. Thenmozhi, Forecasting stock index movement: a comparison of support vector machines and random forest, in: Proceedings of Ninth Indian Institute of Capital Markets Conference, Mumbai, India, 2005, p. 16.
[35] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.
[36] D. Chicco, M.J. Warrens, G. Jurman, The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation, PeerJ Comput. Sci. 7 (2021) e623.
[37] M. Borowski, K. Zwolińska, Prediction of cooling energy consumption in hotel building using machine learning techniques, Energies 13 (23) (2020) 6226.
[38] Addressing energy challenges in Iraq: Forecasting power supply and demand for sustainable development, https://github.com/bvegaus/energy-irak.
[39] K. Shaukat, S. Luo, S. Chen, D. Liu, Cyber threat detection using machine learning techniques: a performance evaluation perspective, in: 2020 International Conference on Cyber Warfare and Security (ICCWS), IEEE, 2020, pp. 1–6.