Comparative Analysis of Time Series Forecasting Models for Hourly Energy Demand in Turkey

F. Alhaj Omar[1]
[1]Department of Electric and Energy, Zonguldak Bulent Ecevit University, 67100 Zonguldak, Turkey
*corresponding author's email: fuad.a@beun.edu.tr

Abstract – *This study conducts an in-depth comparative assessment of five-time series forecasting methods—AutoRegressive (AR), Moving Average (MA), AutoRegressive Moving Average (ARMA), AutoRegressive Integrated Moving Average (ARIMA), and Long Short-Term Memory (LSTM) networks, applied to hourly electricity demand data from Turkey. The objective is to determine which model delivers the highest level of predictive precision and contextual relevance. Model performance was assessed through three key indicators: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE). The results demonstrate that LSTM exhibits the most favorable accuracy, recording the lowest error values (MAE: 389.21, RMSE: 550.98, MAPE: 1.02%) on the test dataset. This superior performance reflects LSTM's ability to capture nonlinear behaviors and abrupt demand shifts influenced by external and internal system dynamics. In contrast, conventional models such as AR, MA, and ARMA reported significantly higher forecasting errors, with MAPE values exceeding 14%, indicating limited adaptability to complex and variable consumption patterns. These insights position LSTM as a highly effective tool for improving forecasting reliability, supporting real-time operational planning, and informing strategic energy policies.*

Keywords: *ARIMA Models, Energy Demand Forecasting, LSTM Networks, Machine Learning in Energy Systems, Statistical Forecasting Models.*

Article History
Received 27 September 2025
Received in revised form 04 November 2025
Accepted 10 November 2025

I. Introduction

Accurately anticipating future electricity demand has become increasingly vital for maintaining the reliability and performance of modern energy infrastructures. As global power systems undergo rapid transformation, driven by the rise of renewable energy sources and the growing complexity of interconnected grids, there is an increasing demand for forecasting frameworks that are both precise and adaptable. Such forecasts are essential not only for enhancing operational efficiency but also for informing long-term policy decisions and strategic energy planning. They enable utilities to optimise generation schedules, reduce resource wastage, and improve cost-effectiveness. Furthermore, effective demand prediction supports the delicate balance between supply and consumption, contributing to the stability and resilience of the broader energy ecosystem [1, 2].

In recent years, increasing grid decentralization, the proliferation of distributed generation technologies, and the emergence of electric vehicles have added further complexity to demand-side behavior, necessitating forecasting tools that can dynamically adapt to multi-scalar and nonlinear data characteristics. Studies such as Deb et al. [1] and Zhang et al. [2] have emphasized the critical role of context-aware, data-driven models in capturing short-term fluctuations and long-term consumption trends in modern energy systems.

From an environmental perspective, energy demand forecasting acquires a critical role in facilitating sustainable energy management. By providing accurate predictions, it is possible to optimize the operation of power plants, reduce fuel consumption, minimize overproduction, and consequently decrease environmental pollutants and greenhouse gas emissions [3]. This aspect of forecasting is increasingly significant as countries commit to international climate targets and shift towards more sustainable, low-carbon energy systems [4].

The context of Turkey, which serves as the geographical and analytical locus for this study, presents a particularly interesting case. Straddling Europe and Asia, Turkey has a rapidly expanding energy market. This market is driven by a growing population, an increasing rate of urbanization, and substantial investments in renewable energy. The characteristics of Turkey's energy market, including its reliance on both conventional and renewable energy sources and the dynamic nature of its energy consumption patterns, make it an ideal candidate for examining the effectiveness of various forecasting models [5, 6]. Additionally, the Turkish grid operates under transitional regulatory frameworks and experiences high load volatility during seasonal peaks, creating an ideal testbed for hybrid and intelligent prediction systems that integrate meteorological, socio-economic, and behavioral data streams. Despite the presence of diverse modeling efforts globally, localized calibration of models for emerging markets like Turkey remains limited, opening avenues for context-specific, performance-oriented research.

Classical time series forecasting techniques, such as AR, MA, and ARMA models, have historically been favored for their straightforward formulation and solid performance in modeling linear, stationary datasets. However, their effectiveness significantly diminishes when applied to real-world energy demand scenarios characterized by irregular fluctuations, evolving trends, and seasonal cycles. These limitations necessitate the exploration of more advanced modeling frameworks capable of capturing complex temporal dynamics [7]. Recent works have proposed hybrid architectures, such as ARIMA–LSTM or Prophet–XGBoost combinations, to leverage the interpretability of statistical models and the pattern-recognition strength of machine learning. However, standardized evaluation benchmarks across model families remain underdeveloped.

Despite the proliferation of forecasting models in the literature, most prior studies have focused on individual model implementations or limited forecasting horizons, often using synthetic or low-resolution datasets. There is a notable lack of integrative comparisons that evaluate both Classical Statistical Techniques (CST) and advanced deep learning (DL) models using high-frequency, real-world energy demand data. Furthermore, existing research rarely contextualizes model performance within the operational realities of dynamic and complex energy markets such as Turkey. This study addresses these limitations by presenting a comprehensive side-by-side evaluation of CST and DL models, applied to an extended and granular dataset, thereby bridging an important methodological and practical gap in the field of energy forecasting.

In contrast to traditional statistical models, the ARIMA and LSTM architectures represent more sophisticated forecasting methodologies [8, 9]. ARIMA extends the capabilities of ARMA by incorporating differencing operations, which allow it to effectively handle non-stationary time series, thereby increasing its versatility. On the other hand, LSTM networks, as part of the deep learning family, are particularly adept at capturing prolonged temporal dependencies and intricate structures embedded in sequential data. This makes them highly suitable for scenarios with substantial variability and volatility, such as those frequently encountered in energy demand forecasting [10, 11]. Moreover, with recent advances in attention-based neural networks, such as Transformer architectures, a new frontier has emerged in time series forecasting, showing promise in handling long-range dependencies with greater efficiency than traditional Recurrent Neural Network (RNN)-based models. This shift necessitates a comparative understanding of where LSTM still holds an advantage and where newer models may outperform.

This paper critically assesses these models both quantitatively and visually through the analysis of forecasting curves generated by each model, providing a nuanced understanding of each model's behavior and accuracy. Such comparative analysis is invaluable as it not only elucidates the strengths and limitations of each model but also highlights the unique contributions of advanced computational techniques like LSTM in improving forecasting accuracy.

Through the comparative investigation of forecasting techniques, this research provides a valuable contribution to the ongoing discourse in energy demand modeling. The insights generated herein are intended to support decision-makers, energy system planners, and utility stakeholders in identifying forecasting approaches best suited for managing complexity in modern power markets. By critically evaluating model performance under realistic operational conditions, the study extends the current understanding of model applicability, thereby enriching methodological development and offering strategic guidance for improving demand-side planning and energy resource allocation in future-oriented systems. Finally, this study contributes to the broader body of energy analytics by proposing an empirically grounded framework for comparing forecasting techniques under realistic, high-resolution data conditions, laying the groundwork for future explorations involving attention mechanisms, real-time anomaly detection, and the integration of economic and environmental indicators into prediction pipelines.

II. Literature Review

Accurate energy demand forecasting plays a critical role in the strategic planning and real-time operation of modern power systems. With growing integration of renewable energy sources, increasing electrification across sectors, and heightened variability in consumption behavior, forecasting methodologies must evolve beyond traditional paradigms. In this context, a thorough examination of both classical and modern forecasting approaches is necessary—particularly within complex and dynamic electricity markets such as Turkey's.

Historically, the energy forecasting domain has been dominated by CST, including AR, MA, and ARMA models. These models, rooted in the assumption of linearity and stationarity, have long served as the backbone of short-term forecasting tasks due to their mathematical simplicity, ease of implementation, and interpretability [7, 12]. AR models excel in capturing autoregressive patterns, while MA models effectively account for stochastic noise and short-term fluctuations. Together, ARMA models integrate both aspects, offering balanced performance under stationary conditions. Multiple regression techniques have also been employed, particularly in long-term load forecasting (LTLF), where they have achieved MAPE values as low as 2.6% in well-conditioned datasets [13].

Despite their early success, these CST approaches face considerable challenges in modern contexts. Their reliance on stationary assumptions and limited capacity to model nonlinear or non-periodic behaviors constrain their adaptability to real-world datasets, which are often affected by socio-economic variability, weather-driven demand spikes, and policy-induced transitions [14]. In response to these limitations, researchers have increasingly turned to data-driven methods that offer enhanced flexibility and learning capacity.

Among the prominent modern approaches, the ARIMA model and LSTM networks stand out. ARIMA, by incorporating differencing operations, extends the classical ARMA framework and allows for effective modeling of non-stationary time series, making it suitable for datasets with deterministic trends or temporal drifts [15]. Meanwhile, LSTM, a special class of RNNs, introduces memory cells capable of preserving long-term dependencies, a feature particularly advantageous in capturing complex dynamics across long time horizons [11, 16].

Beyond these individual models, Artificial Neural Networks (ANNs) and standard RNNs have emerged as leading contenders in the machine learning domain. Their ability to approximate nonlinear mappings and temporal sequences has yielded high precision in short-term load forecasting applications. Reported MAPE values are approximately 3.7% for ANNs and 3.6% for RNNs in recent empirical evaluations, validating their practical utility for operational decision-making and grid dispatch planning [17, 18].

Recently, there has been a growing interest in hybrid forecasting frameworks that combine the strengths of CST and ML models. These systems aim to enhance forecasting accuracy by utilizing the structured interpretability of classical methods alongside the adaptive learning capabilities of deep learning techniques. Studies have proposed integrated ARIMA–LSTM and Seasonal Autoregressive Integrated Moving Average-Gated Recurrent Unit (SARIMA–GRU) configurations that benefit from ARIMA's trend-seasonality decomposition and LSTM's sequence learning, demonstrating improved robustness under volatile load conditions [19, 20]. Despite these promising advances, hybrid models still lack standardized testing across diverse geographical and temporal datasets, limiting their transferability.

The case of Turkey's electricity market presents unique modeling challenges and opportunities. The country's energy system is marked by a hybrid generation mix, combining hydroelectric, thermal, wind, and solar sources, alongside increasing interregional transmission complexity and high seasonal load variability. These characteristics necessitate forecasting approaches that can simultaneously account for short-term fluctuations, medium-term planning, and long-term structural changes [21]. Furthermore, external factors such as climate change, energy pricing reforms, and grid modernization initiatives add layers of uncertainty that classical models are ill-equipped to manage.

While some recent contributions have ventured into hybridization or neural network approaches, comprehensive model validation using long-span, high-resolution datasets remains limited, particularly in Turkey's context. Most prior efforts focus on daily or monthly aggregates, leaving hourly-level forecasting underexplored. This study seeks to fill this gap by employing both CST and Deep Learning (DL) models on a national-scale dataset encompassing six years of hourly electricity demand. The comparative evaluation not only benchmarks predictive performance across methodologies but also investigates the operational viability of each model in a real-world, high-dimensional forecasting task.

In summary, the evolution from simple statistical predictors to complex, learning-based systems reflects the growing sophistication of energy markets and the urgent need for adaptive forecasting. This review establishes the intellectual foundation for the empirical work that follows, positioning the study within the broader discourse on data-driven energy analytics and supporting its methodological relevance for future smart grid implementations.

III. Methodology

A. Data Description

This research leverages an extensive high-resolution dataset that records hourly electricity consumption across the Turkish power grid. The dataset was made available by the Turkish Electricity Transmission Corporation (TEİAŞ) and spans from December 2015 to April 2022, covering nearly 55,500 hourly observations. The breadth and granularity of this dataset make it especially valuable for exploring both transient fluctuations and seasonal dynamics in energy demand across different temporal scales. By capturing variations across years, seasons, and weekdays versus weekends, the dataset enables a multi-dimensional assessment of model adaptability and predictive accuracy in realistic operational contexts.

B. Data Collection Methods

The electricity demand data were acquired through a national network of digitally automated metering systems integrated into the transmission infrastructure at regional substations. These advanced metering systems continuously capture consumption metrics in real-time and communicate them to a central data aggregation unit operated by TEİAŞ. To ensure data fidelity and system-wide consistency, the raw meter outputs are standardized into hourly aggregate demand values, representing national electricity load. This digital infrastructure reduces latency and measurement errors, offering a reliable basis for high-resolution time series modeling. Importantly, the use of synchronized metering across diverse climatic zones ensures that regional consumption behaviors are reflected in the overall data, thus increasing the robustness of model generalizability.

C. Sample Size and Variables Measured

The dataset comprises a set of variables that capture essential attributes influencing electricity consumption. These include:

- Hourly Electricity Demand (in MW): The target variable for forecasting, representing the national power consumption per hour.
- Date and Time: Timestamps associated with each entry allow temporal indexing and feature engineering.
- Ambient Temperature: Hourly temperature records were merged from the Turkish State Meteorological Service to account for the well-documented temperature-load dependency.
- Holiday Flag: A binary indicator identifying national holidays, as demand behavior often deviates significantly during public holidays due to commercial and industrial closures.

By combining meteorological and calendar-based variables, the dataset supports the development of both univariate and multivariate forecasting models, facilitating an in-depth exploration of exogenous factor effects on energy demand.

D. Data Preprocessing Steps

In preparation for model training, the dataset underwent rigorous preprocessing to enhance its analytical reliability and ensure compatibility across both statistical and deep learning modeling paradigms.

- Handling Missing Data: Initial screening identified occasional gaps in the hourly readings. To preserve temporal continuity, missing values were imputed using linear interpolation, which maintains time-series trends without introducing artificial variance.
- Normalization and Scaling: Given the sensitivity of neural models such as LSTM to input scales, Z-score Standardization was applied to the energy demand data before feeding it into the models. This process transforms the data to have a mean of zero ($\mu = 0$) and a standard deviation of one ($\sigma = 1$). It is crucial to note that Z-score standardization, unlike scaling methods such as Min-Max, does not impose a fixed and definite range on the standardized values. Thus, the theoretical range of the standardized values is $(-\infty, \infty)$. However, practically, and assuming a near-normal distribution of the data, the vast majority of standardized values (over 99.7%) fall within the practical range of $[-3, 3]$, which effectively facilitates the convergence of the neural network models.
- Stationarity Testing: Since ARIMA and similar models require stationary inputs, Augmented Dickey-Fuller (ADF) tests were performed to assess unit root presence. When non-stationarity was confirmed, first- or second-order differencing was applied based on test statistics and visual inspection of autocorrelation plots.
- Seasonal Decomposition: Energy demand in Turkey exhibits clear intra-daily and seasonal periodicities due to human activity cycles and weather patterns. To better isolate trends, the series was decomposed into trend, seasonal, and residual components using additive decomposition, enabling models to focus on structure-specific behaviors.
- Temporal Feature Engineering: From the datetime stamps, additional cyclical features were generated: hour of day, day of the week, month, and weekday/weekend categorization. These engineered features were particularly beneficial for models designed to capture periodicity without explicitly modeling seasonality.

These preprocessing stages ensured that the dataset was not only cleansed and complete but also analytically enriched for robust learning across a diverse suite of models. The complete dataset was subsequently partitioned into three sequential subsets for chronological forecasting: the training set spanned from December 1, 2015, to December 31, 2020; the validation set covered January 1, 2021, to December 31, 2021; and the final testing set comprised the most recent period from January 1, 2022, to April 30, 2022.

IV. Models Employed

To facilitate a systematic and comparative analysis of forecasting methods, a variety of classical and machine learning models were implemented within a unified computational pipeline. All experiments were conducted using Python 3.8, chosen for its mature ecosystem of libraries tailored for statistical modeling, deep learning, and visualization.

• Data Handling Tools: Pandas was utilized for dataset organization and manipulation, while NumPy enabled efficient array-based operations. SciPy supported numerical integrations, statistical testing, and filtering functions.

• Model Development Platforms: The deep learning models, particularly the LSTM networks, were built using Keras with TensorFlow backend, allowing flexible architecture definition, GPU acceleration, and model checkpointing. These tools also facilitated hyperparameter tuning via callbacks such as early stopping.

This study implemented a set of five distinct forecasting models, each reflecting a different class of temporal modelling:

1. AR: A linear model utilizing prior lagged values to predict current demand.
2. MA: Captures short-term memory by modelling noise in the error structure.
3. ARMA: Combines both AR and MA for stationary series.
4. ARIMA: Extends ARMA by incorporating differencing, ideal for non-stationary data.
5. LSTM: A type of recurrent neural network with memory gates, designed for long-term temporal dependency learning.

Each model was selected based on theoretical relevance, historical precedence in energy studies, and its anticipated performance under the data characteristics of Turkey's hourly demand. Performance comparisons were performed not only via statistical error metrics but also through visual inspection of forecast curves, providing an intuitive understanding of how well each model captured temporal patterns and extremes.

In the following sections, the mathematical formulations, training details, and empirical outcomes of each model are presented in depth, offering a comprehensive benchmark of their relative effectiveness in the context of national-scale energy forecasting.

A. AR Model

The AR model represents one of the foundational approaches in linear time series forecasting. It operates under the principle that present values in a time series can be effectively estimated based on a weighted summation of preceding observations. Mathematically, it is characterized by the expression:

$$X_t = c + \sum_{i=1}^{p} \phi_i X_{t-i} + \epsilon_t \tag{1}$$

where p is the order of the model, ϕ_i are parameters, and ϵ_t is white noise.

The AR model is particularly suited for stationary datasets exhibiting strong autocorrelation but lacking evident trends or seasonality. Its principal merits lie in computational simplicity and interpretability, making it a reliable tool in scenarios involving stable and predictable temporal dynamics [22, 23]. Nonetheless, the model's linear structure limits its ability to address complex patterns such as nonlinear fluctuations or seasonal shifts, which are common in real-world energy demand systems.

B. MA Model

The MA model is a classical approach within time series analysis that predicts current observations by incorporating past forecast errors rather than past values directly. Its general structure is expressed as:

$$X_t = \mu + \epsilon_t + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{2}$$

where μ is the mean of the series, q is the order of the model, and θ_i are the parameters of the model.

The MA model excels in filtering high-frequency noise, making it effective in stabilizing erratic short-term fluctuations. This feature is particularly valuable in preprocessing or when modeling residual components of more complex forecasting architectures [24]. However, its reliance solely on past error terms limits its capacity to model time series with underlying trends or seasonal effects, thereby reducing its standalone forecasting effectiveness in dynamic energy demand scenarios.

C. ARMA Model

The ARMA model represents a synergistic integration of two classical forecasting approaches: AR and MA. It leverages both the temporal dependencies captured by past values and the stochastic information contained in prior forecast errors to generate future predictions [24]. Its general mathematical representation is:

$$X_t = c + \varepsilon_t + \sum_{i=1}^{p} \phi_i X_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} \tag{3}$$

This model is particularly appropriate for stationary time series that exhibit both persistent patterns (autocorrelation) and residual-based randomness. One of its key strengths lies in its structural balance between memory and error correction, offering robust performance across various practical settings [25]. However, its dependency on stationarity implies that preliminary diagnostics, such as unit root tests, must be conducted to ensure model validity, and differencing might be required to stabilize the series.

D. ARIMA Model

The ARIMA model builds upon the foundational structure of ARMA by incorporating a differencing mechanism designed to address non-stationary behavior in time series data. This integration allows the model to stabilize trends and remove stochastic drift, rendering the data more amenable to linear modeling. The configuration is denoted as ARIMA(p,d,q), where p and q denote the autoregressive and moving average orders, and d specifies the number of differencing steps required to achieve stationarity.

ARIMA is particularly effective for datasets that exhibit underlying trends without seasonal fluctuations, offering a reliable solution in scenarios where classical stationary models fall short [26, 27]. Its strength lies in its ability to model a broad spectrum of real-world processes, including those characterized by gradual changes or evolving patterns. However, the model's efficacy hinges on accurate identification of the differencing order; incorrect parameterization can lead to excessive smoothing (over-differencing), which may obscure essential temporal dynamics or introduce bias into the forecasting process [28].

E. LSTM Model

LSTM networks, an advanced subclass of RNNs, are specifically engineered to capture temporal dependencies in sequential data. Unlike traditional RNNs, LSTMs are equipped with gated architectures, most notably the memory cell, which enables them to selectively retain relevant information across extended time intervals [29, 30]. This capacity renders them particularly advantageous for applications involving time series data with complex, nonlinear, and long-range correlations, such as energy demand forecasting.

The strength of LSTM lies in its adaptability to volatile data structures and its ability to model both short- and long-term influences within input sequences. However, these benefits come at the cost of increased computational demand and architectural complexity, which require considerable tuning and domain knowledge for optimal performance [31, 32].

In this study, the LSTM architecture was developed using the Keras API integrated with a TensorFlow computational backend (refer to Table 1). A two-layer network topology was selected following multiple tuning iterations, incorporating 64 and 32 hidden units in the first and second layers, respectively. The hidden layers utilized the hyperbolic tangent (tanh) activation function, while a linear activation was assigned to the output layer to align with the continuous nature of the target variable. Model training was conducted using the Adam optimizer. The Adam (Adaptive Moment Estimation) algorithm is an adaptive learning rate optimization method that calculates individual adaptive learning rates for different parameters. It effectively combines the benefits of two other extensions of stochastic gradient descent: Root Mean Square Propagation (RMSProp), which uses the exponentially decaying average of past squared gradients (v_t), and Momentum, which utilizes the exponentially decaying average of past gradients (m_t). By leveraging both the first moment (the mean) and the second moment (the uncentered variance) of the gradients, Adam ensures rapid and robust convergence in deep learning tasks like time series forecasting. and performance was guided by minimizing the Mean Squared Error (MSE) as the objective function. The training routine spanned up to 100 epochs with a mini-batch size of 64 samples. To prevent overfitting, an early stopping strategy was applied with a patience threshold of 10 epochs, and a dropout mechanism set at 0.2 was introduced to improve model generalization. The model ingested input sequences representing 24 consecutive hours of historical energy demand to forecast the subsequent hour's load value.

TABLE I
LSTM MODEL ARCHITECTURE AND TRAINING CONFIGURATION

Parameter	Value
Number of LSTM Layers	2
Units per Layer	64 and 32
Activation Function	tanh (hidden), linear (output)
Optimizer	Adam
Loss Function	Mean Squared Error (MSE)

The optimal set of hyperparameters detailed in Table 1 for the LSTM model was determined through an iterative empirical search and tuning process designed to minimize the prediction error (MSE/MAE) on the validation dataset while ensuring generalization. This process was critical given the high complexity and computational demands of deep learning models.

This rigorous empirical approach ensured that the final LSTM architecture was robust and highly optimized for the Turkey hourly energy demand dataset.

V. Evaluation Metrics and Visual Analysis

To rigorously assess the predictive performance of the models implemented in this study, a triad of statistical evaluation indicators was employed: MAE, RMSE, and MAPE. These complementary metrics were selected to capture different dimensions of forecasting accuracy, MAE reflecting average absolute deviation, RMSE emphasizing penalization of larger errors, and MAPE enabling percentage-based interpretability across scales.

This multi-faceted evaluation framework ensures a balanced and robust comparison of model outputs by accounting for both absolute and relative forecasting discrepancies. In addition to the numerical indicators, visual inspection of the prediction trajectories was conducted through plotted forecast curves, which allowed for intuitive identification of lagging responses, overfitting tendencies, and tracking performance across varying demand levels. Such graphical representations serve as a valuable supplement to statistical evaluation by revealing model behavior that may not be fully captured through quantitative measures alone.

A. Mean Absolute Error

The MAE quantifies the average discrepancy between forecasted values and actual observations, using absolute values to neutralize directional bias. [33, 34]. It is mathematically defined as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - y_i'| \quad (4)$$

where y_i are the actual observed values, y_i' are the predicted values, and n is the number of observations.

MAE is especially useful due to its intuitive interpretability, it expresses prediction errors in the same units as the target variable, making the results easily relatable in practical applications. However, one limitation of MAE is its equal weighting of all errors, regardless of magnitude, which can understate the impact of extreme deviations.

B. Root Mean Squared Error

The RMSE serves as a widely accepted metric for quantifying the accuracy of predictive models involving continuous numerical targets. It is computed by taking the square root of the mean of the squared deviations between actual values y_i and predicted values y_i' [34, 35].

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i')^2} \quad (5)$$

This metric is particularly sensitive to large prediction errors due to the squaring operation, which amplifies the effect of outliers and highlights deviations with greater magnitude. Unlike MAE, which assigns equal weight to all errors, RMSE disproportionately penalizes larger inaccuracies, making it especially informative in applications where substantial deviations are critical to detect and mitigate.

C. Mean Absolute Percentage Error

The MAPE is a relative accuracy metric that evaluates forecasting performance by expressing the average absolute deviation between predicted and actual values as a percentage of the actual values. It is mathematically represented as [36]:

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i'}{y_i}\right|\right) \times 100\% \quad (6)$$

The interpretability of MAPE, expressed in intuitive percentage terms, makes it especially suitable for comparing model performance across datasets with different scales [37]. For example, a MAPE of 20% suggests that the average error corresponds to 20% of the observed values. However, this metric can become unstable or undefined when actual values approach zero, a limitation that necessitates careful consideration when applying MAPE to datasets with near-zero entries.

D. Visual Analysis Through Forecasting Curves

Beyond quantitative error metrics, visual inspection of model outputs provides an essential layer of evaluation. Forecasting curves were generated for all implemented models to depict the alignment between predicted and actual energy demand values across the testing period. These visualizations enabled the identification of systematic biases, such as persistent underestimation or overestimation, and offered an intuitive view of the models' ability to follow temporal dynamics, including trends, seasonality, and abrupt shifts.

While metrics like RMSE and MAE summarize prediction accuracy numerically, they may obscure localized forecasting deficiencies, for example, consistent misalignment during demand peaks. Visual plots reveal such nuances and help contextualize the numerical indicators, offering a holistic understanding of model reliability under diverse operating conditions.

In combination, the graphical interpretations and statistical measures provide a robust framework for model performance assessment, guiding both the selection and refinement of forecasting approaches tailored to energy demand modeling.

VI. Results

To comprehensively assess the performance of each forecasting algorithm, a dual evaluation framework was adopted comprising both quantitative metrics, MAE, RMSE, and MAPE, and qualitative visual analysis using forecast versus actual demand curves. This approach enables a more nuanced understanding of each model's capacity to capture real-world consumption behaviors and react to dynamic fluctuations within the energy demand profile.

A. AR Model

The implementation of the AR model yielded the following statistical performance indicators on the test dataset: a MAE of 5779.73, a RMSE of 7138.90, and a MAPE of 14.53%. These results reflect the model's moderate ability to reconstruct baseline demand behaviour but highlight significant challenges in adapting to rapidly shifting consumption patterns. The relatively elevated RMSE points to the model's vulnerability to large prediction errors, particularly in scenarios involving abrupt demand spikes, which are common in heterogeneous and weather-sensitive electricity markets.

From a structural perspective, the AR model operates under the assumption of linear dependencies among past observations, making it well-suited for stationary and stable series but less adept at capturing complex, non-stationary fluctuations. Its reliance on a fixed number of prior lags restricts its memory depth, thereby limiting the capacity to identify long-term seasonal or nonlinear relationships. This constraint becomes especially evident during extreme demand conditions, where deviations from historical norms are more pronounced.

The visual forecasting profile generated by the AR model, illustrated in Fig. 1, demonstrates reasonable alignment with actual energy usage patterns during periods of moderate activity. However, during intervals of peak demand or rapid transitions, the model fails to adequately track the magnitude and direction of change. This lag in response results in systematic underestimation or overestimation, depending on whether the consumption shift is upward or downward.

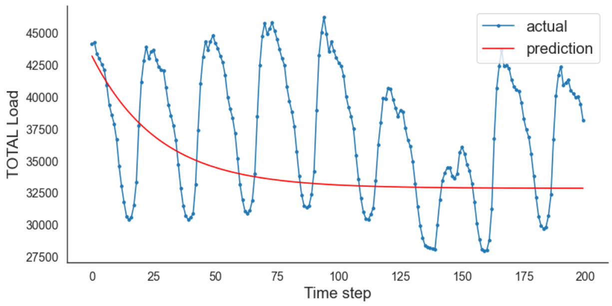

Fig. 1: Actual and Predicted Values Using the AR Model.

Furthermore, the forecast residuals during peak events suggest that the model does not internalize exogenous drivers of demand variability, such as temperature anomalies or sudden industrial load changes. This shortcoming reinforces the model's inherent static learning framework, which treats future conditions as deterministic extensions of the past. While such behaviour may be acceptable in controlled or less volatile systems, it proves inadequate for national-scale applications where real-time adaptability and resilience to shocks are paramount.

In summary, although the AR model serves as a useful benchmark for assessing the baseline forecasting potential of linear models, its performance is substantially constrained under volatile conditions. These findings highlight the need for more sophisticated architectures that can incorporate nonlinearity, extended memory, and contextual awareness, features that are largely absent in traditional autoregressive frameworks.

B. MA Model

The application of the MA model in this study produced distinctly contrasting outcomes across the three evaluation metrics. The model achieved a MAE of 0.09289 and a RMSE of 0.10963, suggesting relatively low average and squared deviations in absolute terms. However, the MAPE reached an anomalously high value of 99.998%, signaling a substantial disconnect between predicted and actual values when scaled relative to the magnitude of the data.

This sharp divergence between absolute and percentage-based error measures reveals a fundamental limitation of the MA model in handling the scaling dynamics of real-world energy demand data. While MAE and RMSE evaluate raw numerical differences, MAPE penalizes forecasting inaccuracies relative to actual values, making it highly sensitive in datasets where the range of values varies significantly over time. The exceptionally high MAPE thus exposes the model's pronounced inability to adapt to non-constant variance, especially during periods characterized by steep demand gradients or volatile peak events.

From a structural standpoint, the MA model operates by averaging past forecast errors over a defined window, assuming that future deviations will mirror historical discrepancies. This approach is inherently smoothing, designed to suppress random noise and enhance signal clarity in stationary time series. However, this very mechanism becomes a liability when applied to non-stationary, nonlinear, or seasonally variable data, as it effectively blunts the model's ability to detect or react to dynamic behavioral shifts.

The forecast trajectory plotted in Fig. 2 vividly illustrates the MA model's limitations in real-time applications. It consistently underpredicts demand spikes and overpredicts demand troughs, creating a lagging response that fails to track abrupt changes in the actual data series. This symmetrical dampening effect stems directly from the MA model's reliance on historical errors rather than raw observations or contextual cues.

Additionally, the MA model lacks any form of embedded memory or recursive structure, rendering it inadequate for capturing sequential dependencies or long-term temporal relationships. Its simplistic nature, while computationally efficient, does not lend itself well to environments characterized by frequent load shifts, temperature-driven consumption changes, or holiday-related anomalies.

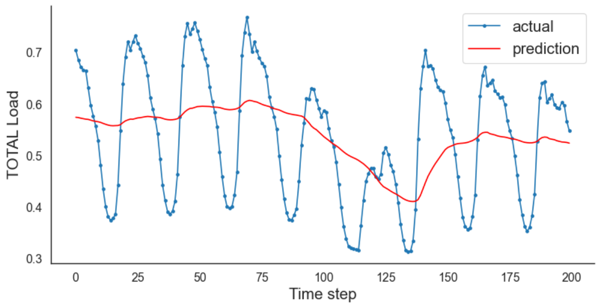

Fig. 2: Actual and Predicted Values Using the MA Model.

In practical terms, the MA model's forecasting framework might prove marginally useful in applications involving stationary, low-variance datasets, where the underlying patterns are relatively stable and free from abrupt discontinuities. However, its standalone applicability in national-level energy forecasting, where real-time responsiveness and adaptability are essential, is highly limited. When used, it is best integrated as a secondary component within hybrid modeling strategies, where it may assist in smoothing outputs or filtering short-term noise in conjunction with more adaptive forecasting architectures.

C. ARMA Model

The ARMA model, which integrates both autoregressive and moving average mechanisms, demonstrated a moderate level of forecasting accuracy in this study. Quantitatively, the model recorded a MAE of 5789.82, a RMSE of 7151.39, and a MAPE of 14.55%. These values suggest a modest improvement over standalone AR and MA models, particularly in capturing short-term dependencies and local error corrections. However, the elevated RMSE value, which penalizes larger deviations more severely, indicates the model's inadequate resilience against large-scale errors, particularly during periods of heightened volatility or structural regime shifts in energy demand.

The ARMA model's framework assumes that future values in a time series can be efficiently predicted by a linear combination of past values and past forecast errors, under the condition that the underlying series is stationary. While this assumption holds in simplified or synthetic datasets, its applicability diminishes in real-world energy systems, where the presence of seasonal fluctuations, exogenous shocks, and nonlinear dynamics is the norm rather than the exception.

The forecast visualization in Fig. 3 provides additional insights into the model's operational behavior. The forecast line generated by the ARMA model tends to smooth over sharp transitions, particularly those associated with peak load intervals or sudden drops in demand due to behavioral or environmental factors. This behavior reflects the model's difficulty in adapting to non-stationary phenomena, where changes in level, variance, or frequency are common and unpredictable.

Despite the inclusion of both AR and MA components, the ARMA structure remains fundamentally linear and deterministic, limiting its ability to represent nonlinear causal relationships or context-dependent anomalies. This limitation becomes especially pronounced in national-scale energy datasets where demand is influenced by a confluence of interacting variables, such as temperature, public holidays, industrial activity, and behavioral patterns, which are not directly modeled within the ARMA paradigm.

From a methodological standpoint, while ARMA offers greater flexibility than its individual constituents, it still operates under the constraints of stationarity and fixed memory horizons, rendering it suboptimal for long-range forecasting or environments with frequent structural breaks. Its primary strength lies in short-term forecasting of relatively stable segments, where its dual components can exploit autoregressive momentum and residual correction simultaneously.

In practical applications, the ARMA model may serve as a baseline benchmark in comparative studies or as a preprocessing filter for more advanced hybrid models. When used alone, however, its utility is constrained to datasets with limited volatility and minimal structural irregularity. In dynamic energy markets like Turkey's, characterized by rapid urban growth, fluctuating renewables penetration, and policy-driven consumption patterns, the ARMA model's performance is inherently limited.

In conclusion, while the ARMA model represents a meaningful progression beyond the simplicity of AR and MA approaches, it falls short in modeling the dynamic and nonlinear nature of modern energy systems, necessitating its augmentation or replacement with more adaptive frameworks such as LSTM or hybrid statistical-ML ensembles.

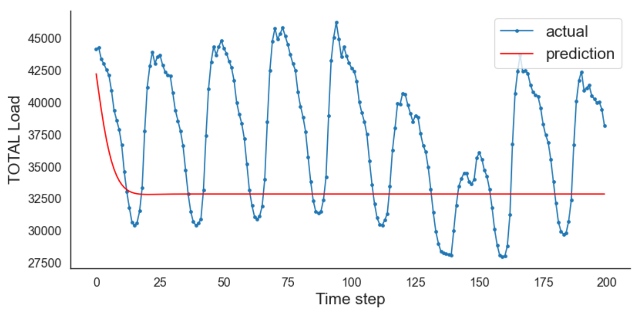

Fig. 3: Actual and Predicted Values Using the ARMA Model.

D. ARIMA Model

The ARIMA model, a sophisticated extension of the ARMA framework, incorporates differencing operations to transform non-stationary time series into stationary form, thereby enabling the application of autoregressive and moving average techniques. In this study, the ARIMA model exhibited a MAE of 5822.69, a RMSE of 7166.29, and a MAPE of 17.50%. These statistical outcomes suggest moderate predictive efficacy, with the relatively high RMSE signaling a susceptibility to large forecasting deviations. The MAPE value, indicating an average deviation of 17.5% between predicted and observed values, reflects the model's limited accuracy in scenarios involving high-frequency variability and abrupt demand changes.

While ARIMA models are well-regarded for their capacity to handle linear time series exhibiting trend-like behavior through differencing, they lack an intrinsic mechanism for capturing seasonal or cyclical patterns unless extended into SARIMA variants. In the context of hourly energy demand forecasting, where both short-term irregularities and longer-term seasonal cycles coexist, the ARIMA model's assumptions may not sufficiently reflect the intricate temporal structure of the data.

The forecast trajectory depicted in Fig. 4 reveals further insights into the model's behavior. Although ARIMA offers modest alignment with general trend movements, it tends to misalign with periodic oscillations and demonstrates difficulty in predicting demand peaks or troughs that deviate from past trends. This is symptomatic of the model's rigid structure, which relies heavily on prior data transformations and presumes consistent statistical properties throughout the forecasting horizon.

Moreover, while the differencing operation is instrumental in eliminating trend-induced non-stationarity, it may also obscure underlying seasonal signatures or degrade information embedded in long-term correlations. Over-differencing, in particular, can result in the attenuation of meaningful signal components, thereby reducing the model’s ability to capture real-world complexities inherent in energy consumption data.

From a computational standpoint, ARIMA requires careful parameter identification (p, d, q), a process that is often non-trivial and data-specific, especially for large datasets with variable temporal structures. In the context of national-scale datasets such as that of Turkey, with influences ranging from meteorological to socio-political, the assumption of stationarity within finite differencing may impose constraints on model scalability and generalizability.

Despite these challenges, ARIMA remains a valuable tool for benchmarking and for understanding the baseline behavior of time-dependent data. Its strength lies in modeling datasets where trends dominate and noise levels are relatively stable. However, in dynamic systems subject to exogenous shocks and nonlinear responses, as is typical in real-time energy markets, ARIMA models often underperform relative to more adaptive deep learning-based alternatives.

In summary, while ARIMA presents a theoretically sound approach for linear, trend-adjusted time series modeling, its application in high-resolution energy forecasting contexts is constrained by its limited flexibility, inability to autonomously adapt to seasonality, and high sensitivity to parameter mis-specification. This reinforces the need to integrate or transition toward hybrid and deep learning models capable of automated feature extraction and nonlinear mapping, particularly in fast-evolving energy systems.

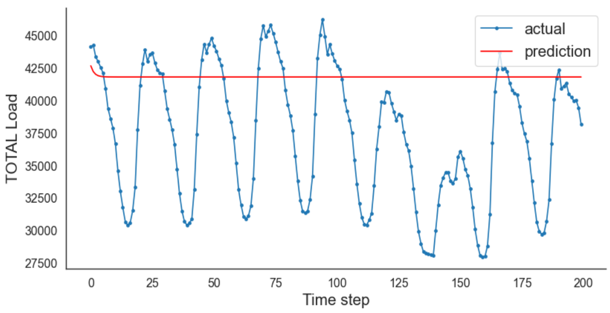

Fig. 4: Actual and Predicted Values Using the ARIMA Model.

E. LSTM Model

The LSTM network, a variant of RNNs designed to mitigate vanishing gradient issues, demonstrated exceptional performance in forecasting hourly electricity demand within the Turkish national grid. Leveraging its gated architecture, which allows for dynamic memory retention across long sequences, the model captured complex dependencies and temporal fluctuations with high precision.

During the training phase, the model achieved a MAE of 330.86 and a RMSE of 464.52. On the testing dataset, unseen by the model during training, the corresponding values slightly increased to 389.21 (MAE) and 550.98 (RMSE). Notably, the MAPE remained remarkably low and stable across both phases, registering 1.03% for training and 1.02% for testing, highlighting the model’s robust generalization capacity and resistance to overfitting. These consistently low error rates affirm the LSTM model’s capability to internalize latent temporal structures and accurately forecast future demand even under highly dynamic conditions.

The forecast curve generated by the LSTM, illustrated in Fig. 5, reflects a high degree of congruence with the actual observed energy demand. Unlike traditional statistical models, the LSTM is proficient in tracking both sharp surges and sudden drops, effectively responding to nonlinear consumption behaviors driven by exogenous influences such as weather shifts, holiday effects, and unexpected load variations. This responsiveness is indicative of the model's intrinsic ability to model nonlinearity and long-range dependencies, which are fundamental characteristics of real-world electricity consumption. Additionally, the learning dynamics of the LSTM model, as visualized in Fig. 6, further corroborate its stability and adaptability. The training and validation loss curves display rapid convergence within the first 50 epochs, followed by a plateauing behavior, indicating that the network has successfully identified the underlying data distribution. The near overlap between the two loss curves implies minimal generalization gap, suggesting the effectiveness of the regularization techniques employed, such as dropout and early stopping, in mitigating overfitting. Based on this behavior, an early stopping mechanism at epoch 50 emerges as an optimal strategy for maintaining forecasting accuracy while minimizing computational costs.

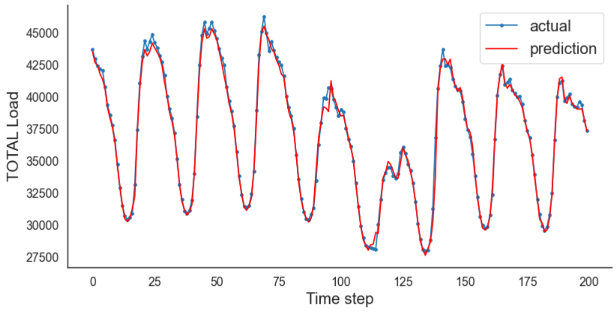

Fig. 5: Actual and Predicted Values Using the LSTM Model.

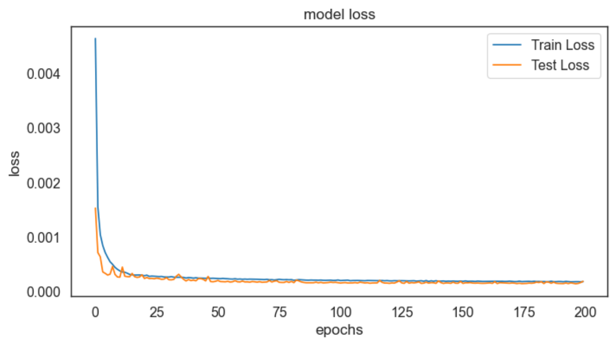

Fig. 6: Training and Testing Loss Curves for the LSTM Model.

The architectural strengths of LSTM are pivotal to its superior performance. Comprising memory cells regulated by input, output, and forget gates, the LSTM is uniquely suited for modeling time series where patterns evolve across variable temporal lags. This architecture enables the network to preserve context across extended time horizons, a critical feature for accurately forecasting demand cycles influenced by daily routines, seasonal weather patterns, and policy interventions.

Furthermore, LSTM's ability to learn from raw sequential data without extensive manual feature engineering makes it highly scalable and transferable to other domains of energy analytics. In contexts such as Turkey's energy market, characterized by a mix of conventional and renewable sources, and subject to climatic and socio-economic variability, such flexibility is especially advantageous.

In conclusion, the LSTM model surpasses traditional and hybrid forecasting techniques in both predictive accuracy and interpretive reliability. Its performance in this study validates the strategic integration of deep learning architectures in energy forecasting systems. As energy systems continue to evolve toward greater decentralization and real-time responsiveness, models like LSTM offer a future-proof solution capable of supporting high-resolution, demand-side planning and grid management strategies.

VII. Discussion

The comparative analysis of time series forecasting models, as evidenced by the statistical performance metrics, MAE, RMSE, and MAPE, presented in Figs. 7, 8, 9, and in Table 2, reveals substantial variation in predictive accuracy among classical and deep learning approaches. These differences underscore the distinct modeling capabilities of each method when faced with the nonlinear, volatile nature of real-world energy demand data. While traditional models offer simplicity and interpretability, their performance deteriorates in capturing complex temporal dependencies. Conversely, advanced models like LSTM demonstrate superior adaptability and accuracy, especially in dynamic environments where precision is critical. This reinforces the growing need to incorporate deep learning techniques in modern energy forecasting pipelines.

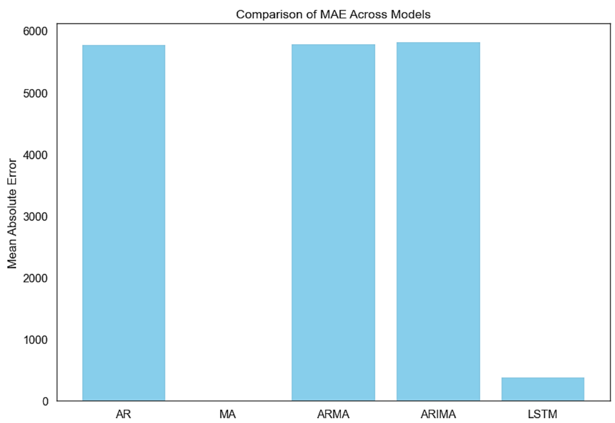

Fig. 7: MAE Comparison Across Employed Models.

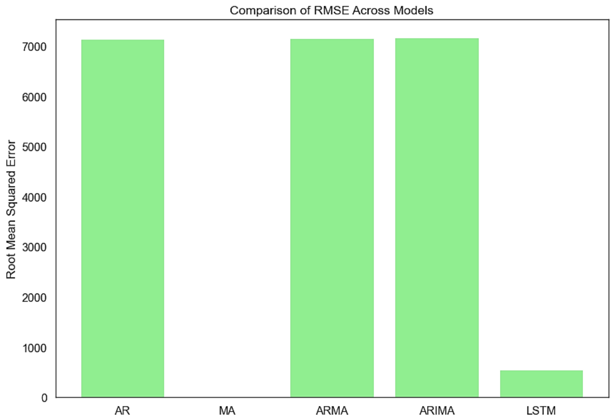

Fig. 8: RMSE Comparison Across Employed Models.

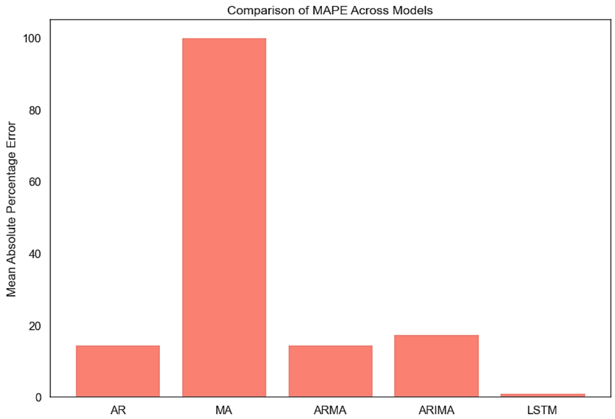

Fig. 9: MAPE Comparison Across Employed Models.

TABLE II
MODEL PERFORMANCE COMPARISON

Model	MAE	RMSE	MAPE	Rank (by MAPE)
LSTM	389.21	550.98	1.02%	1
AR	5779.73	7138.90	14.5%	3
ARMA	5789.82	7151.39	14.55%	4
ARIMA	5822.69	7166.29	17.5%	5
MA	0.09289	0.10963	99.99%	2

The MAE, RMSE, and MAPE graphs clearly demonstrate that traditional models such as AR, MA, ARMA, and ARIMA, while useful in certain scenarios, consistently yield higher error rates compared to the LSTM model. These traditional models show substantial errors, with MAE and RMSE values indicating that the magnitude of their forecasting errors is significantly higher, which could be detrimental in practical applications where precise demand forecasting is crucial. Particularly, the MA model's performance with an unusually high MAPE points to its inefficiency in datasets characterized by volatile fluctuations, confirming its limitations in handling dynamic changes within the energy sector.

In contrast, the LSTM model's performance excels across all evaluated metrics, showcasing markedly lower values in MAE, RMSE, and MAPE. This indicates not only its superior accuracy in capturing the actual energy demand but also its consistency in maintaining this accuracy across different data segments, including both training and testing phases. The LSTM's robust ability to adapt to new data, learn from long-term dependencies, and manage both sudden spikes and drops in energy usage highlights its potential to revolutionize energy demand forecasting.

The findings from the graphical comparisons advocate strongly for the integration of LSTM models in complex, non-linear environments typical of the energy sector. Future research should explore the development of hybrid models that combine the interpretative benefits of traditional statistical methods with the predictive accuracy and adaptability of LSTM networks. Such hybrid approaches could further enhance forecasting precision while providing deeper insights into the factors driving changes in energy demand. This could also include the integration of real-time data adjustments to continually refine forecasts in response to live data, ensuring that predictive models remain as accurate and relevant as possible over time.

VIII. Conclusion

This study provided a comprehensive comparative evaluation of classical statistical models (AR, MA, ARMA, ARIMA) and advanced deep learning techniques (LSTM) in forecasting hourly energy demand within the dynamic context of Turkey's power grid. The empirical analysis highlighted that while classical models exhibit reasonable performance under stationary and linear conditions, they struggle to adapt to the complexity and volatility inherent in real-world energy data.

In contrast, the LSTM model demonstrated a marked superiority, effectively capturing long-range dependencies, handling non-linearities, and maintaining robustness in the face of seasonal and abrupt fluctuations. Its performance underscores the transformative potential of deep learning in energy forecasting, especially when precision and adaptability are critical to operational planning and policy formulation.

However, the rapidly evolving nature of energy systems, driven by renewable integration, demand-side management, and geopolitical uncertainties, demands even more responsive and intelligent forecasting architectures. Therefore, future research should explore the integration of LSTM with real-time data ingestion pipelines and anomaly detection frameworks to enhance adaptability and fault tolerance.

Moreover, comparative investigations involving more recent architectures such as the Transformer and Gated Recurrent Unit (GRU) models are warranted. These models, particularly those utilizing attention mechanisms, have shown remarkable success in other time series domains and may offer further improvements in forecast accuracy and interpretability. Likewise, embedding external drivers such as electricity pricing, economic indicators, or policy variables into model architectures could enrich predictive performance and contextual relevance.

Another promising direction lies in the development of hybrid forecasting systems that combine the statistical rigor and explainability of CST models with the pattern-learning power of LSTM and attention-based architectures. Such hybrid approaches may yield enhanced accuracy while preserving computational efficiency and interpretability, two aspects often traded off in pure deep learning models.

In conclusion, the LSTM model represents a significant advancement in energy demand forecasting, but it should be viewed as a foundation upon which more sophisticated, adaptive, and hybrid forecasting solutions can be built. Integrating LSTM with attention-based mechanisms, economic variables, or real-time anomaly detection may form the next frontier in forecasting research, offering substantial benefits for energy planners, system operators, and policymakers seeking resilient and data-driven solutions in an increasingly uncertain energy landscape.

Conflict of Interest

The authors declare no conflict of interest in the publication process of the research article.

Author Contributions

The author solely conceived the study, designed the methodology, implemented the software, conducted the analyses and investigations, curated the data, created the figures and tables, drafted and revised the manuscript, and approved the final version.

References

[1] C. Deb, F. Zhang, J. Yang, S. E. Lee, and K. W. Shah, "A review on time series forecasting techniques for building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 74, pp. 902-924, 2017.

[2] B. Zhang et al., "Assessment of the economic impact of forecasting errors in Peer-to-Peer energy trading," *Applied Energy*, vol. 374, p. 123750, 2024.

[3] Y. Pan, C.-C. Zhang, C.-C. Lee, and S. Lv, "Environmental performance evaluation of electric enterprises during a power crisis: Evidence from DEA methods and AI prediction algorithms," *Energy Economics*, vol. 130, p. 107285, 2024.

[4] X. Zhang and Y. Zhang, "Environment-friendly and economical scheduling optimization for integrated energy system considering power-to-gas technology and carbon capture power plant," *Journal of Cleaner Production*, vol. 276, p. 123348, 2020.

[5] U. Bulut and G. Muratoglu, "Renewable energy in Turkey: Great potential, low but increasing utilization, and an empirical analysis on renewable energy-growth nexus," *Energy policy*, vol. 123, pp. 240-250, 2018.

[6] E. Toklu, M. Güney, M. Işık, O. Comaklı, and K. Kaygusuz, "Energy production, consumption, policies and recent developments in Turkey," *Renewable and Sustainable Energy Reviews*, vol. 14, no. 4, pp. 1172-1186, 2010.

[7] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, Time series analysis: forecasting and control. *John Wiley & Sons*, 2015.

[8] M. Saqib, E. Şentürk, S. A. Sahu, and M. A. Adil, "Comparisons of autoregressive integrated moving average (ARIMA) and long short term memory (LSTM) network models for ionospheric anomalies detection: a study on Haiti (M w= 7.0) earthquake," *Acta Geodaetica et Geophysica*, pp. 1-19, 2022.

[9] K. ArunKumar, D. V. Kalaga, C. M. S. Kumar, M. Kawaji, and T. M. Brenza, "Comparative analysis of Gated Recurrent Units (GRU), long Short-Term memory (LSTM) cells, autoregressive Integrated moving average (ARIMA), seasonal autoregressive Integrated moving average (SARIMA) for forecasting COVID-19 trends," *Alexandria engineering journal*, vol. 61, no. 10, pp. 7585-7603, 2022.

[10] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mechanical Systems and Signal Processing*, vol. 115, pp. 213-237, 2019.

[11] A. Sagheer and M. Kotb, "Time series forecasting of petroleum production using deep LSTM recurrent networks," *Neurocomputing*, vol. 323, pp. 203-213, 2019.

[12] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice. OTexts, 2018.

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

[14] S. R. Khuntia, J. L. Rueda, and M. A. van Der Meijden, "Forecasting the load of electrical power systems in mid-and long-term horizons: A review," *IET Generation, Transmission & Distribution*, vol. 10, no. 16, pp. 3971-3977, 2016.

[15] C. Chatfield and H. Xing, The analysis of time series: an introduction with R. Chapman and hall/CRC, 2019.

[16] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber, "LSTM: A search space odyssey," *IEEE transactions on neural networks and learning systems*, vol. 28, no. 10, pp. 2222-2232, 2016.

[17] A. Mosavi, M. Salimi, S. Faizollahzadeh Ardabili, T. Rabczuk, S. Shamshirband, and A. R. Varkonyi-Koczy, "State of the art of machine learning models in energy systems, a systematic review," *Energies*, vol. 12, no. 7, p. 1301, 2019.

[18] J. Ma, A. Oppong, K. Nketia Acheampong, and L. A. Abruquah, "Forecasting renewable energy consumption under zero assumptions," *Sustainability*, vol. 10, no. 3, p. 576, 2018.

[19] L. M. De Menezes, D. W. Bunn, and J. W. Taylor, "Review of guidelines for the use of combined forecasts," *European Journal of Operational Research*, vol. 120, no. 1, pp. 190-204, 2000.

[20] R. T. Clemen, "Combining forecasts: A review and annotated bibliography," *International journal of forecasting*, vol. 5, no. 4, pp. 559-583, 1989.

[21] M. Ozturk and Y. E. Yuksel, "Energy structure of Turkey for sustainable development," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 1259-1272, 2016.

[22] J. Kaur, K. S. Parmar, and S. Singh, "Autoregressive models in environmental forecasting time series: a theoretical and application

review," *Environmental Science and Pollution Research*, vol. 30, no. 8, pp. 19617-19641, 2023.
[23] T. Ullrich, "On the autoregressive time series model using real and complex analysis," *Forecasting*, vol. 3, no. 4, pp. 716-728, 2021.
[24] L. Alhmoud and Q. Nawafleh, "Short-term load forecasting for Jordan power system based on NARX-ELMAN neural network and ARMA model," *IEEE Canadian Journal of Electrical and Computer Engineering*, vol. 44, no. 3, pp. 356-363, 2021.
[25] Z. Fang, D. L. Dowe, S. Peiris, and D. Rosadi, "Minimum message length in hybrid ARMA and LSTM model forecasting," *Entropy*, vol. 23, no. 12, p. 1601, 2021.
[26] C. Tarmanini, N. Sarma, C. Gezegin, and O. Ozgonenel, "Short term load forecasting based on ARIMA and ANN approaches," *Energy Reports*, vol. 9, pp. 550-557, 2023.
[27] E. Chodakowska, J. Nazarko, and Ł. Nazarko, "Arima models in electrical load forecasting and their robustness to noise," *Energies*, vol. 14, no. 23, p. 7952, 2021.
[28] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, and G. K. Matsopoulos, "A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks," *Future Internet*, vol. 15, no. 8, p. 255, 2023.
[29] J. Q. Wang, Y. Du, and J. Wang, "LSTM based long-term energy consumption prediction with periodicity," *energy*, vol. 197, p. 117197, 2020.
[30] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural computing and applications*, vol. 31, pp. 2727-2740, 2019.
[31] Y. Wang, S. Zhu, and C. Li, "Research on multistep time series prediction based on LSTM," *in 2019 3rd international conference on electronic information technology and computer engineering (EITCE)*, 2019: IEEE, pp. 1155-1159.
[32] N. L. M. Jailani et al., "Investigating the power of LSTM-based models in solar energy forecasting," *Processes*, vol. 11, no. 5, p. 1382, 2023.
[33] W. Wang and Y. Lu, "Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model," *in IOP conference series: materials science and engineering*, 2018, vol. 324: IOP Publishing, p. 012049.
[34] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geoscientific Model Development Discussions*, vol. 2022, pp. 1-10, 2022.
[35] D. S. K. Karunasingha, "Root mean square error or mean absolute error? Use their ratio as well," *Information Sciences*, vol. 585, pp. 609-629, 2022.
[36] A. De Myttenaere, B. Golden, B. Le Grand, and F. Rossi, "Mean absolute percentage error for regression models," *Neurocomputing*, vol. 192, pp. 38-48, 2016.
[37] G. A. Busari and D. H. Lim, "Crude oil price prediction: A comparison between AdaBoost-LSTM and AdaBoost-GRU for improving forecasting performance," *Computers & Chemical Engineering*, vol. 155, p. 107513, 2021.