International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning with Multi-Headed Attention for Forecasting Residential Energy Consumption

[1]Dr. Bhalchandra M. Hardas, [2]Shivkant Kaushik, [3]Dr. Ankur Goyal, [4]Dr. Yashwant Dongare, [5]Dr. Mithun G. Aush, [6]Dr. Shwetambari Chiwhane

Submitted: 18/07/2023 **Revised**: 08/09/2023 **Accepted**: 25/09/2023

Abstract: Accurate household energy consumption predictions are crucial for efficient resource allocation and optimal energy management. In recent years, time series forecasting problems have seen encouraging outcomes from deep learning models. However, it is extremely difficult to make precise projections due to the energy consumption patterns' intrinsic complexity and non-linearity. This research suggests a novel method for estimating household energy usage based on deep learning and multi-headed attention to address these issues. To capture the complex temporal correlations and consumption patterns in household energy data, the proposed model makes use of deep neural networks and attention processes that are interpretable. In specifically, the model learns different representations and accurately captures both short-term and long-term relationships by simultaneously attending to many characteristics of the input data via multi-headed attention. The model design combines convolutional and recurrent network neural network layers to extract valuable features from the source time series data and capture changes in time. This study makes a significant contribution to the field of energy forecasting by developing a new model using deep learning with multi-headed attention, producing precise estimates of residential energy consumption, and facilitating effective energy management and allocation of resources in residential settings.

Keywords*: Efficient energy prediction, deep learning, multi headed attention, recurrent neural network*

1. Introduction

The International Energy Agency (IEA) emphasized the predicted annual rise of 1.3% in global energy demand until 2040 in the World Energy Outlook 2019, emphasizing the necessity for increased efficiency initiatives [1]. Residential power consumption is one of the industries that significantly contributes to this demand, making up about 27% of the world's electricity consumption [2]. As a result, the forecasting and analysis of household power consumption become essential for efficient power supply planning, making machine learning technologies, particularly deep learning algorithms, a desirable option [3]. A multivariate time series prediction issue must be solved in order to anticipate residential electricity usage [4]. As shown in Figure 1, sensor-level signals are used to extract distinct properties, such as attributes related to energy use. Using a prediction model, these parameters are then used to forecast power consumption levels [5]. This procedure is crucial for energy management systems (EMS) and smart grid services because it makes it possible to anticipate future power demand using information about past consumption and other factors relating to electricity.

This study attempts to solve the modelling issues related to energy usage in residential settings in order to get around these constraints. A more complicated and accurate prediction model can be created by utilizing cutting-edge deep learning algorithms. Deep learning algorithms are used in the proposed model to capture the complex temporal relationships and non-linear patterns present in residential power use. The model also combines multivariate time series data, including energy consumption variables. The drawbacks of conventional naive time-series models can be avoided by using this method, which will result in better energy management and more effective power supply planning in residential settings. Machine learning techniques are used in the energy sector to forecast future energy demand by utilizing historical customer data.

[1]*Assistant Professor, Department of Electronics and Computer Science, Shri Ramdeobaba college of Engineering and Management, Nagpur, Maharashtra, India*
hardasbm@rknec.edu
[2]*Associate Professor, Department of Computer Science and Engineering, Sat Kabir Institute of Technology and Management, Jhajjar, Haryana, India*
kaushikshivkant@gmail.com
[3]*Associate Professor, Department of Computer Science and Engineering, Symbiosis Institute of Technology, Symbiosis International (Deemed) University, Pune, Maharashtra, India*
ankur_gg5781@yahoo.co.in
[4]*Assistant Professor, Computer Engineering Department, Vishwakarma Institute of Information Technology Pune, Maharashtra, India*
yashwant.dongare@viit.ac.in
[5]*Assistant Professor, Department of Electrical Engineering, Chh. Shahu College of Engineering, Aurangabad, Maharashtra, India*
mithun.csmss@gmail.com
[6]*Assistant Professor, Computer Science and Engineering, Symbiosis Institute of Technology, Pune, Maharashtra, India*
shwetambari.chiwhane@sitpune.edu.in

Deep learning-based power forecasting models have excelled in this situation [10]. But they must overcome two significant obstacles. The active power consumption patterns and other power parameters are multicollinear, to start [11], [12]. Second, the use of electronic devices is the main driver of the transitory and impulsive behaviour displayed by power consumption. The concurrent utilization of power-consuming services can impair the performance of deep learning models, even if convolutional operations have been designed to learn filters that capture local correlations. In real-world circumstances, this method is anticipated to deliver predictions that are more solid and trustworthy, facilitating effective energy management and planning. The local link between electric characteristics and active power in time-series modelling is captured by a unique deep learning model that we offer in this study [13], [14]. It makes use of multi-headed attention.

We illustrate the potency of our suggested strategy through in-depth tests and evaluations. The ability to capture and utilize the key characteristics of the power consumption patterns demonstrates the effectiveness of the multi-headed attention mechanism, enabling precise predictions, including energy peaks. Our findings outperform the effectiveness of current approaches and demonstrate the superiority of the suggested strategy for simulating and forecasting power consumption trends.

2. Review of Literature

In this section, we go over crucial deep learning methods for estimating energy use. We examine both traditional signal processing techniques and more current deep learning-based power prediction studies. The bulk of methods utilized prior to the invention of deep learning concentrated on time series modelling using a symbolic-dynamic framework [15], [16]. These techniques, however, were unable to handle minute changes in the time axis, leading to large discrepancies between time series. Lin et al. created a bag-of-patterns format for feature extraction and incorporated rotation-invariant symbols to address this issue [17]. A fundamental obstacle developed despite the effectiveness of machine learning-based power demand models: these techniques were predominantly assessed for short-term forecasting horizons, omitting medium and long-term projections [7].

Neural networks regularly outperformed other machine learning-based power demand forecasting techniques, such as autoregressive integrated moving average (ARIMA) and decision trees [7][8], in large part because of their non-linear mapping abilities [19]. Previous research mainly concentrated on time series modelling with symbolic-dynamic techniques. However, the development of deep learning methods, notably neural networks, opened the door to the possibility of stronger and more precise predictions of power usage. In addition to outperforming earlier approaches, these models also addressed their shortcomings, notably with regard to medium- and long-term forecasting horizons. Two deep learning models that have excelled at classifying energy patterns and predicting energy use are the LSTM (Long Short-term Memory) and convolution neural networks [10]. In time series data, LSTM models can identify patterns and long-term dependencies due to their ability to learn temporal gating functions. On the other hand, CNN models excel at extracting regional relationships and spatial characteristics from power spectra.

Table 1: Summary of related work in residential energy consumption using deep learning

Paper	Area Domain Data	Prediction Method	MSE Value for prediction
[21]	Household Power at UCI	Conditional RBM That Was Made	0.7211 (1-h)
[22]	Household Power at UCI	LSTM Model	0.5420 (1-h)
[20]	Victoria and New South Wales, Australia	RNN-SVR Group	0.6059 (1-h)
[23]	Australian smart grid and smart city	LSTM, and RNN Method	0.4903 (1-h)
[24]	Building energy applied	ELM, Stacked Autoencoder	4256747 (30-min)
[7]	Competition Time Series Data for M3	k-NN, NN, GP, SVR, and ARIMA	0.4252 (1-h)
[25]	Building for Salt Lake City's Public Safety	LSTM Method	0.2913 (1-h)

Other deep learning models including autoencoders (AE) and generative networks of adversarial networks (GAN), in addition to LSTM and CNN, have become more well-liked in the context of predicting power usage. Deep learning is effective at capturing and reconstructing patterns of power usage, as shown by autoencoders, which are built as a probabilistic technique using CNN and LSTM layers as building blocks [24]. These deep learning models are well suited for tackling challenging energy forecasting problems and energy pattern classification, as evidenced by their extensive acceptance and successes. These models offer promising ways to increase the precision and efficacy of consumption prediction in a variety of energy-related applications by making use of their capacity to learn complicated temporal dependence, extract local correlations, and record complex patterns. Due to the intricate traits displayed by demand time series, anticipating electricity demand is extremely difficult. These traits include, among others, jumps, numerous periodicities, calendar effects, non-constant mean and variance, and high volatility. Researchers have investigated numerous methodologies and optimization methods to address these problems. In order to estimate power usage using a neural network, Mocanu et al. created a stochastic pertaining stage and assessed its effectiveness in various temporal contexts [21]. RBMs (restricted Boltzmann machines) are typical unsupervised learning models that strive to reduce the Kullback-Leibler divergence between layers. When learning the prior distribution from data on power usage, stacked RBMs have proven particularly useful for enhancing prediction accuracy.

The RNN-CNN cells, specifically created for time series modelling, were used to estimate power usage [22], [23]. These research demonstrated the viability and efficiency of boosting prediction performance by utilizing neural networks with memory cells. For forecasting models, researchers have developed optimization techniques to accommodate the varied properties of demand series. For instance, Li et al. pioneered the use of RNNs with pooling procedures for selective gradient updates and the use of autoencoders for obtaining multidimensional features prior to time series modelling [24][30]. Critical time lag factors that significantly affect prediction performance were optimized using genetic algorithms (GA) [31]. Additionally, particle swarm optimization (PSO), which outperforms current deep learning models and is recognized for faster convergence and greater search space exploration compared to GA, has shown potential in optimizing power prediction models [29].

3. Data-Set Used

1. Household Electric Power Consumption:

The amount of active energy spent per minute (in watt-hours) by electrical equipment in the home that is not counted by submeterings 1, 2, or 3 is denoted by the formula (global_active_power * 1000 / 60 - sub_metering_1 - sub_metering_2 - sub_metering_3).

The measurements in the dataset have missing values, which make up about 1.25 percent of the rows. All calendar timestamps are present in the collection, although some of them lack measurement values. The absence of a value between two consecutive semi-colon attribute separators serves as a signal for these missing values. The dataset, for instance, has missing values for April 28, 2007.

Table 2: Dataset Information

Sr. No	Characteristics	Subject Area	Task	Attribute	Record
1	Time Series	Physical	Classification Prediction	9	2075259

4. Proposed Methodology

The CNN-LSTM network with multi-headed focus, which is used to extract spatiotemporal information and predict power consumption based on the power spectrum, is shown in this section's architecture. In order to include them into an end-to-end neural network architecture, we accept and change two crucial elements of the standard power prediction model. Notably, capturing fleeting and impulsive values in electricity demand is a key function of the multi-headed attention mechanism. The recovered features from the CNN are then processed further by the LSTM component, which is renowned for its capacity to capture long-term dependencies. Taking into account the sequential character of time series data, it simulates the temporal dynamics of power use. The model can successfully learn both spatial and temporal patterns thanks to the integration of CNN and LSTM, which boosts the accuracy of power consumption prediction.

Our suggested method combines these components into an end-to-end neural network architecture to extract important spatiotemporal information from the power spectrum and anticipate the transient and impulsive values of electricity consumption. The multi-headed attention mechanism enhances the model's ability to focus on key elements and spot key trends in data on electricity usage.

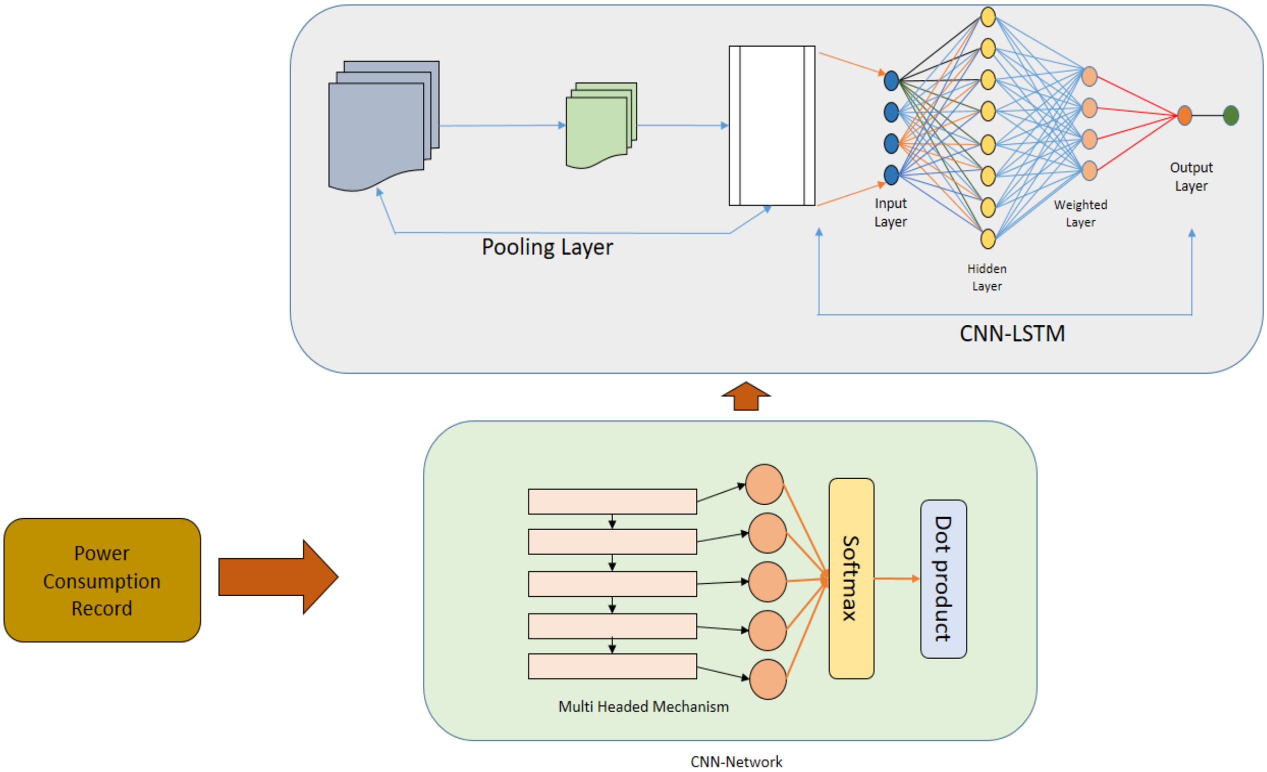

Fig 1: Proposed CNN-LSTM based Model for Energy Consumption Prediction

A. CNN-LSTM Network:

Our suggested approach uses a CNN-LSTM network to estimate electricity demand as its primary goal. The multi-headed attention mechanism, which is created using the function φ that extracts spatiotemporal information and forecasts future energy need, is incorporated to do this. We utilize a direct forecast technique [5] in light of the intricate non-linear mapping stated by stacking numerous levels in the network. This method stays clear of the bias accumulation that recursive sequential forecasting methods frequently use can cause. In the direct forecast tactics, we create a direct model h that, using the input features, directly forecasts the future energy demand. The direct model's associated parameter set is represented by the symbol. We seek to increase the precision and dependability of the forecasting of power demand by utilizing this direct forecast technique.

Our suggested strategy, represented by the function, efficiently captures spatiotemporal characteristics and gives precise forecasts for future energy demand by combining CNN, LSTM, and the multi-headed attentiveness mechanism. The model's learnable weights and biases are captured by the parameter set, allowing it to generate precise predictions according to the input data.

$$\hat{y}t = \varphi h(X\omega t - h;\ \Theta h) + et, h$$

Convolutional recurrent neural networks with trained neurons (CNN-LSTM) are well known for their benefits in learning data-driven filters, notably in extracting spatiotemporal characteristics for signal processing tasks like predicting power consumption [10][6]. An overview of the proposed CNN-LSTM model with multi-headed attention for power consumption prediction is shown in Figure 1. The model consists of two main phases, each of which has a specific function. Convolutional layers are used in the initial stage to separate out spatial information from the input power data. The model can comprehend the spatial linkages and recognize significant features thanks to these layers' application of filters to capture local correlations and patterns within the power spectrum.

Hyperparameters necessary for training the CNN-LSTM model are defined at the data preprocessing stage. Each power attribute is subjected to min-max normalization, and a sliding-window method is used. The time lag, also known as the length of the input signal, and the stride parameter, which controls how many time steps overlap, create the sliding window. The sampling temporal resolutions for the power usage data are 1 to 60 minutes, 1 day, and 1 week as shown in figure 1. The inclusion of LSTM layers in the second stage makes use of their capacity to identify temporal dependencies and enduring patterns in time series data. Taking into account the sequential nature of the data, the LSTM layers process the retrieved spatial information and predict the dynamics of power use over time. The multi-headed attention mechanism is incorporated into the architecture to further improve the performance of the model. The model may concentrate on the most instructive elements of the data, even transitory and impulsive values, because to this attention mechanism's selective weighting and attention to pertinent features.

B. Convolution Neural Recurrent Networks:

The extraction of spatiotemporal information from small consumption samples is one of the key difficulties in modelling power consumption with neural regressor [21]. We use a combination of CNN and LSTM models for feature learning from time-series power consumption data to get around this problem. Taking use of the complementary nature of these two deep learning models in capturing spatiotemporal data, they are successively combined. Convolution and pooling techniques are used in the CNN model to extract spatial data. The ability to identify spatial linkages between various power parameters is made possible by its suitability for collecting local correlations and patterns within the data on power usage.

As opposed to this, the LSTM model is created primarily to identify long-term trends and temporal dependencies in time-series data. The LSTM model can accurately represent the dynamics and temporal relationships within the power consumption sequence by include memory cells. We may take advantage of the advantages of both CNN and LSTM in capturing spatiotemporal features by sequentially merging these two models. While the LSTM component simulates temporal interdependence, the CNN component concentrates on spatial relationships. The intricate patterns in the power usage data may now be fully understood thanks to this integration.

Proposed Deep Learning CNN-LSTM Algorithm:

Step 1: Data preparation

- Apply min-max normalization to the time series data of power consumption, X.
- Create input-output pairs using the sliding-window method.
- Time series data are divided into overlapping windows of length stride.
- The input sequence should be defined as X = x1, x2,..., x, and the corresponding produce y = xn+1.

Step 2: Modeling:

- Convolutional Neural Network (CNN), 2.1
  - Convolutional filters should be applied to the input sequence X.
  - Create feature maps by activating function c for each filter.
  - Use the feature maps' pooling procedure to retrieve pertinent spatial characteristics.
- LSTM stands for Long Short-Term Memory.
  - Pass the feature maps obtained from the CNN to the LSTM model as input.
  - Create the cell states and hidden states for the LSTM.
  - Use the LSTM layers to process the input sequence while recording temporal dependencies.
  - Produce the LSTM model's output sequence h = h1, h2,..., hn+1.
- Multi-headed
  - Attention should be used to recognize the significance of various input sequence components.
  - Calculate attention weights utilizing dot product operations and softmax.
  - Determine the LSTM outputs' weighted sum.

3. Training:

- Set up the model's parameters.
- To determine the difference between the output that was predicted and the actual output, y, use a loss function, L.
- To reduce the loss, modify the model's parameters using an optimization approach (such as stochastic gradient descent).

4. Prediction:

- Pre-process the data using a fresh input sequence, Xnew.
- Pass the trained CNN-LSTM model via the preprocessed input.
- Utilize the multi-headed attention mechanism to calculate the final prediction for the following time step.

5. Evaluation:

- Measure the performance of the CNN-LSTM model using appropriate evaluation metrics

Widely used in signal processing, the convolution operation φc(·) and pooling operation in CNNs are excellent for modelling power consumption sequences and extracting features by capturing local connection between windowed signals. By using filters to identify hidden correlations, the convolution procedure preserves the spatial link among power attributes while lowering translational variation between features [22]. An m 1 sized filter W is used to apply the convolutional operation sequence at the t-th time step. The filter is convolved with the input power attribute sequence R during the convolution operation, with an emphasis on the i-th node in the lst layer and the i-th element of the power attribute sequence. This procedure enables CNN to pick up and record pertinent geographical information.

$$\varphi_{l}^{c(X_t)} = \left[\sum_{\tau=1}^{mX-1} \sum W_{l-1}^{a} \; R_{t-\omega+\tau} \right]$$

Our suggested model uses a 1D convolution-pooling operation to separate out specific properties from the power spectrum in order to extract spatial data. The following LSTM layers receive these spatial characteristics as a sequence of encoded vectors. According to the sliding-window preprocessing, the convolution-pooling function _c catches the time-series data inside a window size. Our model's LSTM component includes gating operations, which are represented by an input gate, forget gate, and output gate (abbreviated as ot). These gates regulate the information flow inside the LSTM cell in an adaptive manner. The LSTM generates an encoded vector L(i) based on the cell state ct and the hidden value ht at each time step t.

$$\varphi L(\cdot) = ht = ot \circ \tanh(ct)$$

where, b stands for the bias term and represents the element-wise product after the CNN-LSTM extracts the spatiotemporal features

$$\varphi \, l \, (Xt) = \sigma \; Wl\varphi \, l - 1 \; L \, \varphi \, l - 2 \; c \; (Xt) + b \, l$$

A linear activation function is used in the last layer of the Multilayer Perceptron (MLP) in the proposed CNN-LSTM regressor. By selecting this activation function, the output scalar value is guaranteed to be directly read as the power prediction.

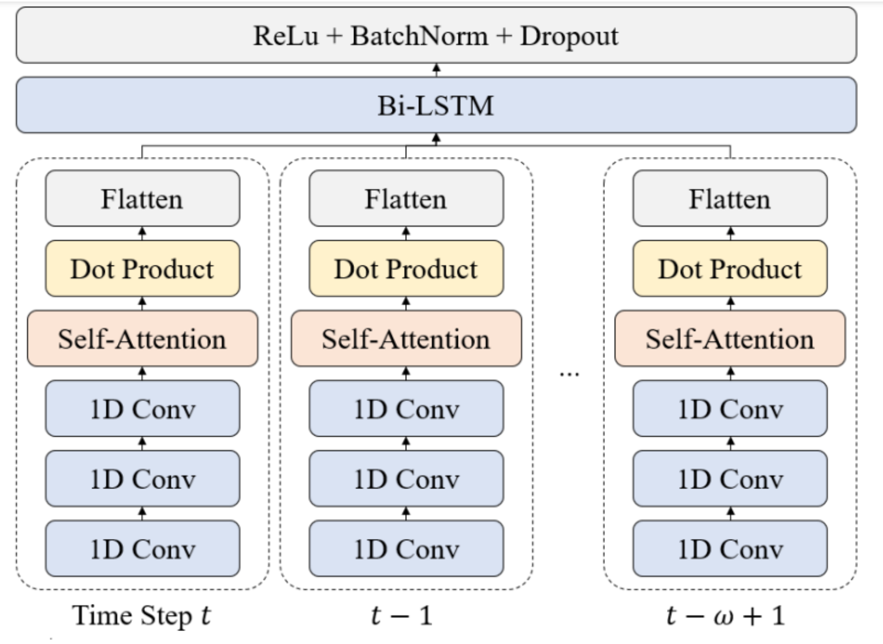

Fig 2: Proposed model CNN-LSTM using multi-headed attention

The backpropagation approach is used in conjunction with gradient descent optimization to update the CNN-LSTM regressor.

$$L_{MSE} = \frac{1}{n} \sum_{j=1}^{n} (Xi - X\hat{}i)^2$$

The objective is to reduce the mean squared error (MSE) metric, which stands for the loss function. The average squared difference between the true values y and the predicted values for a set of n observations is used to calculate the MSE. The CNN-LSTM regressor gains the ability to produce more accurate power forecasts based on the input data by minimizing the MSE loss function using the backpropagation algorithm and gradient descent optimization. To enhance the model's functionality and produce more accurate predictions, the procedure iteratively refines the model's parameters. The gradients of the parameters of the model with respect to the loss function are calculated via the backpropagation algorithm. In order to iteratively reduce the loss and increase the precision of the power estimates, these gradients are then utilized to update the weights of the model and bias through the process of gradient descent.

3. Multi Headed Mechanism:

The attention mechanism determines the alignment score by using a compatibility function f(R_t, q) given a window at time step t, designated as X_t = [R_t,..., R_t-, and the spatiotemporal feature vector representation of a query, indicated as q. The correlation or resemblance between the element R_t and the query q is measured by this compatibility function. A_t = [f(R_, q)] is the alignment score vector.The compatibility function measures the correlations between the query and key items in _=t, which is made up of correlations between those elements.

$$\mathrm{At} = [\, f(\mathrm{R}\tau, \mathrm{Q})]$$

The probability distribution p(z) is used to represent the compatibility function using a softmax operation on the variables (X, Q), and defining the indicator variable z as follows:

$$p(z|\mathrm{Rt}\,, \mathrm{Q}) = \mathrm{softmax}(\mathrm{At})$$

The attention score s can be used as a random variable to describe the relationship between the query Q and the key, which is represented on a scale of [0,1]. The anticipated amount of the use of energy, which is sampled depending on its importance, can be expressed as this attention score.

$$p(z = t|\mathrm{Q}) = \frac{\exp\big(f(\mathrm{Rt}\,, \mathrm{Q})\big)}{t\exp\big(f(\mathrm{R}\tau, \mathrm{Q})\big)}\mathrm{Pt} - \omega\,\tau$$

The energy consumption is dependent on the expectation (E), which indicates the average value. Based on the association between the query and key, the attention score determines the relevance or importance of the energy usage. By taking into account its significance in relation to the supplied inquiry, it enables us to predict the projected energy consumption value.

$$f(\mathrm{Rt}\,, \mathrm{Q}) = \varphi\, l\, \mathrm{cp}(\mathrm{Rt}) \cdot \varphi\, l\, \mathrm{cp}(\mathrm{Q})$$

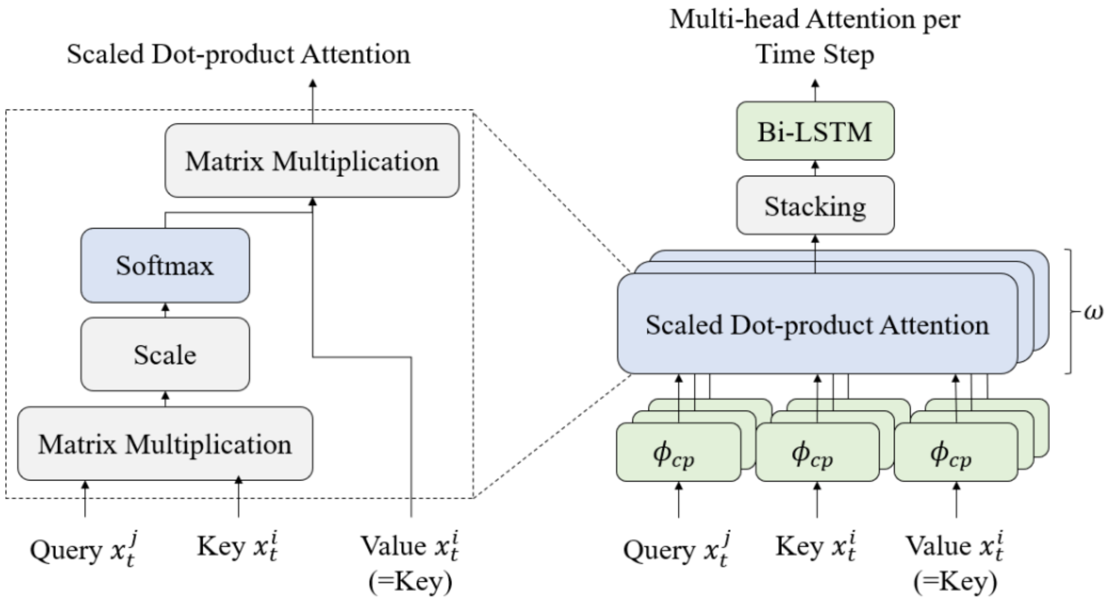

Fig 3: Internal Operation of Multi Headed Mechanism

The deep learning architecture incorporates specialized layers for the attention mechanism. The attentive layer is specifically positioned between the time series and the modelling stages that capture the power spectrum. With the addition of attention layers, the problem of foreseeing abrupt increases in power facility demand, which can be challenging for standard deep learning models, is addressed.

5. Result and Discussion

Convolutional neural networks (CNN) and long short-term memory (LSTM) neural networks were among the machine learning models whose prediction performance results were assessed. A notable improvement in error reduction was shown by our suggested model, which includes a selective modelling strategy for spatiotemporal features. Our suggested model specifically produced a significant error reduction of 22.82% when compared to the traditional CNN-LSTM neural network. This demonstrates the efficacy and superiority of our model in precisely identifying and simulating the intricate spatiotemporal patterns inherent in the data.

Table 3 shows the evaluation findings for our suggested model at various temporal resolutions, from 1 minute to 1 week. Each method's prediction error is represented by its mean squared error (MSE) values, which were obtained by 10-fold cross-validation. We contrast our model with other machine learning approaches for predicting power consumption that have been released in the previous two years. The loss of short-term temporal variables that can affect the modelling of long-term trends may be the reason why the largest prediction errors are shown at time resolutions of 45 minutes and 1 hour. The proposed strategy, however, consistently outperforms the alternative approaches at all temporal resolutions, outperforming both modern machine learning and deep learning techniques.

Table 3: MSE comparison of Different method

Resolution	LR	ARIMA	DT	RF	SVR	CNN	LSTM	Proposed Method
1M	0.0744	0.0738	0.1226	0.0792	0.0798	0.0566	0.0641	0.0616
15M	0.2415	0.2328	0.4434	0.3962	0.3229	0.2023	0.2108	0.1938
30M	0.295	0.2892	0.5399	0.3941	0.362	0.2584	0.2673	0.2466
45M	0.3221	0.3331	0.5004	0.4305	0.4248	0.3083	0.312	0.2938
1H	0.3298	0.3152	0.5459	0.4334	0.406	0.2765	0.2803	0.2762

This indicates how well our method works at capturing both short- and long-term behaviours when estimating power demand. Non-linear mapping techniques like SVR and neural networks reduce mistakes in shorter timescales, whilst conventional methods like ARIMA excel at capturing broad patterns over longer periods. Our suggested technique outperforms existing methodologies in the field of energy forecasting across all temporal resolutions, demonstrating its superiority.

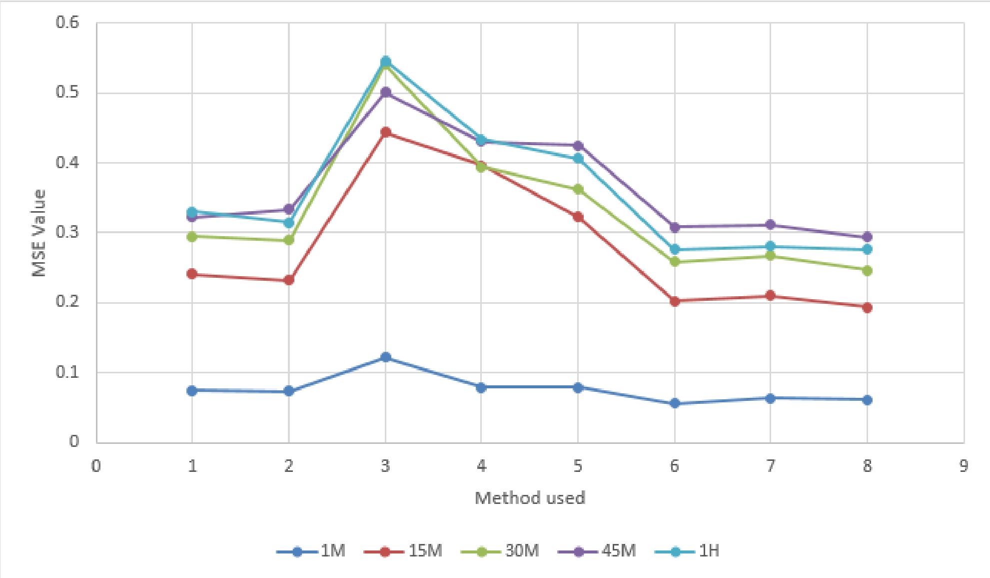

Fig 4: MSE value comparison of different method with proposed method

At various resolutions, the performance of several models' predictions was assessed. Across all resolutions, our suggested approach consistently outperformed competing models such as LR, ARIMA, DT, RF, SVR, CNN, and LSTM. The results show that our approach is superior to previous models in accurately forecasting power usage, with much lower mean squared error values.

Table 4: Different attention processes (MSE) effects on various neural network designs

Attention Type	CNN (1D)	CNN (2D)	LSTM	CNN-LSTM
None	0.0724	0.0656	0.0641	0.0576
Single-attention	0.0801	0.0632	0.0606	0.056
Multi-headed Attention	0.0788	0.0528	0.0583	0.0416

For the CNN (1D), CNN (2D), LSTM, and CNN-LSTM models, the performance of several attention types—including None, Single-attention, and Multi-headed Attention was assessed. The CNN (1D) model had an MSE of 0.0724 for the None attention type, while the CNN (2D), LSTM, and CNN-LSTM models had MSEs of 0.0656, 0.0641, and 0.0576, respectively.

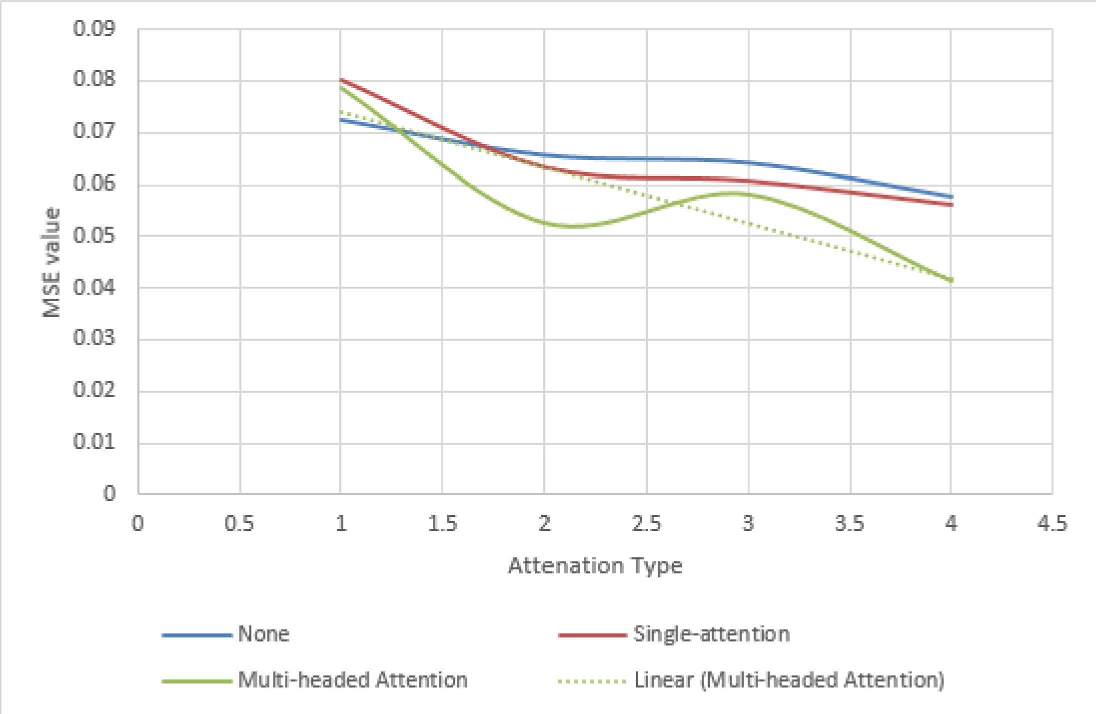

Fig 5: Different attention processes (MSE) effects on various neural network designs

The CNN (1D) model demonstrated an increase in MSE to 0.0801 when utilizing the Single-attention mechanism. The CNN (2D), LSTM, and CNN-LSTM models, with MSE values of 0.0632, 0.0606, and 0.056, respectively, outperformed the others. The performance was further enhanced by the Multi-headed Attention method. The MSE for the CNN (1D) model was 0.0788, while the MSE for the CNN (2D), LSTM, and CNN-LSTM models were even lower at 0.0528, 0.0583, and 0.0416.

Figure 5 shows the results of our comparison of our proposed model's prediction performance to that of a competing CNN-LSTM model. The black line represents the actual power usage figures, while the red line displays the estimated figures. It is clear that the anticipated values and the actual power consumption values closely match, especially when transient and impulsive changes are present. This demonstrates how effective our algorithm is at identifying and forecasting such dynamic patterns in power use data.

6. Conclusion

We established the usefulness of our suggested paradigm by experimental assessments. It performed better across a range of temporal resolutions than deep learning models like CNN and LSTM as well as more conventional machine learning techniques like regression techniques and time-series models. Our model demonstrated its capacity to capture both short-term and long-term patterns, achieving improved prediction accuracy while lowering mean squared error. Multi-headed attention's incorporation was essential for collecting complicated correlations and locating key elements in the power consumption prediction process. The attention mechanism made it possible to localize data probabilistically and extract specific patterns of power use by utilizing the softmax and dot product processes. This assisted in addressing problems with multicollinearity between power consumption patterns and other characteristics, as well as transitory and impulsive behaviour. Our results demonstrate the importance of attention processes in energy forecasting by showing how they can enhance prediction accuracy and get around the drawbacks of traditional deep learning models. The suggested model has potential for use in smart grids, energy management systems, and related fields where precise forecasting is essential.

Additional contextual data can be incorporated, various attention mechanisms can be investigated, and the model's applicability to various domains of energy use can all be explored in future research. Overall, by combining deep learning with multi-headed attention, our study advances the field of energy forecasting.

References:

[1] Marino, D.L.; Amarasinghe, K.; Manic, M. Building energy load forecasting using Deep Neural

Networks. In Proceedings of the IECON 2016—42nd Annual Conference of the IEEE Industrial Electronics Society, Florence, Italy, 23–26 October 2016; pp. 7046–7051.

[2] Nejat, P.; Jomehzadeh, F.; Taheri, M.M.; Gohari, M.; Majid, M.Z.A. A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). Renew. Sustain. Energy Rev. 2015, 43, 843–862.

[3] Zhao, G.; Liu, Z.; He, Y.; Cao, H.; Guo, Y.B. Energy consumption in machining: Classification, prediction, and reduction strategy. Energy 2017, 133, 142–157.

[4] Deb, C.; Zhang, F.; Yang, J.; Lee, S.E.; Shah, K.W. A review on time series forecasting techniques for building energy consumption. Renew. Sustain. Energy Rev. 2017, 74, 902–924.

[5] Arghira, N.; Hawarah, L.; Ploix, S.; Jacomino, M. Prediction of appliances energy use in smart homes. Energy 2012, 48, 128–134.

[6] Prashar, A. Adopting PDCA (Plan-Do-Check-Act) cycle for energy optimization in energy-intensive SMEs. J. Clean. Prod. 2017, 145, 277–293.

[7] Makridakis, S.; Spiliotis, E.; Assimakopoulos, V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS ONE 2018, 13, e0194889.

[8] Gonzalez-Briones, A.; Hernandez, G.; Corchado, J.M.; Omatu, S.; Mohamad, M.S. Machine Learning Models for Electricity Consumption Forecasting: A Review. In Proceedings of the 2019 2nd International Conference on Computer Applications & Information Security, Riyadh, Saudi Arabia, 19–21 March 2019; pp. 1–6.

[9] Burgio, A.; Menniti, D.; Sorrentino, N.; Pinnarelli, A.; Leonowicz, Z. Influence and Impact of Data Averaging and Temporal Resolution on the Assessment of Energetic, Economic and Technical Issues of Hybrid Photovoltaic-Battery Systems. Energies 2020, 13, 354.

[10] Kim, T.-Y.; Cho, S.-B. Predicting residential energy consumption using CNN-LSTM neural networks. Energy 2019, 182, 72–81.

[11] Lago, J.; De Ridder, F.; De Schutter, B. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. Appl. Energy 2018, 221, 386–405.

[12] Fan, H.; MacGill, I.; Sproul, A. Statistical analysis of driving factors of residential energy demand in the greater Sydney region, Australia. Energy Build. 2015, 105, 9–25.

[13] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Polosukhin, I. Attention is all you need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.

[14] Zheng, H.; Fu, J.; Mei, T.; Luo, J. Learning Multi-attention Convolutional Neural Network for Fine-Grained Image Recognition. In Proceedings of the 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5219–5227.

[15] Ray, A. Symbolic dynamic analysis of complex systems for anomaly detection. Signal Process. 2004, 84, 1115–1130.

[16] Rajagopalan, V.; Ray, A. Symbolic time series analysis via wavelet-based partitioning. Signal Process. 2006, 86, 3309–3320.

[17] Lin, J.; Khade, R.; Li, Y. Rotation-invariant similarity in time series using bag-of-patterns representation. J. Intell. Inf. Syst. 2012, 39, 287–315.

[18] Tso, G.K.; Yau, K.K.; Tso, G.; Yau, K.K.W. Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. Energy 2007, 32, 1761–1768.

[19] Ekonomou, L. Greek long-term energy consumption prediction using artificial neural networks. Energy 2010, 35, 512–517.

[20] Li, W.; Yang, X.; Li, H.; Su, L. Hybrid Forecasting Approach Based on GRNN Neural Network and SVR Machine for Electricity Demand Forecasting. Energies 2017, 10, 44.

[21] Mocanu, E.; Nguyen, P.H.; Gibescu, M.; Kling,W.L. Deep learning for estimating building energy consumption. Sustain. Energy Grids Netw. 2016, 6, 91–99.

[22] Wang, Y.; Gan, D.; Sun, M.; Zhang, N.; Lu, Z.-X.; Kang, C. Probabilistic individual load forecasting using pinball loss guided LSTM. Appl. Energy 2019, 235, 10–20.

[23] Kim, T.-Y.; Cho, S.-B. Particle Swarm Optimization-based CNN-LSTM Networks for Forecasting Energy Consumption. In Proceedings of the 2019 IEEE Congress on Evolutionary Computation, Wellington, New Zealand, 10–13 June 2019; pp. 1510–1516

[24] Shi, H.; Xu, M.; Li, R. Deep Learning for Household Load Forecasting—A Novel Pooling Deep RNN. IEEE Trans. Smart Grid 2017, 9, 5271–5280.

[25] Bouktif, S.; Fiaz, A.; Ouni, A.; Serhani, M.A. Optimal Deep Learning LSTM Model for Electric Load Forecasting using Feature Selection and Genetic Algorithm: Comparison with Machine Learning Approaches. Energies 2018, 11, 1636.

[26] Guo, Z.; Zhou, K.; Zhang, X.; Yang, S. A deep learning model for short-term power load and probability density forecasting. Energy 2018, 160, 1186–1200.

[27] Fan, C.; Sun, Y.; Zhao, Y.; Song, M.; Wang, J. Deep learning-based feature engineering methods for improved building energy prediction. Appl. Energy 2019, 240, 35–45.

[28] Hinton, G.E. Deep belief networks. Scholarpedia 2009, 4, 5947.

[29] Taieb, S.B. Machine Learning Strategies for Multi-Step ahead Time Series Forecasting; Universit Libre de Bruxelles: Bruxelles, Belgium, 2014; pp. 75–86. 36. Wang, K.; Qi, X.; Liu, H. Photovoltaic power forecasting based LSTM-Convolutional Network. Energy 2019, 189, 116225.

[30] Bu, S.-J.; Cho, S.-B. A convolutional neural-based learning classifier system for detecting database intrusion via insider attack. Inf. Sci. 2020, 512, 123–136.

[31] Qing, X.; Niu, Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. Energy 2018, 148, 461–468.

[32] Chorowski, J.; Bahdanau, D.; Serdyuk, D.; Cho, K.; Bengio, Y. Attention-Based Models for Speech Recognition. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 7–12 December 2015; pp. 577–585.

[33] Patil, R. A. ., & Patil, P. D. . (2023). Skewed Evolving Data Streams Classification with Actionable Knowledge Extraction using Data Approximation and Adaptive Classification Framework. International Journal on Recent and Innovation Trends in Computing and Communication, 11(1), 38–52. https://doi.org/10.17762/ijritcc.v11i1.5985

[34] Brown, R., Brown, J., Rodriguez, C., Garcia, J., & Herrera, J. Predictive Analytics for Effective Resource Allocation in Engineering Education. Kuwait Journal of Machine Learning, 1(1). Retrieved from http://kuwaitjournals.com/index.php/kjml/article/view/91

[35] Sharma, R., Dhabliya, D. A review of automatic irrigation system through IoT (2019) International Journal of Control and Automation, 12 (6 Special Issue), pp. 24-29.