29th Electrical Power Distribution Conference, April 29- May 1, 2025, Tehran, Iran

An Attention-Based Deep Learning Model for Multi-Horizon Prediction of Load, Price, and Wind Power Generation in Smart Grids

S. Mozhgan Rahmatinia
Department of Computer Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
mozhgan.rahmatinia@mail.um.ac.ir

Seyed-Amin Hosseini-Seno
Department of Computer Engineering
Ferdowsi University of Mashhad
Mashhad, Iran
hosseini@um.ac.ir

***Abstract* – Accurate forecasting of critical parameters in power grids, such as electricity load, price, and renewable energy production, plays a pivotal role in implementing demand response strategies and maintaining grid stability within smart grid systems. In this paper, we propose a novel deep learning approach based on an encoder-decoder architecture with an attention mechanism to simultaneously predict these essential parameters. To enhance the model's ability to capture complex patterns and long-term dependencies in time series data, we incorporate a preprocessing step using the Fast Fourier Transform (FFT), which transforms the data into the frequency domain. This transformation aids in reducing noise and extracting latent patterns, thereby improving the model's predictive performance. The proposed model is evaluated on real-world electricity consumption datasets from multiple regions, including Austria, Italy, Sweden, and the UK. It is compared against several state-of-the-art time series forecasting models, such as LSTM, CNN-LSTM, LSTM-Attention, and a basic encoder-decoder structure. The results demonstrate that our approach achieves significantly higher accuracy in most cases, particularly for short- to medium-term predictions (12 to 48 hours ahead). Additionally, visualizations of the predictions reveal that the proposed method closely follows the actual trends, confirming its effectiveness in capturing both short-term fluctuations and long-term dependencies. This study contributes to the field by introducing a flexible and robust framework that integrates attention mechanisms and frequency-domain preprocessing to improve time series forecasting in the electricity domain. The proposed approach not only enhances prediction accuracy but also provides valuable insights into the underlying patterns of electricity consumption and production, making it a promising tool for smart grid applications. The source code and implementation details of the proposed model are publicly available on GitHub[1].**

Keywords— Deep Learning, Multivariate Time Series Prediction, Encoder-Decoder Architecture, Attention Mechanism, Temporal Patterns, Demand Side Management.

I. Introduction

The increasing energy consumption and the expansion of electric vehicles have necessitated the transition to smart power grids, introducing challenges such as stability, efficiency, and reliability. Accurate forecasting of electricity load and prices, along with clean energy generation, plays a fundamental role in key grid management processes and Demand Response (DR) programs.

One of the most crucial approaches to maintaining the stability of smart grids is the implementation of demand response strategies, which are categorized into incentive-based and price-based models [1]. In price-based models, such as real-time pricing (RTP), electricity prices fluctuate based on demand levels. However, uncertainties in forecasting electricity prices and grid load pose a significant challenge, as variations in consumption can unexpectedly impact prices. Consequently, simultaneous forecasting of load, electricity prices, and clean energy generation is essential for improving decision-making.

Moreover, predicting electricity prices without considering fluctuations in load and clean energy generation, and vice versa, results in inaccurate estimates and suboptimal decisions. This issue arises due to the dynamic and interdependent nature of electricity load and prices; increased consumption during peak hours leads to higher prices, whereas reduced consumption during off-peak hours results in lower prices. Therefore, simultaneous forecasting of electricity load, prices, and clean energy generation is critical for enhancing prediction accuracy and optimizing decision-making in demand-side management programs and electricity markets.

Classical time series methods such as ARIMA [2-4], dynamic regression [5], and linear regression [6], despite their efficiency in linear data, face limitations in modeling nonlinear relationships and complex patterns with seasonal and trend features. Additionally, these models often fail to adequately capture long-term dependencies and perform poorly in the presence of extreme fluctuations and sudden changes in data. In multivariate time series forecasting, variables such as weather conditions, date, time, and electricity prices play a crucial role in prediction accuracy. Moreover, general data trends, seasonal variations, and irregular fluctuations pose fundamental challenges in forecasting smart grid data. These complexities have rendered classical and statistical methods less effective and constrained in extracting precise patterns.

Neural network-based models, particularly LSTM and GRU, have addressed some of these challenges, yet they still exhibit limitations in learning long-term dependencies [7, 8]. Combining these models with CNN has enhanced feature extraction but remains constrained in certain aspects [9, 10]. Integrating deep learning architecture with the Attention mechanism has enabled a focus on key features, significantly improving forecasting accuracy[11-13].

[1] https://github.com/mozhgan-Rahmatinia/Attention-based-deep-for-load-price-wind-forecasting-in-Smart-grid

TABLE I. TABLE OF ABBREVIATIONS

FFT	Fast Fourier Transform
S2S	Sequence-to-Sequence architecture
STLF	Short-term Load Forecasting
STPF	Short-term electricity Price Forecasting
DSM	Demand Side Management
DR	Demand Response
RTP	Real-Time Pricing
RNN	Recurrent Neural Networks
LSTM	Long Short-term Memory
BiLSTM	Bidirectional Long Short-Term Memory
GRU	Gated Recurrent Unit
CNN	Convolutional Neural Network
ARIMA	Autoregressive Integrated Moving Average
RBF	Radial Basis Function
BSO	Bird Swarm Optimizer
BWO	Beluga Whale Optimization
PSO	Particle Swarm Optimization
BO	Bayesian Optimization
FL	Federated Learning

Sequence-to-Sequence (S2S) architectures with an Encoder-Decoder mechanism provide high flexibility and the ability to learn complex variable interactions. Additionally, models within this family, by utilizing the Multi-head Attention mechanism, have been able to improve the simultaneous prediction of multiple features[14].

This research presents an attention-based deep learning model that simultaneously forecasts electricity load, electricity prices, and wind energy generation for future time periods. The proposed model leverages the complex and nonlinear correlations between these variables and incorporates contextual features such as weekdays and historical trends. Compared to traditional methods and conventional machine learning algorithms, the model provides more accurate predictions. This enhanced forecasting capability helps reduce consumer costs, improve grid stability, and increase the efficiency of demand-side management programs.

The structure of the proposed model follows the Encoder-Attention-Decoder architecture. The Encoder employs a stack of Bidirectional Long Short-Term Memory (BiLSTM) networks to extract temporal features and recognize behavioral patterns. However, due to the long-term dependencies in the data, these extracted features alone are insufficient for precise predictions. Therefore, an attention mechanism is incorporated into the model to focus on key features, better capture long-term dependencies, and improve forecasting accuracy. Finally, the Decoder utilizes the extracted information to provide accurate predictions of electricity load and prices.

In addition to initial processing, the data is transformed into the frequency domain to reduce noise and better extract temporal patterns. To achieve this, FFT is employed as a complementary preprocessing step, enhancing model performance and stability.

The key contributions of this research can be summarized as follows:

- Proposing a deep learning model capable of simultaneously predicting electricity consumption and prices with high accuracy.
- Utilizing effective features such as temporal information and historical data to improve forecasting precision.
- Employing Fast Fourier Transform (FFT) in the data preprocessing stage, enabling better extraction of seasonal patterns and trends, balancing short-term and long-term dependencies.
- Introducing an Attention Mechanism within the Encoder-Decoder architecture that, through stacked BiLSTM networks, effectively models complex temporal dependencies and enhances prediction accuracy.
- Presenting a flexible architecture capable of forecasting electricity load and prices over variable time horizons with accurate short-term and long-term predictions.

The structure of the paper includes a review of related works (Section 2), problem formulation and proposed model (Section 3), model architecture explanation (Section 4), data description and results analysis (Section 5), and conclusion and future work (Section 6).

II. RELATED WORK

Short-term load forecasting (STLF), short-term electricity price forecasting (STPF), and wind power generation forecasting are essential for the optimal management of smart grids. Load forecasting contributes to grid stability and reduces production and distribution costs, while electricity price forecasting plays a crucial role in demand-side management (DSM) and balancing supply and demand. Wind power generation forecasting presents additional challenges due to its dependence on weather conditions. The fluctuations in load, price, and wind generation—driven by environmental and economic factors—highlight the need for accurate forecasting models.

Developing precise forecasting models for electricity-related data, particularly through hybrid approaches and machine learning techniques, has been a key focus of researchers. In Ref [15], the authors employed fuzzy clustering to segment input data into different clusters and used a hybrid model comprising a neural network with aggregation layers, convolutional layers, and a radial basis function (RBF) to forecast the next week's load.

The effectiveness of recurrent neural networks (RNNs) such as long short-term memory (LSTM) and gated recurrent unit (GRU) in time-series forecasting, due to their ability to capture dependencies within data, has led to increasing interest in integrating these methods with other learning models. Ref [10], LSTM was employed as a hybrid long-term and short-term forecasting method, where a convolutional neural network (CNN) was used to identify local dependencies, while an RNN captured long-term dependencies in multidimensional inputs. Additionally, in Ref [16], a genetic algorithm was utilized to optimize LSTM network parameters, including the number of layers and time delays. Furthermore, in Ref [17], hybrid models integrating ARIMA and LSTM were developed.

In recent years, more advanced hybrid models have been introduced. In Ref [18], a Hybrid Stacking-based model was proposed for short-term electricity price forecasting. This model combined XGBoost, CatBoost, NODE, LightGBM, GRU, and LSTM, incorporating weather data and time-series features to significantly improve forecasting accuracy. Similarly, Ref [19], a deep learning-based hybrid model (LFS-HDLBWO) was proposed for STLF in smart grids. By integrating CBLSTM-AE with the beluga whale optimization (BWO) algorithm, the model automatically optimized hyperparameters, enhancing forecasting accuracy compared to LSTM and ANN models while also achieving improved execution efficiency. Ref [20] investigates residential energy management and electricity price forecasting in smart homes based on the Internet of Energy (IoE). The objective of this research is to utilize a combination of the Gated Recurrent Unit (GRU) and the Bird Swarm Optimizer (BSO) algorithm for electricity price prediction and energy consumption scheduling, which has successfully reduced both energy consumption and costs.

Integrating attention mechanisms into deep networks has been widely explored as an effective approach to improving the accuracy of time-series forecasting. These mechanisms enhance the ability of deep learning models by strengthening the identification of meaningful patterns and preserving long-term dependencies in data. Consequently, numerous studies have focused on combining attention mechanisms with deep learning models.

Reference [21], a multitask attention mechanism was incorporated into an LSTM-based model for electricity load forecasting. Similarly, [22] introduced an attention-based encoder-decoder structure, where the attention-driven encoder extracts correlations among input loads, and the decoder learns temporal dependencies.

Other studies have demonstrated that combining CNN and LSTM networks with multi-head attention can further enhance forecasting performance. In [23], a hybrid CNN-LSTM model was proposed, utilizing multi-head attention to extract significant time-series features. Furthermore, [24] leveraged a rolling update (RU) approach alongside an attention mechanism and BiLSTM to improve the accuracy of load forecasting across different countries.

In [25], the authors aim to find a balance between interpretability and prediction accuracy in models by comparing the interpretability of two non-RNN methods, N-BEATSx and TFT, with the recurrent method LSTM. This study demonstrates that although the prediction accuracy is higher in RNN-based approaches, the interpretability of non-recurrent methods is greater, which is highly significant for practical applications. Furthermore, the variable indicating weekends was found to be the most important factor in the forecasting process.

Several studies have also examined the integration of data decomposition techniques with deep learning networks. For instance, Ref [8] proposed a hybrid model combining Ensemble Empirical Mode Decomposition (EEMD) with BiLSTM and an attention mechanism for load and electricity price forecasting. This model employed Bayesian Optimization (BO) alongside Random Forest Regressors for automated hyperparameter tuning. Additionally, Ref [26] introduced an encoder-decoder model with a mixture of attention mechanism and pinball loss function for multi-step forecasting tasks.

In reference [27], Pentsos and colleagues aim to enhance the accuracy of electricity load forecasting for 650 households in a smart grid by considering various factors such as user behavioral patterns, geographical conditions, and temporal constraints. To achieve this goal, they propose a hybrid model that leverages the capabilities of LSTM to understand temporal dependencies and employs the transformer attention mechanism to extract significant long-term features. Additionally, the article [28] presents a hybrid neural network that uses time-series clustering to categorize similar electricity consumption patterns. Subsequently, these similar patterns are analyzed separately using a transformer model with the goal of STLF. This approach leads to improved accuracy in the prediction method.

Beyond these approaches, some research efforts have explored the combination of deep learning and federated learning to enhance forecasting accuracy while preserving data privacy. Federated Learning (FL), as a decentralized approach, enables data processing at its source without requiring direct data sharing between computational nodes, thereby ensuring security and privacy. In [29], a hybrid 1D-CNN-GRU model with an attention mechanism was proposed, integrating hyperparameter optimization via the Particle Swarm Optimization (PSO) algorithm and leveraging FL to improve load forecasting accuracy. Additionally, pruning techniques in FL were employed to reduce computational resource consumption. Along the same lines, the FedGrid framework [30] introduced a federated learning-based approach for forecasting electricity load and renewable energy generation. This model utilized LSTM and DSS-LSTM to predict electricity consumption and solar and wind power generation while ensuring data security and privacy without sharing sensitive information. The results demonstrated enhanced data security, reduced reliance on fossil fuels, and optimized demand-supply management in smart grids.

Moreover, certain deep learning models have achieved superior performance in load forecasting by integrating multiple advanced approaches. Ref [31], the DenseNet-AM-LSTM model was proposed for smart grid load forecasting. By combining DenseNet, LSTM, and an attention mechanism, this model improved forecasting accuracy. A comparative analysis against traditional methods such as ARIMA, ResNet, and CNN indicated that DenseNet-AM-LSTM outperformed these approaches in managing smart grid load dynamics.

III. Problem Formulation

This study focuses on developing a model for simultaneous forecasting of electricity consumption, electricity prices, and wind power generation based on historical data over the past k time intervals. The goal is to estimate the corresponding values for the next n time intervals. To achieve this, an attention-based deep learning network is employed.

In this approach, the input data for each training and forecasting step is represented as a matrix R_{K*I}, which is defined by the following Equation:

$$R = \{l^t, p^t, w^t, f_i^t \mid t \in \{m-k, \ldots, m-1\}, i \in \{1,2,\ldots,n\}\}$$

Where k represents the number of past time intervals, i denotes the number of features influencing the forecasting process, l^t is the electricity consumption at time interval t, p^t is the electricity price at time interval t, and w^t represents wind power generation at time interval t. Additionally, f includes other relevant features affecting the forecasting process, such as temperature, hour of the day, day of the week, and month, each identified by i.

The forecasting of electricity consumption, electricity price, and wind power generation for the next n time intervals is represented as a Pr_{n*3} matrix, which is defined by Equation (2):

$$Pr = \{ l^{t'}, p^{t'}, w^{t'} \mid t' \in \{m, \ldots, m+n\} F \} \quad (2)$$

In simple terms, our model predicts three things at once—electricity demand, price, and wind energy production—for the next n time steps using data from the past k time steps. It analyzes historical patterns to make accurate forecasts.

From a scientific view, the model uses a sequence-to-sequence approach, leveraging multiple input features to capture complex relationships between load, price, and wind energy, improving prediction quality.

Figure 1 illustrates the steps undertaken in this study for Multivariate-Multihorizon Time Series Prediction, which is structured into three main phases: data preprocessing, model development, training and evaluation, and finally, testing the model on unseen data.

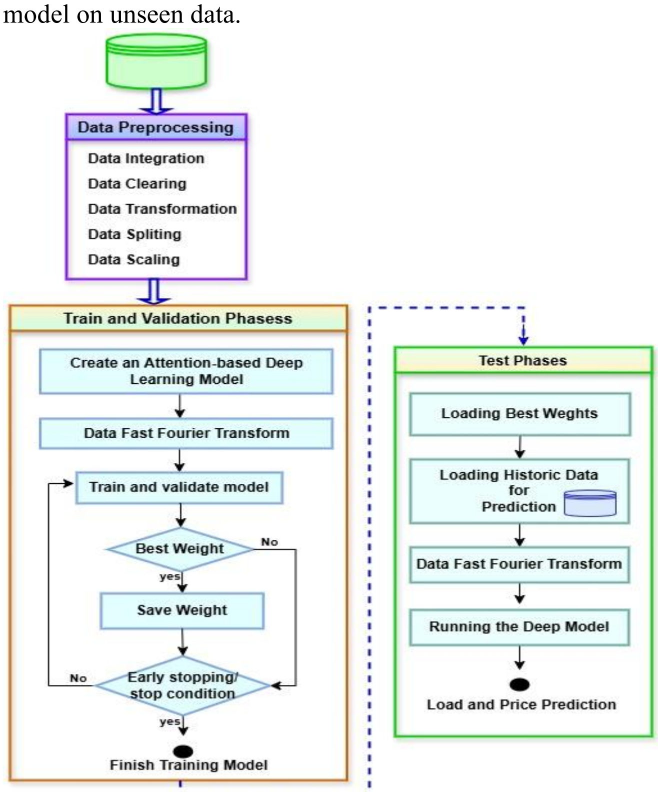

Fig. 1 Flowchart of the proposed method.

VI. Model Architecture

The Attention-Based Deep Learning architecture proposed in this study consists of four main components:

Preprocessing with Fast Fourier Transform (FFT): To identify periodic and trend-based patterns in time series data, FFT is applied as a preprocessing step. This transformation converts data from the time domain to the frequency domain, improving the extraction of key information. Importantly, FFT is applied only to the encoder's input, while the decoder receives the original data without an inverse transformation, preventing information leakage.

Encoder: The encoder is responsible for transforming input data into a feature space where meaningful patterns and long-term dependencies can be effectively extracted. It consists of multiple layers of BiLSTM networks, which excel at capturing temporal dependencies. Additionally, a Linear Layer is incorporated to extract local features, enhancing the model's ability to detect short-term patterns in the data.

Attention Mechanism: The attention mechanism is integrated to enhance the impact of significant features while reducing the influence of less relevant ones. By dynamically assigning weights to different parts of the input sequence, the attention mechanism enables the model to focus on the most critical information for prediction. This not only improves the understanding of long-term dependencies but also enhances prediction accuracy and memory retention.

Decoder: The decoder leverages the feature representations extracted by the encoder, along with the attention mechanism, to generate accurate time series forecasts. It decodes this enriched feature space into predictions for target variables such as electricity consumption, price, and wind power generation, ensuring a more precise and informed forecasting process.

To enhance clarity, we provide a comprehensive explanation of the model's operation. During the prediction process, an input sequence of fixed length, l_{seq}, denoted as seq, is fed into the encoder at each time step t. This sequence is extracted from the interval $[t - l_{seq}, t]$ and serves as the basis for forecasting a future sequence, $pred$, spanning l_{pred} time steps within the interval $[t+1, t+l_{pred}+1)$

The encoder processes sequences to extract meaningful temporal patterns, which are then leveraged by the decoder to generate the predicted sequence. During training and evaluation, the predicted sequence is compared against actual values to optimize model performance. Notably, while the Fast Fourier Transform is applied to the input sequence to enhance feature extraction, the predicted sequence bypasses the inverse FFT step. This ensures that the decoder generates predictions in the original data space, effectively preventing information leakage and preserving the integrity of the forecasting process.

Figure 2 illustrates the overall architecture of the proposed model, which will be discussed in the following sections.

A. *Feature Extraction and Prediction Mechanism*

Encoder: In this stage, the input data is processed to extract key features and long-term dependencies. This is achieved using BiLSTM networks, which capture temporal information in both forward and backward directions. The extracted information is then transformed into a compressed representation, referred to as the feature space. To further refine this representation, an attention mechanism is applied, emphasizing the most influential features.

Decoder: The decoder utilizes the extracted feature space to reconstruct future data with high accuracy. By leveraging BiLSTM networks, it processes the identified patterns within the data and generates precise forecasts for the upcoming time steps. The objective of this stage is to translate the feature space into reliable and well-aligned predictions. Consequently, the decoder plays a crucial role in ensuring accurate and trustworthy forecasting.

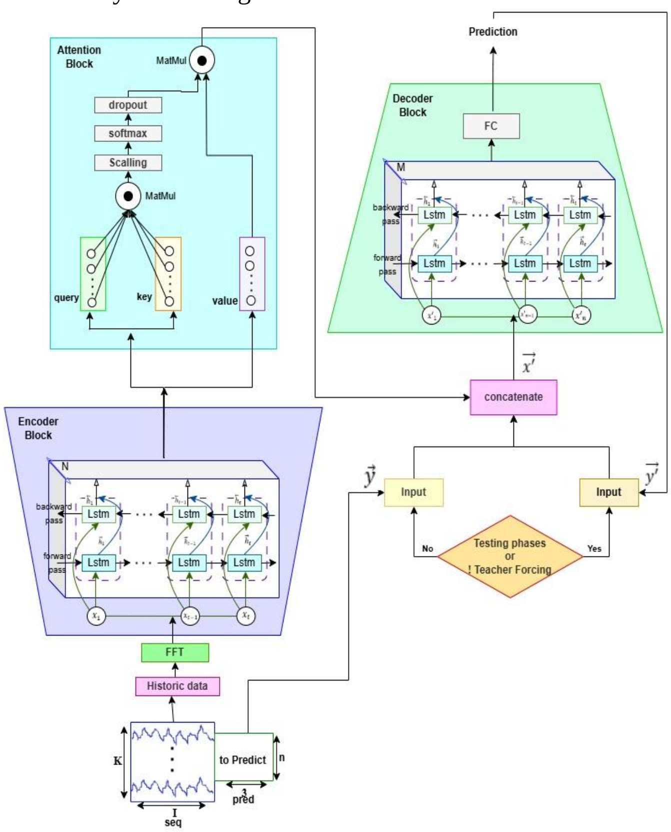

Fig. 2 Main architecture.

B. Internal Deep Networks

Long Short-Term Memory (LSTM) networks, introduced as a specialized type of recurrent neural network (RNN) [32], have become widely used in tasks such as time series forecasting and text translation. Their popularity stems from their ability to capture long-term dependencies and address challenges like the vanishing gradient problem, which often occurs in traditional RNNs when processing long sequences.

LSTMs achieve this by incorporating a **cell state**, which acts as a memory unit capable of storing information over extended time steps. Three key gates regulate the flow of information:

***Input Gate*:** Determines how much new information should be added to the cell state. It is calculated as follows:

$$i_1(t) = \sigma(x(t)u_i + h(t-1)w_i) \tag{3}$$

$$i_2(t) = \tanh\,(x(t)u_g + h(t-1)w_g) \tag{4}$$

$$i(t) = i_1(t) * i_2(t)) \tag{5}$$

Forget Gate: Controls what information should be discarded from the cell state:

$$f(t) = \sigma(x(t)u_f + h(t-1)w_f) \tag{6}$$

Cell State Update: The cell state is updated based on the retained and newly added information:

$$c(t) = \sigma(\, f(t) * c(t-1) + i(t)\,) \tag{7}$$

Output Gate: Determines the final output of the LSTM unit:

$$o(t) = \sigma(\, x(t)u_o + h(t-1)w_o) \tag{8}$$

$$h(t) = \tanh(c_t) * o(t) \tag{9}$$

Here, σ represents the sigmoid activation function, while $tanh$ denotes the hyperbolic tangent activation function. x_t is the input at time step t, u, and w are learnable weight matrices. Figure 3 illustrates the structure of an LSTM cell.

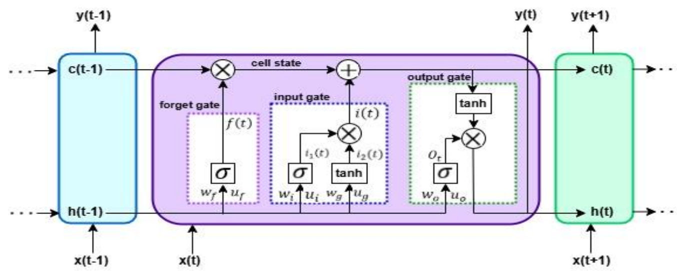

Fig. 3 LSTM unit.

Traditional LSTM networks process information in a unidirectional manner, considering only past data for predictions. However, in many applications, understanding both past and future contexts is crucial for accurate forecasting. To address this limitation, Bidirectional LSTM is employed.

BiLSTM enhances the learning process by incorporating both forward and backward LSTM layers. While the forward LSTM processes the input sequence from past to future, the backward LSTM scans the sequence in reverse, capturing future dependencies. By merging these two hidden states, BiLSTM effectively utilizes information from both the past and future, leading to more comprehensive feature extraction. The computational steps for BiLSTM are as follows:

- Forward LSTM pass:

$$\vec{h}_t = \overrightarrow{LSTM}(\, h_{t-1}, x_t, c_{t-1}), t \in [1, T] \tag{10}$$

- Backward LSTM pass:

$$\overleftarrow{h}_t = \overleftarrow{LSTM}(\, h_{t+1}, x_t, c_{t+1}), t \in [1, T] \tag{11}$$

- Final output combining both directions:

$$h_t = \vec{h}_t \oplus \overleftarrow{h}_t \tag{12}$$

C. Scaled Attention Mechanism

The scaled attention mechanism enhances deep learning models by effectively capturing complex relationships and identifying meaningful patterns in time series data. In this study, we integrate scaled attention between the encoder and decoder to extract critical information from the encoder's hidden states, improving predictions for load and electricity price. The proposed attention model, illustrated in Figure 2, operates based on three key vectors: query (Q), key (K), and value (V).

Unlike conventional approaches, we derive both Q and K from the encoder output and process them through linear layers to compress the encoded information. The attention weights are computed via element-wise multiplication and then scaled based on model dimensions, i.e., the hidden layer size (13). To normalize the scores, we apply the SoftMax function, ensuring the sum of the weights to one and forming a probability distribution (14):

$$score = \frac{(\omega_q Q . \omega_k K^T)}{\sqrt{d_{model}}} \quad (13)$$

$$P_{att} = \frac{e^{score_i}}{\sum_{j=1}^{n} e^{score_j}} \quad (14)$$

In our proposed model, the encoder is built upon a stack of Bidirectional LSTM (BiLSTM) layers, where each layer processes the input sequence in both forward and backward directions. To generate the value (v) for subsequent processing, we utilize the output sequences produced by the top BiLSTM layer. To transform the concatenated outputs into a suitable representation for downstream tasks, we apply a learnable linear layer to create value (V).

Finally, the attention weights are applied to the encoder's hidden states to extract seasonal and trend-related features (15), producing an attention vector that highlights the most influential patterns in the input data:

$$Att = P_{att} . \omega_v V \quad (15)$$

Here, ω represents learned weight vectors, and d_{model} denotes the dimensionality of the hidden layers. This attention mechanism allows the model to assign greater importance to key temporal features, improving prediction accuracy and interpretability.

D. Frequency Domain Transformation

In this study, the FFT is employed as a preprocessing step to enhance model accuracy by leveraging its ability to extract meaningful frequency-domain features. Notably, FFT is applied only to the input data of the encoder, while the decoder processes the original time series data without transformation. This ensures that the model learns to predict directly in the time domain, avoiding potential distortions from inverse FFT operations.

The integration of FFT serves several key purposes:

- **Frequency-Domain Representation:** Transforming data into the frequency domain allows the model to capture periodic patterns and seasonal trends more effectively.
- **Noise Reduction:** By analyzing frequency components, FFT helps distinguish essential patterns from noise, improving prediction reliability.
- **Feature Selection:** FFT highlights dominant frequency components while suppressing less relevant ones, optimizing feature extraction and enhancing model efficiency.
- **Enhanced Temporal Dependencies:** Incorporating frequency information aids in balancing short-term fluctuations with long-term dependencies, leading to more robust forecasts.

By leveraging FFT as a preprocessing step, the proposed approach improves pattern recognition, reduces noise interference, and enhances the model's predictive capabilities.

V. Result and Discussion

A. Data Description

This study employs the time-series dataset from the Open Power System Data (OPSD)[2] platform, which provides free access to European electricity data categorized into conventional power plants, renewable energy, and other relevant categories. The dataset, with a 60-minute resolution, covers four regions: Austria (AT), Central-North Italy (IT-CNOR), Sweden (SE_2), and the UK (GB). Our analysis focuses on forecasting electricity load demand, prices, and wind energy production using date-related information.

The dataset spans over five years, starting from January 1, 2015, with approximately 50400-time intervals. We allocate 80% of the data for training, and 20% for testing, and reserve an additional 20% of the training data for validation to implement early stopping and prevent overfitting. To further address overfitting, the data is randomly shuffled, and feature scaling is performed using the StandardScaler method, which standardizes the data to have a mean of zero and a standard deviation of one.

In all three phases, FFT operations were used as a preprocessing step for the historical data. By transforming the data into the frequency domain, this approach not only helps reduce noise in the time series but also enables the model to capture long-term patterns more efficiently. This method plays a key role in improving the overall accuracy of the model.

B. Environment and Metrics

The experiments in this study were conducted using Python 3.10, leveraging the powerful libraries of PyTorch and scikit-learn to implement the proposed algorithms. The computational environment was set up on the free version of Google Colab, utilizing a T4 GPU. This setup provided access to 12.6 GB of system RAM, 15.0 GB of GPU RAM, and 78.2 GB of disk space, ensuring efficient processing of large datasets and complex computations. All code, data, and experimental configurations are publicly available on the article's GitHub repository for reproducibility.

To assess the performance of the proposed model, we employed several statistical metrics commonly used in forecasting tasks. These include the Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE), and the Pearson correlation coefficient (r). These metrics provide a comprehensive evaluation of the model's accuracy and reliability. In 13-15 equations, N represents the number of samples, y_i is the actual $i-th$ sample, and $\hat{y}_i$ is the predicted $i-th$ sample. $\bar{x}$ is the mean of the x variable and $\bar{y}$ is the mean of the y variable. Additionally, the Pearson correlation coefficient varies between -1 and 1. If $r = 1$, it indicates a perfect direct relationship between two variables. The hyperparameters used in this study were determined through trial and error and are documented in the GitHub repository for transparency.

$$MAE = \frac{1}{N} \sum_{i=1}^{N} | y_i - \hat{y}_i | \quad (16)$$

[2] https://data.open-power-system-data.org/

$$MSE = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (17)$$

$$r = \frac{\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2 \quad \sum_{i=1}^{N} (y_i - \bar{y})^2}} \quad (18)$$

C. Experimental Result

In this paper, we present a deep learning model based on attention mechanisms for simultaneous forecasting of three features—electricity load, electricity price, and wind energy production—for future time horizons of 12, 24, and 48-time steps, using the previous 98-time steps as input. To evaluate the proposed method, we compare the accuracy of our model with several benchmark models, including an LSTM-based model similar to the approach in [9], a CNN-LSTM architecture like the one presented in [32], and a hybrid LSTM-Attention model with a query-key-value structure. Additionally, we compare our model with a seq2seq (sequence-to-sequence) model that does not incorporate attention mechanisms or FFT preprocessing. The results of these comparisons are summarized in Table 1.

D. Result Analysis

The simulation results indicate that the proposed model generally outperforms other models, particularly for Austria, Italy, and the UK. This model demonstrates higher accuracy across all time horizons and evaluation metrics compared to its counterparts. As the prediction horizon increases from 12 to 48 hours, the accuracy of all models decreases. However, the proposed model continues to exhibit superior performance in most cases, maintaining a higher level of precision relative to the other models.

One intriguing aspect of this study is the performance of the models for Sweden. While the proposed model achieves the best results for other countries, the LSTM-Attention model performs better for Sweden.

This difference in model performance can be attributed to two primary factors:

- **Characteristics of Data:** The data for SE_2 may contain patterns that align more effectively with the Query-Key-Value attention mechanism employed in the LSTM-Attention model.
- **Complexity of the SE_2 Power Grid:** The structure of SE_2's power grid and its energy production-consumption patterns might favor simpler models that achieve higher accuracy in predictions.

In summary, it is evident that increased model complexity does not always guarantee improved accuracy across all scenarios. The characteristics and patterns of the data play a crucial role in model selection. In some cases, simpler models may even yield higher accuracy when dealing with less complex datasets. This highlights the importance of tailoring the choice of model to the specific properties of the data being analyzed.

To visually compare the performance of the proposed method with other approaches examined in this paper, we present graphical illustrations of the predictions made by the discussed methods on the AT dataset for the next 12-time steps. We aggregate and plot the prediction results over 500-time steps to provide a comprehensive overview. Figure 4 displays the predicted and actual values of electricity load for the mentioned methods. Figure 5 illustrates the predictions for electricity prices, while Figure 6 shows the forecasts for wind energy production. As can be observed, the method proposed in this paper generates predictions that most closely follow the actual patterns, confirming the accuracy of the presented model. This visual analysis further supports the effectiveness of the proposed approach in capturing real-world trends across all three features.

TABLE II. MULTIVARIATE RESULTS WITH DIFFERENT PREDICTION LENGTHS (12, 24, 48)

		AT			IT-Corn			SE_2			GB		
Model	Len	MAE	MSE	r	MAE	MSE	r	MAE	MSE	r	MAE	MSE	r
proposed Model	12	**0.4071**	**0.3442**	**0.8091**	**0.3499**	**0.2757**	**0.8525**	*0.3450*	*0.239*	*0.8730*	**0.4087**	**0.3426**	**0.8144**
	24	**0.4579**	**0.4432**	**0.7467**	**0.4374**	**0.3994**	**0.7764**	*0.4129*	*0.3853*	*0.7891*	**0.4913**	**0.4371**	**0.7476**
	48	**0.5840**	**0.6333**	**0.6278**	**0.5189**	**0.4899**	**0.7275**	*0.4829*	*0.4713*	*0.7293*	*0.5480*	*0.5896*	*0.6919*
LSTM [9]	12	0.5042	0.4711	0.7252	0.4383	0.3804	0.7881	0.4038	0.3234	0.8235	0.4313	0.3564	0.8035
	24	0.5468	0.5456	0.6718	0.4964	0.4777	0.7243	0.5356	0.5590	0.6713	0.5175	0.4862	0.7237
	48	0.6490	0.7106	0.5573	0.5624	0.6570	0.6370	0.5752	0.5913	0.6495	0.5848	0.6172	0.6404
CNN-LSTM [24]	12	0.5240	0.4948	0.7113	0.5157	0.4793	0.7691	0.5921	0.5732	0.7443	0.4939	0.4764	0.7403
	24	0.5508	0.5395	0.6769	0.5622	0.5720	0.6834	0.6442	0.6688	0.6651	0.5580	0.5296	0.6937
	48	0.6030	0.6542	0.5899	0.6071	0.6890	0.5984	0.7124	0.8260	0.5666	0.6038	0.6257	0.6333
LSTM-Attention	12	*0.4216*	*0.3463*	*0.8094*	*0.3559*	*0.2545*	*0.8656*	**0.3143**	**0.2094**	**0.8900**	0.4097	0.3500	0.8062
	24	0.5245	0.4957	0.7116	*0.4403*	*0.3914*	*0.7672*	**0.3968**	**0.3247**	**0.8221**	0.4914	0.4436	0.7338
	48	*0.5898*	*0.6415*	*0.6255*	0.5288	0.4982	0.7142	**0.4420**	**0.3778**	**0.7892**	**0.5170**	**0.4713**	**0.7280**
Seq2Seq	12	0.4465	0.3806	0.8115	0.3989	0.3134	0.8567	0.4073	0.3152	0.8556	0.4102	0.3590	0.8400
	24	*0.5131*	*0.4760*	*0.7365*	0.4706	0.4323	0.7776	0.4618	0.4130	0.6423	0.4981	0.4392	0.7473
	48	0.5913	0.6364	0.6141	*0.5244*	*0.4923*	*0.7164*	0.5927	0.5719	0.6722	0.5813	0.5863	0.6496

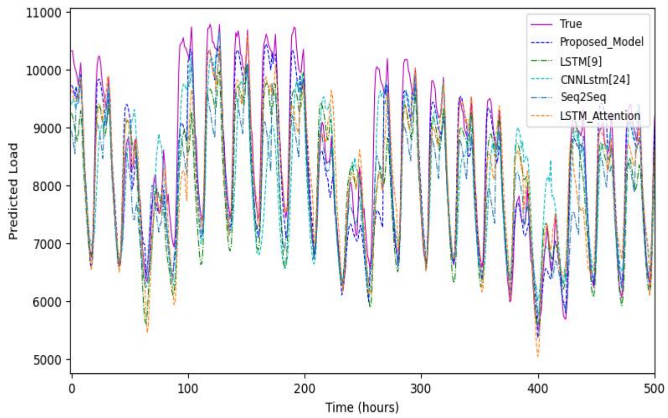

Fig. 4 Predicting load consumption for the next 12-time intervals.

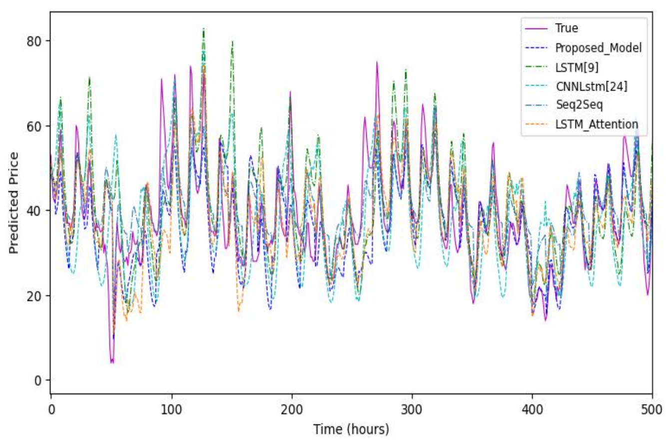

Fig. 5 Predicting Price for the next 12-time intervals.

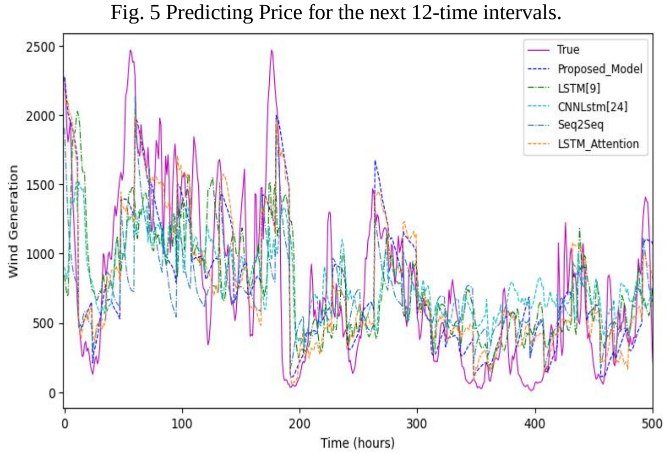

Fig. 6 Predicting Wind generation for the next 12-time intervals.

E. Practical Benefits and Implementation Considerations

In addition to improving forecasting accuracy, the proposed model offers several practical advantages that can positively impact power system operation from economic, environmental, and industrial perspectives.

From an economic standpoint, more accurate demand forecasting reduces the need for extra reserve capacity and helps lower the operational costs of generation units. It also enables better load scheduling, which can result in cost savings for both distribution companies and consumers.

Environmentally, improved forecasts can limit the use of high-emission sources, such as diesel generators or thermal plants, during peak hours. This reduction in fossil fuel reliance leads to lower carbon emissions and supports sustainability goals.

In terms of practical application, the model can be used by utility companies in their control centers for real-time planning and integration of renewable energy sources. Reliable short-term predictions provide valuable input for grid operators to make timely and informed decisions, enhancing system stability and reducing the risk of outages.

Regarding implementation, the model can be integrated with existing smart grid infrastructure using data collected from smart meters and supervisory control systems. Since many distribution networks are already adopting digital technologies, deploying the proposed method is technically feasible and aligns well with ongoing grid modernization efforts.

V. Conclusion and Future Work

In this paper, we introduced novel architecture for the simultaneous prediction of key features in smart electricity grids, including load, electricity price, and wind power generation, using a multi-variate, multi-horizon approach. The proposed deep learning model is based on an encoder-decoder structure, incorporating a stack of BiLSTM layers to effectively learn and extract temporal patterns within the data. Additionally, we employed a self-attention mechanism between the encoder and decoder to enhance the model's ability to capture dependencies among the data and improve its predictive performance. To further preprocess the time-series data, we utilized the FFT to transform it into the frequency domain, enabling more accurate identification and forecasting of underlying data structures.

Simulation results on real-world datasets demonstrate that the proposed method achieves higher accuracy in most cases compared to other approaches under investigation. Overall, the proposed model achieved an average of 0.4969 across the 12 proposed scenarios (4 datasets and 3 prediction lengths), which represents the best improvement compared to CNN-LSTM by 19.18%, followed by the seq2seq model with an improvement of 13.54%.

For future work, we aim to enhance the attention mechanism by exploring the multi-head cross-attention approach to further improve the model's performance. Furthermore, given that evaluations were conducted on datasets from multiple countries, we plan to investigate the potential of federated learning. This approach not only preserves data privacy but also has the potential to significantly boost the predictive capabilities of our model by leveraging collaborative learning across diverse datasets.

References

[1] Paterakis, N.G., O. Erdinç, and J.P. Catalão, *An overview of Demand Response: Key-elements and international experience.* Renewable and Sustainable Energy Reviews, 2017. **69**: p. 871-891.

[2] Hammad, M.A., et al., *Methods and models for electric load forecasting: a comprehensive review.* Logist. Sustain. Transp, 2020. **11**(1): p. 51-76.

[3] Khan, P.W., et al., *Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources.* Energies, 2020. **13**(18): p. 4870.

[4] Lu, D., D. Zhao, and Z. Li, *Short-term nodal load forecasting based on machine learning techniques.* International Transactions on Electrical Energy Systems, 2021. **31**(9): p. e13016.

[5] Zhang, X. and J. Wang, *A novel decomposition-ensemble model for forecasting short-term load-time series with multiple seasonal patterns.* Applied Soft Computing, 2018. **65**: p. 478-494.

[6] Bhinge, R., et al., *Toward a generalized energy prediction model for machine tools.* Journal of manufacturing science and engineering, 2017. **139**(4): p. 041013.

[7] Alhussein, M., K. Aurangzeb, and S.I. Haider, *Hybrid CNN-LSTM model for short-term individual household load forecasting.* Ieee Access, 2020. **8**: p. 180544-180557.

[8] Li, H., et al., *Ultra-short-term load demand forecast model framework based on deep learning.* Energies, 2020. **13**(18): p. 4900.

[9] Muzaffar, S. and A. Afshari, *Short-term load forecasts using LSTM networks.* Energy Procedia, 2019. **158**: p. 2922-2927.

[10] Wang, R., J. Wang, and Y. Xu, *A novel combined model based on hybrid optimization algorithm for electrical load forecasting.* Applied Soft Computing, 2019. **82**: p. 105548.

[11] Chen, W., F. Rong, and C. Lin, *A multi-energy loads forecasting model based on dual attention mechanism and multi-scale hierarchical residual network with gated recurrent unit.* Energy, 2025: p. 134975.

[12] Hosseini, S.-M., et al. *Adaptive-A-GCRNN: Enhancing Real-Time Multi-Band Spectrum Prediction Through Attention-Based Spatial-Temporal Modeling*. in *2024 14th International Conference on Computer and Knowledge Engineering (ICCKE)*. 2024. IEEE.

[13] Rahmatinia, M., S.M. Hosseini, and S.A. Hosseini Seno. *A Context-Aware Attention-Based Model for Enhanced Long term Traffic Prediction on Non-Graph-Structured Data*. In Proceedings of the 29th International Conference on Computer Society of Iran (CSI), 2025.

[14] Rahmatinia, S.-M. and S.-A.H. Seno. *Enhancing NTMA with Simultaneous Multi-QoS Parameter Prediction using Transformer-Based Deep Learning*. in *2024 11th International Symposium on Telecommunications (IST)*. 2024. IEEE.

[15] Aly, H.H., *A proposed intelligent short-term load forecasting hybrid models of ANN, WNN and KF based on clustering techniques for smart grid.* Electric Power Systems Research, 2020. **182**: p. 106191.

[16] Tang, L., Y. Yi, and Y. Peng. *An ensemble deep learning model for short-term load forecasting based on ARIMA and LSTM*. in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 2019. IEEE.

[17] Qin, J., et al., *Multi-task short-term reactive and active load forecasting method based on attention-LSTM model.* International Journal of Electrical Power & Energy Systems, 2022. **135**: p. 107517.

[18] Chen, J., J. Xiao, and W. Xu. *A hybrid stacking method for short-term price forecasting in electricity trading market*. in *2024 8th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. 2024. IEEE.

[19] Asiri, M.M., et al., *Short-term load forecasting in smart grids using hybrid deep learning.* IEEE Access, 2024. **12**: p. 23504-23513.

[20] Zhao, S., et al., *Residential energy consumption and price forecasting in smart homes based on the internet of energy.* Sustainable Energy Technologies and Assessments, 2025. **73**: p. 104081.

[21] Lin, J., et al., *Short-term load forecasting based on LSTM networks considering attention mechanism.* International Journal of Electrical Power & Energy Systems, 2022. **137**: p. 107818.

[22] Dai, Y., et al., *Improving the Bi-LSTM model with XGBoost and attention mechanism: A combined approach for short-term power load prediction.* Applied Soft Computing, 2022. **130**: p. 109632.

[23] Bu, S.-J. and S.-B. Cho, *Time series forecasting with multi-headed attention-based deep learning for residential energy consumption.* Energies, 2020. **13**(18): p. 4722.

[24] Wu, K., et al., *An attention-based CNN-LSTM-BiLSTM model for short-term electric load forecasting in integrated energy system.* International Transactions on Electrical Energy Systems, 2021. **31**(1): p. e12637.

[25] Verheijen, N.N.D.P., *Explainable non-recurrent models for load forecasting*. 2025, University of Twente.

[26] Oyedokun, J.T., *Big data analytics for demand response in smart grids*. 2021, University of Glasgow.

[27] Pentsos, V., et al., *A Hybrid LSTM-Transformer Model for Power Load Forecasting.* IEEE Transactions on Smart Grid, 2025.

[28] Liu, Y., et al., *Short-Term Load Forecasting Model Based on Time Series Clustering and Transformer in Smart Grid.* Electronics, 2025. **14**(2): p. 230.

[29] Sarker, M.A.A., et al., *Enhancing smart grid load forecasting: An attention-based deep learning model integrated with federated learning and XAI for security and interpretability.* Intelligent Systems with Applications, 2024. **23**: p. 200422.

[30] Gupta, H., et al., *FedGrid: A Secure Framework with Federated Learning for Energy Optimization in the Smart Grid.* Energies, 2023. **16**(24): p. 8097.

[31] Zhang, D., *Optimization and research of smart grid load forecasting model based on deep learning.* International Journal of Low-Carbon Technologies, 2024. **19**: p. 594-602.

[32] Rafi, S.H., S.R. Deeba, and E. Hossain, *A short-term load forecasting method using integrated CNN and LSTM network.* IEEE access, 2021. **9**: p. 32436-32448.