The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Electricity Energy Consumption Forecasting Using Different Deep Learning Models

Karwan Salih Rasul[1], Nzar Abdulqadir Ali[2]

[1]Technical Institute of Sulaimani, Sulaimani Polytechnic University, Sulaimani, Iraq
[2] Department of Statistics and Informatics, University of Sulaimani, Sulaimani, Iraq

Email: karwan.rasul@spu.edu.iq[1], nzar.ali@univsul.edu.iq[2]

Abstract:

Accurate forecasting of electricity energy requirements has become critical for current energy systems, which are facing increasing challenges due to industrial development, population growth and the integration of green energy. The research evaluates the capacity of machine learning and deep learning algorithms to forecast demand of electricity energy consumption using dataset information from central electricity control office of Iraqi Kurdistan Region Government (KRG). Electricity energy consumption data analysis requires predictive models that are more sophisticated than traditional methods such as Autoregressive Integrated Moving Average (ARIMA) and exponential sequences, as these methods fail to deal with complex nonlinear patterns and high-frequency oscillations. Therefore, this study presents deep learning models using Long Short-Term Memory (LSTM) and Recurrent Neural Network (RNN), where the test results on the electricity energy data from KRG showed the superiority of the LSTM model in terms of accuracy and stability compared to RNN, the data was analyzed using multiple performance measures such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to confirm the efficiency of the proposed model. The results show LSTM superior performance than the RNN model based on the metrics provided, as the RMSE (187.25) and MAE (139.78) values are lower compared to the RNN (RMSE = 230.34, MAE = 232.276), indicating that LSTM forecasts are more accurate with fewer errors. In addition, the coefficient of determination (R^2) of the LSTM model (0.961) is higher than that of RNN (0.941) Finally, this model can be applied in intelligent energy systems to improve load management efficiency and reduce waste.

Keywords: Electricity Energy Demand Forecasting, LSTM, RNN, Machine Learning, Deep Learning, Forecasting.

DOI: http://dx.doi.org/10.25098/9.1.29

الملخص:

أصبح التنبؤ الدقيق بمتطلبات الطاقة الكهربائية أمرا بالغ الأهمية لأنظمة الطاقة الحالية، التي تواجه تحديات متزايدة نتيجةً للتطور الصناعي والنمو السكاني ودمج الطاقة الخضراء. يُقيّم البحث قدرة خوارزميات التعلم الآلي والتعلم العميق على التنبؤ بالطلب على استهلاك الطاقة الكهربائية باستخدام معلومات من قاعدة بيانات دائرة التحكم المركزي للكهرباء في حكومة إقليم كردستان العراق.

يتطلب تحليل بيانات استهلاك الطاقة الكهربائية لنماذج تنبؤية أكثر تطورا من الطرق التقليدية، مثل المتوسط المتحرك المتكامل الانحداري الذاتي (ARIMA) والمتتاليات الأسية، نظرا لعجز هذه الطرق عن التعامل مع الأنماط غير الخطية المعقدة والتذبذبات عالية التردد. لذلك، تقدم هذه الدراسة نماذج تعلم عميق باستخدام الذاكرة طويلة المدى قصيرة المدى (LSTM) والشبكة العصبية المتكررة (RNN). أظهرت نتائج الاختبارات على بيانات الطاقة الكهربائية من (KRG) تفوق نموذج LSTM من حيث الدقة والاستقرار مقارنةً بالشبكة العصبية المتكررة (RNN). وقد تم تحليل البيانات باستخدام مقاييس أداء متعددة، مثل متوسط الخطأ المطلق (MAE) وجذر متوسط مربع الخطأ (RMSE)، لتأكيد كفاءة النموذج المقترح.

تُظهر النتائج أداءً أفضل لنموذج LSTM مقارنةً بنموذج RNN بناءً على المقاييس المُقدمة، حيث إن قيمتي RMSE (187.25) وMAE (139.78) أقل مقارنةً بنموذج RNN (RMSE=230.34) ، RNN(MAE= 232.276 ، مما يُشير إلى أن تنبؤات LSTM أكثر دقةً مع أخطاء أقل. بالإضافة إلى ذلك، فإن معامل التحديد (R^2) لنموذج LSTM (0.961) أعلى من معامل التحديد (RNN) (0.941). وأخيرًا، يُمكن تطبيق هذا النموذج في أنظمة الطاقة الذكية لتحسين كفاءة إدارة الأحمال وتقليل الهدر.

الكلمات المفتاحية: التنبؤ بالطلب على الطاقة الكهربائية، LSTM، RNN، التعلم الآلي، التعلم العميق، التنبؤ.

پوختە:

پێشبینیکردنی وردی پێداویستییەکانی وزەی کارەبا بووەتە شتێکی گرینگ بۆ سیستەمی وزەی ئێستا، کە بەهۆی گەشەسەندنی پیشەسازی و زیادبوونی ژمارەی دانیشتووان و یەکخستنی وزەی سەوز ڕووبەڕووی ئاستەنگەکانی زیاتر دەبنەوە. توێژینەوەکە توانای فێربوونی ئامێر و ئەلگۆریتمەکانی فێربوونی قووڵ هەڵدەسەنگێنێت بۆ پێشبینیکردنی خواست لەسەر بەکارهێنانی وزەی کارەبا بە بەکارهێنانی زانیارییەکانی داتا سێت لە فەرمانگەی ناوەندی کۆنترۆڵکردنی کارەبای حکومەتی هەرێمی کوردستانی عێراق.

شیکاری داتاکانی بەکارهێنانی وزەی کارەبا پێویستی بە مۆدێلی پێشبینیکراو هەیە کە ئاڵۆزتر بێت لە شێوازە تەقلیدییەکانی وەکو تێکڕای جووڵەی یەکگرتووی خۆپاشکەوتن (ARIMA) و زنجیرە ڕیزبەندییەکان، چونکە ئەم ڕێگایانە شکست دەهێنن لە مامەڵەکردن لەگەڵ نەخشە ناهێڵییە ئاڵۆزەکان و لەرزینەکانی فریکوێنسی بەرز. بۆیە ئەم توێژینەوەیە مۆدێلی فێربوونی قووڵ دەخاتە ڕوو بە بەکارهێنانی بیرگەی کورتخایەنی درێژخایەن (LSTM) و تۆڕی دەماری دووبارەبووەوە (RNN)، کە ئەنجامی تاقیکردنەوەکان لەسەر داتاکانی وزەی کارەبا لە (KRG)ەوە باڵادەستی مۆدێلی LSTMی لە ڕووی وردبینی و سەقامگیرییەوە بەراورد بە RNN نیشان دا، داتاکان بە بەکارهێنانی چەندین پێوەرەکانی ئەدای کارکردن وەک هەڵەی مامناوەندی ڕەها (MAE) و هەڵەی چوارگۆشەی مامناوەندی ڕەگ شیکرانەوە (RMSE) بۆ پشتڕاستکردنەوەی کارایی مۆدێلی پێشنیارکراوی.

ئەنجامەکان ئەدای باشتری LSTM لە مۆدێلی RNN نیشان دەدەن بە پشتبەستن بە پێوەرەکانی پێشکەشکراو، چونکە بەهاکانی RMSE =(187.25) و MAE= (139.78) کەمترن بە بەراورد بە(RNN (RMSE =230.34، MAE = 232.276، ئەمەش ئاماژەیە بۆ ئەوەی پێشبینییەکانی LSTM وردترن لەگەڵ هەڵەی کەمتر. سەرەڕای ئەوە، ڕێژەی دیاریکردن (R^2)ی مۆدێلی LSTM (0.961) زیاترە لە ڕێژەی RNN (0.941) لە کۆتاییدا، دەتوانرێت ئەم مۆدێلە لە سیستەمی وزەی زیرەکدا بەکاربهێنرێت بۆ باشترکردنی کارایی بەڕێوەبردنی بار و کەمکردنەوەی بەفیڕۆدان.

کلیلە وشە: پێشبینی داواکاری وزەی کارەبا، LSTM، RNN، فێربوونی ئامێر، فێربوونی قووڵ، پێشبینی.

DOI: http://dx.doi.org/10.25098/9.1.29

1. Introduction

Energy consumption forecasting has become increasingly important for better managing energy usage and optimizing demand response strategies. In large scale, the efficient management of urban areas and energy consumption in buildings is the major challenge for the future. [1].

Accurate scalable and timely forecasting methods are essential for the world to handle escalating energy demand from industrialization and population growth as well as renewable energy deployment. Reputable energy demand predictions allow improved grid operations and minimize operational costs that support sustainability through alternative energy systems [2].

Throughout numerous years the energy sector depended on three traditional predictive models including ARIMA together with exponential smoothing and seasonal decomposition models. Ordinary statistical prediction models perform well within limited circumstances yet they cannot provide reliable outcomes when processing nonlinear data with high-frequency changes and complicated seasonal patterns which characterize modern energy consumption [3].

Machine learning in time series is a branch of artificial intelligence that focuses on analyzing time-related data to predict future trends or detect patterns. This technology is used in areas of predicting energy consumption [4].

Deep learning models particularly Recurrent Neural Networks (RNN) and Long Short-Term Memory (LSTM) demonstrate superior performance in forecasting applications because they learn sophisticated temporal patterns while handling changing patterns in data. The models demonstrate better performance than traditional methods particularly when they receive sufficient training data with numerous features [5].

The effectiveness of deep learning models depends heavily on input data requirements while their success also depends on suitable feature selection because this process advances both performance and generalization capabilities [6].

2. Literature Review

Khan et al. developed a new power cost forecast which implemented multi-head self-attention mechanisms together with Convolutional Neural Networks (CNNs). The researchers applied their proposed model to Ontario energy market datasets from 2020 while obtaining outstanding results through best average Mean Absolute Percentage Error (MAPE) of 1.75% and Root Mean Square Error (RMSE) of 0.0085. This research proved the capability of integrating state-of-the-art machine learning algorithms for power price forecasting to deliver accurate results suitable for real-world energy market applications [7].

The error-compensation model developed by Ghimire et al. integrates Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN) along with Variational Mode Decomposition (VMD) algorithms for predicting half-hourly electricity prices. These advanced algorithms received a fresh configuration to enhance forecasting capabilities. The researchers validated their solution through the accuracy results which indicated substantial forecast advancements according to the Legates and McCabe Index versus benchmark models [8].

DOI: http://dx.doi.org/10.25098/9.1.29

Biligili et al. implemented four predictive models which included ANFIS-GP, ANFIS-SC, ANFIS-FCM as well as LSTM for Turkey's energy consumption forecasting while assessing their predictive abilities. LSTM demonstrated the most successful performance according to the study results and yielded a 1.91% MAPE measurement. Results proved that LSTM surpasses ANFIS-based models because ANFIS-based models delivered MAPE values of 4.47%, 3.21% and 2.34%. The research showed that LSTM succeeded in anticipated energy consumption trends while proving suitable for Turkish energy market applications [9].

Kuo et al. presented a short-term EP prediction system by incorporating CNN technology into LSTM architecture and applying it to real EP dataset. The model produced superior performance than SVM, RF, MLP, CNN and LSTM when measured by MAE at 8.84. The predictability of the model proved higher through its MAE value of 8.84 which exceeded LSTM (9.82), CNN (9.80), MLP (9.86), RF (9.20), and SVM (28.98) respectively [10].

The CNN-LSTM model built by Heidarpanah et al. served for EP forecasting in Iran's electricity market. The CNN-LSTM model received performance evaluation through its comparison with Multivariate Linear Regression (MLR), SVM, ANN, ANFIS, and ANN-Genetic Algorithm. Their research showed CNN-LSTM provided the most suitable solution because it demonstrated superior robustness compared to the MLR and SVM models unable to handle EP time-series oscillations [11].

Farsi et al. proposed an electricity load consumption and power consumption prediction system based on Convolutional Neural Networks (CNN) with Long Short-Term Memory (LSTM) networks and PLCNet. Malaysian hourly load consumption data and daily power consumption data from Germany formed the basis of the research. The authors determined the hybrid PLCNet model provided superior performance than other models by showing substantial accuracy gains through an enhancement from 83.17% to 91.18% for German data and reaching 98.23% for Malaysian data. Research evidence demonstrates that hybrid models have potential to boost accuracy measurements across different energy markets [12].

Khafaf et al. Proposed Long Short-Term Memory (LSTM) model to forecast electricity consumption for three- and fifteen-days energy demand for different energy load across each month in the year. The research utilized machine learning approaches for time series prediction, employing the LSTM model and introducing an innovative approach to transform time into a feature for the training phase to enhance performance. LSTM demonstrate a strong architecture for both short and median forecasting. Results shows that how to better manage peak energy demand with MAPE at 3.15% [13].

Azzone et al. They propose a new modelling approach that incorporates trend, seasonality and weather conditions, as explicative variables in a shallow Neural Network with an autoregressive feature, by emphasizing probabilistic forecasts and utilizing illustrative variables such as direction, seasonality, and weather conditions. They introduce A novel model utilizing a shallow neural network with a self-regressive characteristic, integrating trend, seasonality, and meteorological parameters as determinants of prediction. Outstanding outcomes in the rigorous forecasting of energy consumption were achieved with a one-year test cohort (daily data from New England, US), demonstrating the model's superiority over conventional models. [14]

DOI: http://dx.doi.org/10.25098/9.1.29

Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Mahjoub et al. They compare three machine learning algorithms (LSTM, GRU, Drop-GRU) for the prediction of short-term energy consumption utilizing time series data, with the objective of enhancing energy management and minimizing waste. Energy consumption data from select French cities were utilized in a time series model, employing three machine learning algorithms (LSTM, GRU, Drop-GRU) to generate forecasts and evaluate their efficacy. LSTM forecasts can be utilized for proactive decision-making in electrical load control, particularly when consumption surpasses allowable thresholds, hence enhancing power quality and equipment maintenance. Their results indicated that the Drop-GRU was superior to the GRU and the LSTM. [15]

Ghojogh et al. This paper elucidates recurrent neural networks (RNN) and LSTM networks, addressing issues of vanishing and exploding gradients in long-term dependencies, along with recommended remedies such as semi-fixed weight arrays and portal cells in LSTM. The paper commences with the dynamic system and temporal reverse propagation of RNNs, then addressing solutions like leak units and echo networks, and concluding with an in-depth elucidation of LSTM gates and cells. This study serves as an extensive instructional resource on RNN and LSTM, emphasizing the challenges and solutions in modelling long-term dependencies. [16]

3. Background Theory

The Figure (1) methodology uses raw data from a CSV file which contains Date Time and Demand columns during its initial data preprocessing phase. A specific format is applied to read the Date Time column to maintain correct date presentation while processing data with day-first date organization. Data processing on the Date Time column allows chronological sorting that maintains proper time series arrangement for analysis.

The data receives Min-Max scaling normalization before the model building process where sequences are created for LSTM and RNN. The data separates into training which contains 80% of data while the remaining 20% goes to test mode. The design of the LSTM model incorporates 16 to256 units in its LSTM layer and a dense output layer that applies an Adam optimizer with an MSE loss function. The training process continued for 50-100 epochs while checking for overfitting risks.

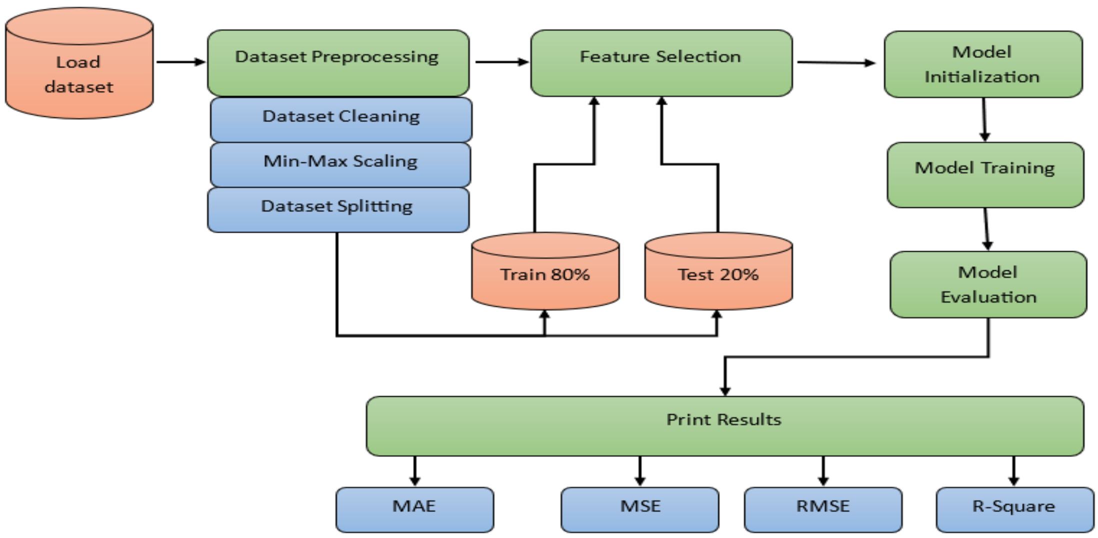

Figure 1. Model Methodology

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya PP: 62-81
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

3.1 Dataset Description

The data set consists of 4596 rows with two columns: Date Time and Demand. The Date Time column is the observation date of energy demand, and the Demand column is the energy demand for the respective date.

1. Date Time: Initially of object type, this column contains the timestamp of the energy demand data.
2. Demand: This column has integer values for the energy demand in megawatt (MW) units.

The descriptive statistics reveal a description of the Demand column, providing useful information such as the mean, standard deviation, minimum, maximum, and percentiles.

1. Count: The Demand column contains 4596 non-null values, confirming that there are no missing values in the data set.
2. Mean: The average energy demand is 4131.96 MW.
3. Standard Deviation: The figures for demand have a deviation of 1318.39 MW units from the mean.
4. Min and Max: The lowest demand is 1647, and the highest recorded demand is 7658 MW.
5. Percentiles: 25th, 50th (median), and 75th percentiles indicate that the distribution of demand is rather scattered, ranging between 3166.5 MW and 4952.25 MW.

3.2. Feature Engineering and modeling

Feature engineering is one of the most important steps in the methodology. In general, it is the process and transforming the features for a modeling problem. A feature is a measurable aspect of a phenomenon which is relevant to the prediction. Features should be carefully aggregated to improve the performance of the model. In time-series problems, it is important to extract and select the right features that capture the relevant variables and patterns of the observed phenomenon. For LSTM model analysis, we are conducting feature engineering by extracting day, monthly, and yearly trends from the "Datetime " column. This includes creating features such as day of the month, month and year, along with the original demand values. These engineered features will help the LSTM model capture temporal patterns and improve forecasting accuracy. As show in table 1. This domain knowledge guides the construction of new features to test, often in an experimental way the data is 4596 daily, 660 weeks and 151 months.

Table 1 Date Time indexing and extracting features

	Date Time	**Demand**	**Year**	**month**	**day**
0	01/02/2012	3111	2012	2	1
1	02/02/2012	3022	2012	2	2
2	03/02/2012	2841	2012	2	3
...	...		...	...	...
4593	29/08/2024	5140	2024	8	29
4594	30/08/2024	4815	2024	8	30
4595	31/08/2024	4976	2024	8	31

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

3.3. LSTM Model Configuration

An essential component of long-term networks, the LSTM (Long Short-Term Memory) cell's structural structure is seen in Figure (2). The cell is made up of a number of primary gateways, such as the input gate, which chooses which fresh data should be stored, the output gate, which chooses which information should be discarded, and the forget gate. This manages the data that is forwarded to the following stage. To control the information flow, these gates use activation functions like Tanh and Sigmoid. Point-wise multiplication and point-wise addition are used to describe in-cell computations, which help the cell avoid the vanishing gradient issue that traditional neural networks face and preserve significant information over time. Because of their intricate design, LSTM cells process information efficiently. Sequential data such as energy demand forecasting or time series analysis.

In addition to the hidden state h(t) and the state of cell c(t−1), the image includes the LSTM cell's primary gates, which are the (ft) forgetfulness gateway, the (it) input gate, and the (Ot) output gateway. An overview of the information flow within the cell is provided by the input, which is also known as x(t). Machine translation and speech recognition are two applications that benefit from this methodology.

LSTM Cell Architecture

The following figure illustrates the internal operations of a single LSTM unit, highlighting the Forget, Update, and Output gates, along with their associated computations.

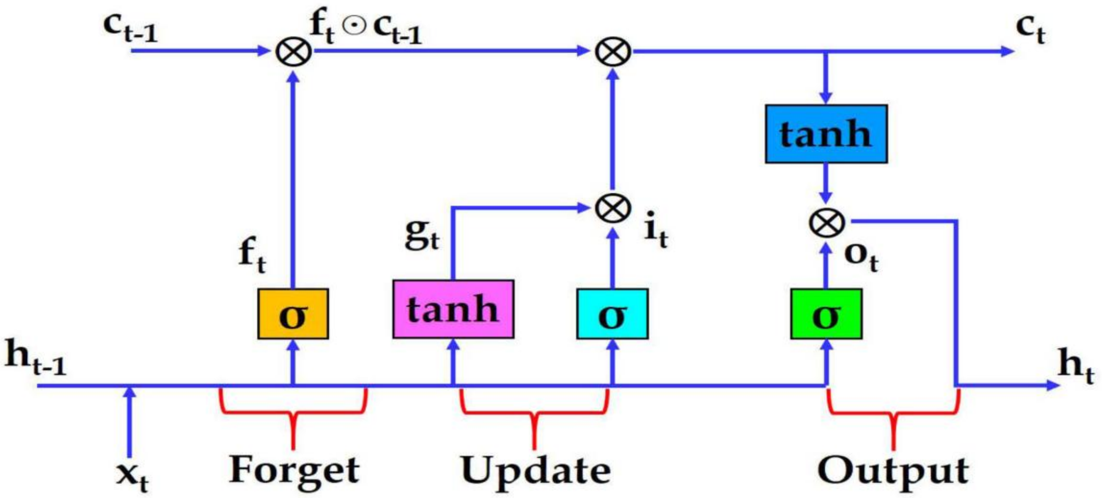

Figure (2) General structure of LSTM model [19]

DOI: http://dx.doi.org/10.25098/9.1.29

Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

3.3.1 Forget Gate

Each LSTM cell consists of several gates. The most widely known are the forget gate, the input gate, and the output gate. The forget gate first takes the element-wise multiplication of the previous cell state and the output of the forget gate. Next, a value between 0 and 1 is forced by applying the sigmoid activation function, deciding which information from the previous cell state will be discarded. Thus, the forget gate is an important gate regarding memory information. In an LSTM-based system, some use cases make a better performance by forgetting quickly the previous information at some points, while some cases require a better memory of even small differences in two adjacent inputs The forget gate plays a role in convincingly learning long-range dependencies, and striking a proper balance between overfitting and redundancy problems, determines whether data should be removed from the cell state. The equation for the forget gate is:

$ft = \sigma(Wf \cdot [ht-1, xt] + bf)$ ………………………………………………………. (1)

where:

- W_f : symbolizes the forget gate's weight matrix.
- $[\mathbf{h_{t-1}}, \mathbf{x_t}]$: indicates that the current input and the prior concealed state have been concatenated.
- $\boldsymbol{b_f}$: Is the Forget Gate biased.
- **σ**: is the activation function of the sigmoid.

3.3.2 Input Gate

Among the four gates in a long short-term memory cell, the input gate is the one that navigates when new information should be introduced. While the forget gate controls the component-wise scale of the cell state, the input gate enables learning immediately from the latest available inputs with element-wise control over the introduction of new values to the cell state. The input gate prominently uses an element-wise multiplication operation. While the individual components of the incoming and cell state vectors can be high, their element-wise multiplication keeps the balance of the overall scale between the new values and the existing state. Important information is added to the cell state by the input gate. The data is first controlled by the sigmoid function, and then it is remembered using inputs h_(t-1) and x_t in a fashion similar to the forget gate. After that, a vector with an output range of -1 to +1 is created using the tanh function, which contains every conceivable value from ht-1 and x_t. Lastly, the vector values are multiplied by the prescribed values to acquire the pertinent information. The equation for the input gate is:

$it = \sigma(Wi \cdot [ht-1, xt] + bi)$ ……………………………………………………………. (2)

- *Wi* is the weight matrix for the input gate. xt: Current input
- ht−1: Previous hidden state. ct−1: Previous cell state
- *bi* is the bias term for the input gate.

The candidate cell state (gt) is computed

gt=tanh (Wg·[ht−1,xt]+bg) ………………………………………………………. (3)

DOI: http://dx.doi.org/10.25098/9.1.29

Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

3.3.3 Output Gate

Information is stored in the cell state that is potentially useful and can be slightly modified or preserved by the gates.

The output gate of an LSTM is the mechanism by which a recurrent neural network cell decides what to output at a time step. The output at each hidden unit is a nonlinear, saturating function of the product of the activation value of that unit and its output weights. Output gates also play a role in moving information back to input layers in sequence generation tasks. A similar mechanism is at play in the output gate of an LSTM. The job of the output gate is to govern the flow of information in a recurrent neural network from the layer with LSTM cells to the next layer or the output.

the output gate is a central factor in deciding the behavior of the LSTM cell, similar to the effect of input gates on the hidden gate in a multiplicative LSTM. The output gate plays a crucial role in the articulation of the model actions and demonstrates the delicate balance that emerges in how a long-term sequence memory is organized into relevant, actionable outputs. Ample experimental validation of this hypothesis is provided throughout a set of common LSTM model evaluations. Finally, a simple, bi-directional attention mechanism with clearer interpretative value is presented to compute the Output Gate Activation.

The equation for the input output is:

$ot=\sigma(Wo\cdot[ht-1,xt]+bo)$ ………………………………………………………… (4)

The hidden state (*ht*) is then computed as

$ht=ot\odot\tanh(ct)$ ……………………………………………………………… (5)

where:

- *Wo* is the weight matrix for the output gate. ct: New cell state
- *bo* is the bias term for the output gate. ht: New hidden state (also the output of the LSTM cell)
- σ → Sigmoid activation function

3.3.4 Cell State Update

The equation for the cell state update is:

$ct = ft\odot ct-1 + it\odot \tilde{c}t$

……………………………………………………………………………………… (6)

- σ is the sigmoid activation function, tanh is the hyperbolic tangent activation function.
- ⊙ denotes element-wise multiplication,
- [ht−1, xt] is the concatenation of the previous hidden state and the current input

Table (2) provides a short description of the layers used in the LSTM model. Each layer is designed to take and process the sequential data with some specific settings of units, activation function, and output shape. The LSTM layer is the most important part of the model that acquires temporal dependencies in the input sequences, and the Dense layer provides the final output prediction. Knowing the structures and parameters of these layers facilitates the interpretation of the model's performance and workings.

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya PP: 62-81
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 2. LSTM Model Configuration

Layer	Units Shape	Activation Function	Output Shape
LSTM Layer	16,32,64,128,256 units	ReLU	(batch_size, 16,32,64,128,256)
Dense Layer	1 unit		(batch_size, 1)

3.4. RNN Model Structure

This Figure (3) show with an emphasis on hidden weights and units, the image displays a diagram illustrating the architecture of a recurrent neural network (RNN). Weights, X inputs, Y outputs, and hidden units h are examples of major components. The connections between hidden units at various time steps, such as ht−1, ht, and ht+1, depict the iterative process and demonstrate how the inputs Xt−1, Xt, and Xt+1 impact them. Additionally, elements like "unfold" are mentioned, which show how the network is being deployed over time.

Redesigned RNN Architecture

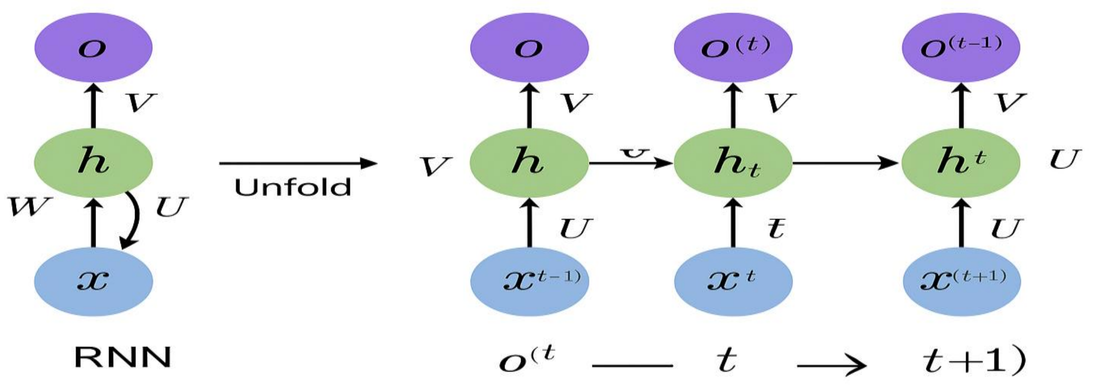

Figure (3) General structure of RNN model [20]

RNN Unfolded in Time:

RNNs process sequences one element at a time, maintaining a hidden state that is updated at each time step. When you "unfold" an RNN, you visualize each time step explicitly.

- RNN Components: x: Input vector, h: Hidden state vector, o: Output vector, (U, W, V): Weight matrices

3.4.1 Hidden state update:

$h(t) = tanh(Wh(t-1) + Ux(t))$ ………………………………………………………………………. (7)

- h(t): hidden state at time t, h(t−1): previous hidden state, x(t): input at time t
- W: hidden-to-hidden weight matrix, U: input-to-hidden weight matrix
- tanh: activation function (commonly used)

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

3.4.2 Output computation

o(t)=SoftMax (Vh(t)) …………………………………………………………… (8)

- o(t): output at time t , V: hidden-to-output weight matrix
- SoftMax: used to get a probability distribution over the output

The Simple RNN model architecture is shown in Table (3) which depicts the Simple RNN model architecture, its layers, respective units or shapes, activation functions, and output shapes. The Simple RNN layer is used to handle temporal dependencies in the input sequence, and the Dense layer gives the final output prediction. All these layers handle different things in sequential data processing, from pattern capture to output prediction.

Table 3. RNN Model Structure

Layer	**Units Shape**	**Activation Function**	**Output Shape**
Simple RNN Layer	16,32,64,128,256units	Relu	(batch_size, 16,32,64,128,256)
Dense Layer	1 unit	-	(batch_size, 1)

3.5. Training Parameters

The key training parameters used in the model appear in Table (4) to control different training features including number of epochs, batch size, optimization method and loss function. The training procedure occurs more than 100 times as the entire dataset passes through the model. The model weight updating process occurs after processing each batch size value set by default. The Adam optimizer functions as the weight adjustment method during training because it demonstrates remarkable efficiency in gradient-based optimization. Mean Abustle Error (MAE) and Root Mean Squared Error (RMSE) functions as the loss metric since it is a traditional regression selection for minimizing the squares of actual and predicted value discrepancies. The model benefits from data shuffling that enables random mixing of training data at epoch beginnings to enhance its ability to predict unknown data.

Table 4. Deep Learning Training Parameters

Parameter	**Value**
Epochs	100 epochs
batch_size	16,32,64,128,256
Optimizer	Adam optimizer
loss_function	MAE,RMSE
Shuffle	True

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya PP: 62-81
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

4. Results and Discussions

The following sections provide a cautious analysis of the forecasting models and training results, including LSTM and RNN, each model's performance is contrasted using training curves, observing their ability to learn and generalize from previous data. A rigorous comparison of the models by critical performance metrics such as MAE, RMSE, and R^2 is also provided. This is followed by a comparison of our study with similar research in the field, illustrating several forecasting methods and their ability to predict electricity demand. The comparison draws the superiority of the models that have been employed in this research and offers knowledge on how these can be utilized for real energy markets.

4.1. LSTM Training Results

The figure (4) presents the training curve for the LSTM model which depicts both the training loss and validation loss over 100, we use stops training function to stop validation loss when doesn't improve for 5 consecutive epochs. The training loss drops quickly at the beginning because the model absorbs the data patterns effectively. The model demonstrates a major drop in performance which shows it effectively identifies the key relationships within the data. The training loss (blue line) alongside validation loss (orange line) reaches tiny values before maintaining this steady point at zero. The model has successfully minimized training errors to such an extent that it stops overfitting. The model shows effective generalization ability across unseen data because the training loss matches closely with the validation loss. The model achieved satisfactory results by successfully learning time-series patterns in the data because its loss values stayed at consistent low levels.

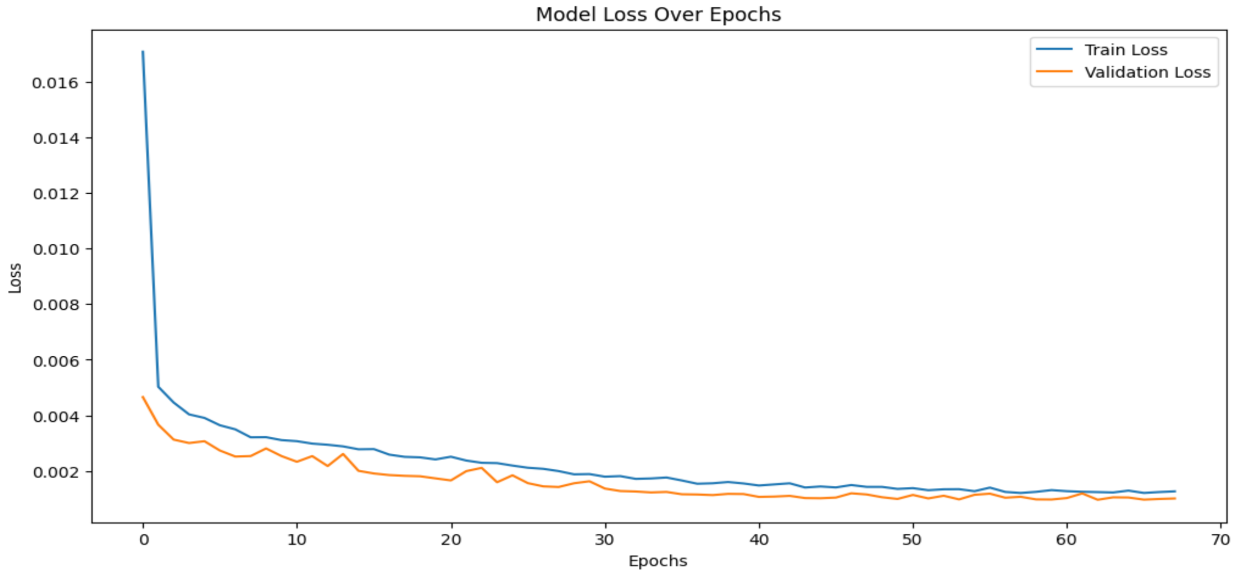

Figure 4. LSTM Training Results

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya PP: 62-81
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

4.2. RNN Training Results

The training curve of RNN in Figure (5) shows both training loss and validation loss over 100 epochs. Similar to the training curve of LSTM, training loss has a steep drop-off in the first few epochs before it plateaus at a point close to a low value. This is an indication that the model picks up the significant patterns in the data quickly during early training.

But unlike the LSTM model, there is a small oscillation in the validation loss, especially after the initial steep drop. Such small oscillations in the validation loss during the course of training may indicate some instability in the model's performance or responsiveness to small changes in the validation set. Despite these small oscillations, both training and validation loss values are low and continue to converge with progressing training.

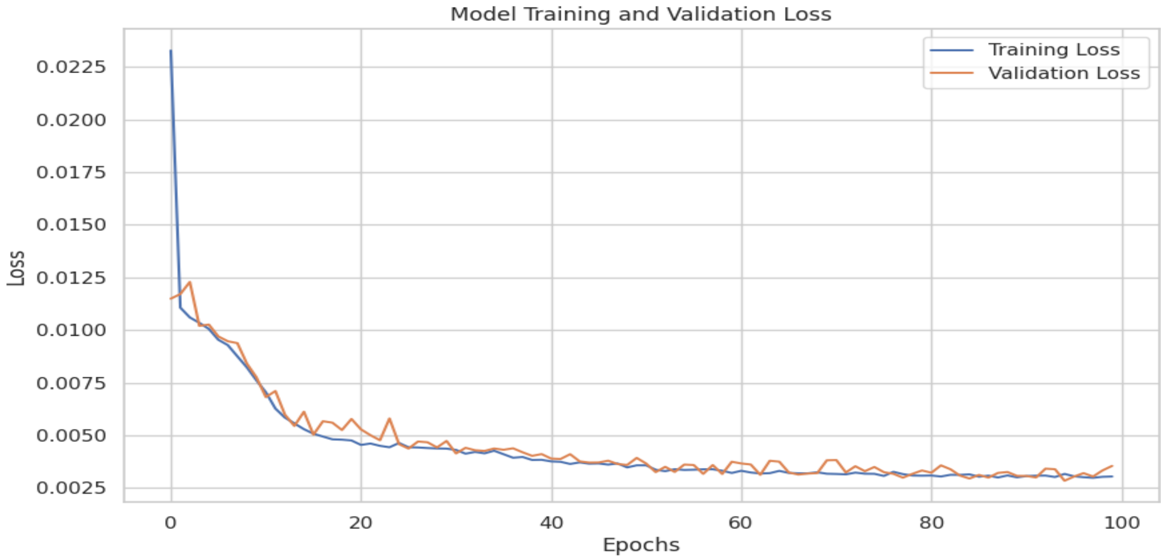

Figure 5. RNN Training Results

Figure (6) shows the fluctuation in the demand for the date from February 2. 2012 to 31/08/2024. We can see that the data is a bit cyclic and projects no trend or seasonal pattern. The plot shows that the peak energy consumption was around December (2019), and the minimum energy consumption was around Apr 2012.

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya PP: 62-81
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

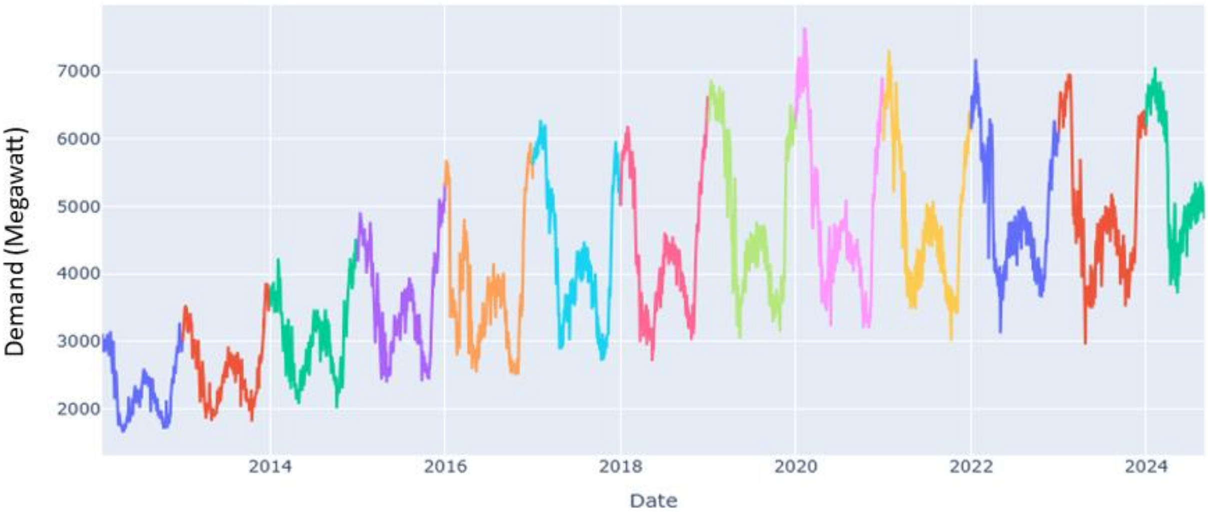

Figure 6. energy consumption according to daily

The Figure (7) shows electricity demand data measured in megawatts over several years, starting from 2012 to 2024. It can be seen that demand has experienced marked fluctuations during this period, rising and falling between different years. In some years such as 2014 and 2018, demand seems to have peaked, while in others such as 2016 and 2020, it may have declined significantly Average demand during this period shows relative stability, with a general trend of increasing as the years progress, especially as 2024 approaches. These changes may reflect factors such as population growth, industrial development, or changes in energy efficiency Data provides valuable insight into energy infrastructure development plans, helping to anticipate future needs and allocate resources effectively.

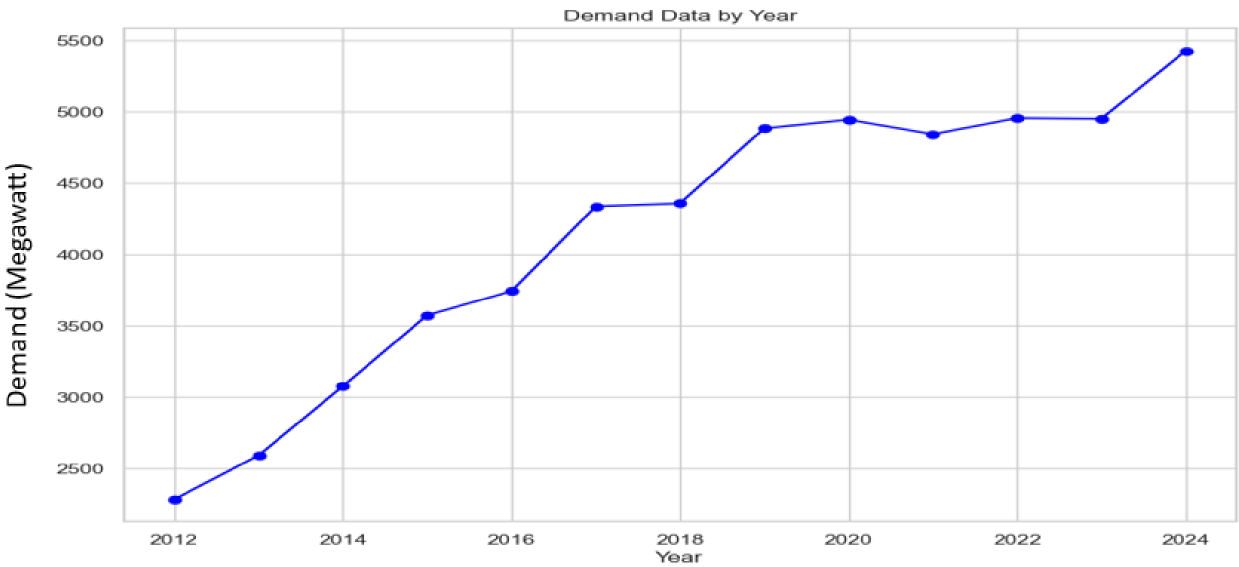

Figure (7) Demand data measured in megawatts over several years

The Figure (8) shows monthly demand data for a particular product or service over 12 months. According to the graph, average demand ranges between 3,000 and 5,000 units, starting from about 3,000 units in the first month, peaking at 5,000 units in a given month, and then gradually declining. It can be seen that demand is witnessing clear fluctuations between months, with peaks and lows

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

indicating seasonal patterns or external factors affecting demand, Average overall demand looks relatively stable, with most data ranging around the range of 3,500 to 4,500 units. Months 2 to 10 show marked variation, with a sudden drop in some months such as month 6 or 8, while month 10 sees a clear rise.

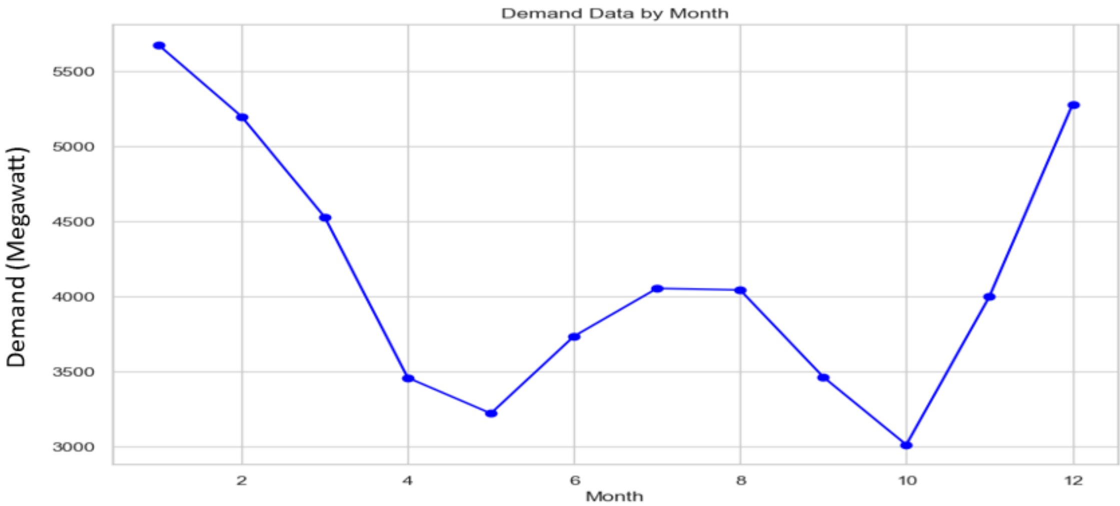

Figure (8) demand data for a particular product or service over 12 months

5. Metrics Analysis

Findings derived from the implementation of the LSTM models are presented along with necessary background to the results. Good experimental design should mean that results stand on their own, but enough information is given to allow others to repeat the experiments. The LSTM models clearly depict the performance metrics used for calculating accuracy and reliability of the forecasting results root mean squared error (RMSE), R-squared (R^2) and Mean Absolute Error (MAE) is defined as the ratio between the sum of the absolute errors and the number of forecasts Forecasting results are compared for other machine learning models run under identical conditions; Predictive capability of the models is depicted clearly with results shown in the form of graphs. All models' predictions for presented along with original monthly consumption highlighting differences between models resulting in deeper insight. Table (5) contrasts two models of forecasting LSTM and RNN, against three performance metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-Square (R^2). These metrics are often used to quantify the accuracy and effectiveness of forecasting models.

Table 5 provides a comparison of the performance of two deep learning models, LSTM and RNN, based on the RMSE, MAE and R-Square evaluation criteria. For the LSTM model, the best results for RMSE and MAE were recorded when using units of 64 and batch size of 64, with RMSE 187.25 and MAE 139.78, indicating high prediction accuracy. As for the model RNN, which performed best when using units with a value of 256 and a batch size of 16, with a value of RMSE of 230.34 and a value of MAE of 232.276. Overall, the LSTM model shows a significant outperformance compared to the RNN model, scoring lower RMSE and MAE values in most cases, making it the best prediction choice in this experiment.

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 5. metric of LSTM AND RNN model

EPOCHS	BACTCH - SIZE	UNITS	LSTM			RNN		
			RMSE	MAE	R-Square	MAE	RMSE	R-Square
100	16	16	194.42	145.91	0.957	293.55	389.73	0.915
100	16	32	188.08	143.56	0.960	282.47	388.66	0.916
100	16	64	196.67	147.39	0.956	267.92	365.88	0.925
100	16	128	196.36	152.95	0.957	232.28	328.02	0.940
100	16	256	188.25	142.79	0.960	230.34	327.42	0.941
100	32	16	314.75	227.41	0.888	343.72	448.27	0.889
100	32	32	189.85	141.57	0.959	300.55	404.26	0.909
100	32	64	197.90	147.68	0.956	291.93	401.83	0.911
100	32	128	188.47	141.75	0.960	265.15	361.55	0.928
100	32	256	236.77	172.14	0.937	236.11	334.03	0.938
100	64	16	193.10	144.12	0.958	340.02	450.35	0.888
100	64	32	189.34	143.36	0.960	310.56	416.59	0.904
100	64	64	187.25	139.78	0.961	294.61	399.39	0.912
100	64	128	199.95	142.21	0.959	277.83	380.37	0.920
100	64	256	199.95	150.05	0.955	283.32	379.17	0.920
100	128	16	312.54	220.67	0.890	437.04	550.41	0.832
100	128	32	271.82	203.74	0.917	366.67	468.09	0.879
100	128	64	191.30	141.88	0.959	392.32	501.65	0.861
100	128	128	198.13	147.76	0.956	324.13	431.10	0.897
100	128	256	244.81	175.01	0.933	308.74	419.63	0.902
100	256	16	317.15	225.93	0.887	549.23	662.26	0.757
100	256	32	236.77	172.14	0.937	421.96	537.92	0.840
100	256	64	278.81	197.91	0.913	462.84	565.41	0.823
100	256	128	242.47	177.19	0.934	391.05	487.71	0.868
100	256	256	193.96	143.03	0.958	399.68	512.31	0.855

5.1. Experimental Results

This table (6) provides a comparison of actual and projected values for a period of time from March 3, 2021 to August 31, 2024. The data shows a discrepancy between the actual figures and the model's forecasts, as it can be seen that the expectations were higher than the actual values most of the time, especially in the last period of August 2024. This table is useful for analyzing the accuracy of a predictive model and identifying gaps between expectations and reality.

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

Table 6. Actual consumption and Predicted consumption in MW

Date	Actual	Prediction
2021-03-03	6128	6051.8701
2021-03-04	6142	6118.2412
2021-03-05	6093	6195.8936
…	….	…
2024-08-29	5140	5235.3926
2024-08-30	4815	5201.9473
2024-08-31	4976	5149.5293

Figure 9 shows the results of the LSTM model's prediction of energy demand, showing the timeline of the actual demand data compared to the model's forecasts during the training predictions and testing predictions. The graphic shows that the model's forecasts follow well the actual patterns of energy demand, with some minor differences, especially in peak or low periods. The graphic covers the period from 2012 to 2024, with energy demand values ranging from 2,000 to 7,000. Model performance indicates LSTM's ability to learn complex patterns in data and provide accurate forecasts, enhancing its effectiveness in energy demand forecasting applications

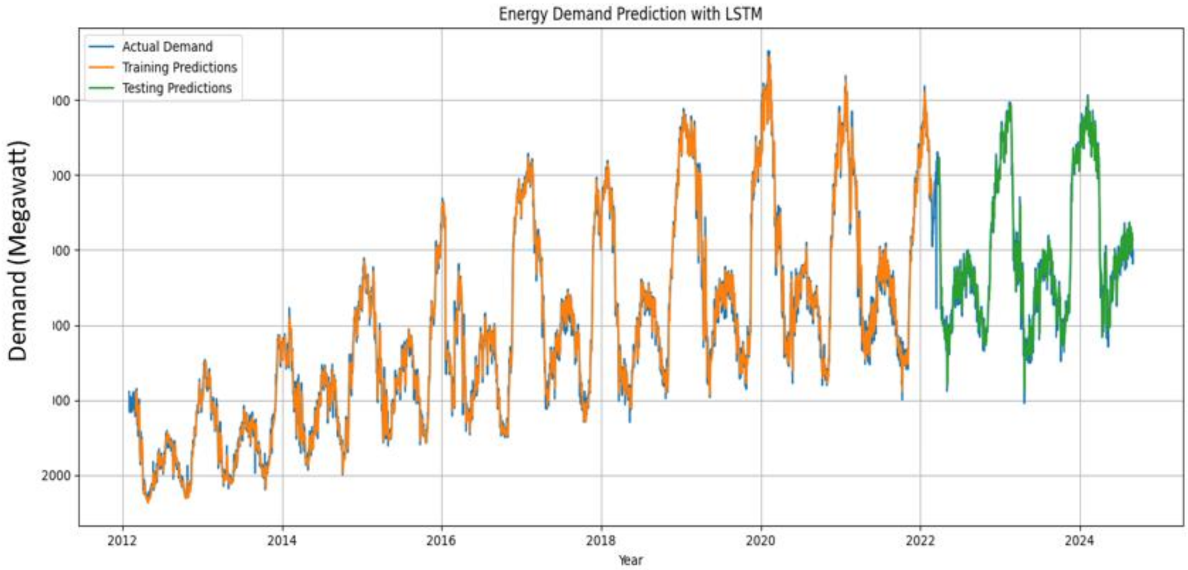

Figure 9. Energy Demand prediction with LSTM

The figure (10) shows a 90-day energy demand forecast from September 1, 2024 to November 29, 2024, based on historical data from February 1, 2012 to August 29, 2024. Demand is expected to increase from 4948 MW to 5645.49 MW during this period, due to the seasonal impact that usually leads to increased consumption during these months. This forecast reflects a recurring pattern of increase in demand during certain periods of the year, which indicates the importance of taking into account Seasonal factors in energy planning.

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

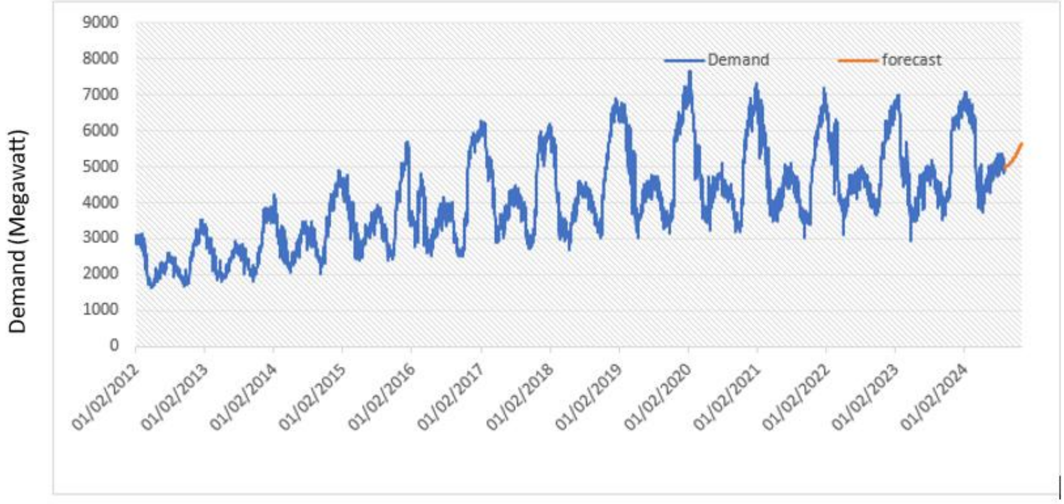

Figure 10. represents forecast energy demand trend over 90 days

6.conclusion

This study explores two different deep learning methods for electricity energy demand for data collected by central electricity control office of KRG from years 2012 to 2014. LSTM shows the most effective performance since it achieves the lowest MAE and RMSE scores with higher R^2. The data exhibited seasonal variations and exogenous factors, rendering temporal behavior analysis essential. The results indicated that the model can learn intricate patterns in data, particularly following the implementation of feature engineering techniques, including the extraction of daily, monthly, and yearly trends. The trials demonstrated that modifying hyperparameters, such as the number of units and batch size, significantly affected model performance. The results of LSTM model show better performance than the RNN model based on the metrics provided, as the RMSE (187.25) and MAE (139.78) values are lower compared to the RNN (RMSE = 230.34, MAE = 232.276), indicating that LSTM forecasts are more accurate with fewer errors. In addition, the R^2 coefficient of the LSTM model (0.961) is higher than that of RNN (0.941).

DOI: http://dx.doi.org/10.25098/9.1.29

Reference

[1] Zhou, Y., Li, X., Wang, J., & Chen, H. Advances in energy consumption forecasting for smart buildings and cities: Methods and applications. Renewable and Sustainable Energy Reviews, 169, 113012, 2023

[2] A. O. Aderibigbe, E. C. Ani, P. E. Ohenhen, N. C. Ohalete, and D. O. Daraojimba. "Enhancing energy efficiency with ai: a review of machine learning models in electricity demand forecasting." *Eng. Sci. Technol. J.* vol. 4, no. 6, pp. 341-356, 2023.

[3] S. F. Ahmed, M. S. B. Alam, M. Hassan, M. R. Rozbu, T. Ishtiak, N. Rafa, M. Mofijur, A. B. M. Shawkat Ali, and A. H. Gandomi. "Deep learning modelling techniques: current progress, applications, advantages, and challenges". *Artif Intell Rev* vol 56, no 11, pp. 13521–13617, 2023.

[4] Zhang, W., Liu, H., Zhang, J., & Wang, Y. A hybrid deep learning model for short-term load forecasting using improved VMD and CNN–BiLSTM with attention mechanism. *Expert Systems with Applications*, *239*, 121837. (2025).

[5] W. Zhao, Y. Gao, T. Ji, X. Wan, F. Ye and G. Bai, "Deep Temporal Convolutional Networks for Short-Term Traffic Flow Forecasting," *IEEE Access*, vol. 7, pp. 114496-114507, 2019.

[6] Utku, A., & Kaya, S. K. Deep learning based on a comprehensive analysis for waste prediction. Operational Research in Engineering Sciences: Theory and Applications, 5(2), 176-189,2022.

[7] Z. A. Khan, et al. "Short-Term Electricity Price Forecasting Through Convolutional Neural Network (CNN). In: Barolli, L., Amato, F., Moscato, F., Enokido, T., Takizawa, M. (eds) Web, Artificial Intelligence and Network Applications. WAINA 2020. Advances in Intelligent Systems and Computing, vol 1150. Springer, Cham,2020.

[8] S. Ghimire, R. C. Deo, D. Casillas-Pérez, and S. Salcedo-Sanz. "Two-step deep learning framework with error compensation technique for short-term, half-hourly electricity price forecasting". *Appl. Energy,* vol. 353, p.122059, 2024.

[9] M. Bilgili, N. Arslan, A. ŞEKERTEKİN, and A. YAŞAR. "Application of long short-term memory (LSTM) neural network based on deep learning for electricity energy consumption forecasting". *Turkish J. of Elect. Eng. Com. Sci.,* vol. 30, no.1, pp.140-157, 2022.

[10] P-H. Kuo, C-J. Huang. "An Electricity Price Forecasting Model by Hybrid Structured Deep Neural Networks". *Sustainability*, vol. 10, no. 4, p. 1280, 2018.

[11] M. Heidarpanah, F. Hooshyaripor, and M. Fazeli. "Daily electricity price forecasting using artificial intelligence models in the Iranian electricity market. *Energy*, vol, 263, p.126011. 2023.

[12] B. Farsi, M. Amayri, N. Bouguila and U. Eicker, "On Short-Term Load Forecasting Using Machine Learning Techniques and a Novel Parallel Deep LSTM-CNN Approach". *IEEE Access*, vol. 9, pp. 31191-31212, 2021,

[13] N. Al Khafaf, M. Jalili, and P. Sokolowski, "Application of Deep Learning Long Short-Term Memory in Energy Demand Forecasting. In: Macintyre, J., Iliadis, L., Maglogiannis, I., Jayne, C. (eds) Engineering Applications of Neural Networks. EANN 2019. Communications in Computer and Information Science, vol, 1000. Springer, Cham. https://doi.org/10.1007/978-3-030-20257-6_3 ,2019.

DOI: http://dx.doi.org/10.25098/9.1.29

The Scientific Journal of Cihan University – Sulaimaniya **PP: 62-81**
Volume (9), Issue (1), June 2025
ISSN 2520-7377 (Online), ISSN 2520-5102 (Print)

[14] M. Azzone and R. Baviera, "Neural Network Middle-Term Probabilistic Forecasting of Daily Power Consumption," J. Energy Mark., vol. 14, no. 1, 2020.
[15] Mahjoub, S., Chrifi-Alaoui, L., Marhic, B., & Delahoche, L. Predicting energy consumption using
LSTM, multi-layer GRU and Drop-GRU neural networks. *Sensors*, *22*(11), 4062, 2022.

[16] B. Ghojogh and A. Ghodsi, "Recurrent Neural Networks and Long Short-Term Memory Networks: Tutorial and Survey," Available at: http://arxiv.org/abs/2304.11461, 2023.
[17] Fatovatikhah, F., Ahmedy, I., & Noor, R. M. (2024). Waste Prediction Approach Using Hybrid Long Short-Term Memory with Support Vector Machine. International Journal of Computational Intelligence Systems, 17(1), 103.
[18] Yazdani, M., Kabirifar, K., & Haghani, M. (2024). Optimizing post-disaster waste collection by a deep learning-enhanced differential evolution approach. Engineering Applications of Artificial Intelligence, 132, 107932.
[19] Siami-Namini, Sima, Neda Tavakoli and Akbar Siami Namin. "The performance of LSTM and BiLSTM in forecasting time series." IEEE International conference on big data (Big Data). IEEE, 2019.
[20] Zheng, Jian, et al. Mar. "Electric Load Forecasting in Smart Grid Using Long-Short-Term-Memory
based Recurrent Neural Network." *IEEE Global Smart Grid Conference (GSS),2017*

DOI: http://dx.doi.org/10.25098/9.1.29