

Deep Learning for Energy Forecasting Using Gated Recurrent Units and Long Short-Term Memory

E. T. Sivadasan^{1, 2,*}, N. Mohana Sundaram¹, R. Santhosh¹

¹Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India

²Vidya Academy of Science and Technology, Thrissur, Kerala, India

Emails: santhoshrd@gmail.com; itismemohan@gmail.com; etsivadasan@gmail.com

Abstract

Forecasting energy demand is essential for efficient grid management as it promotes steady operations, efficient markets, and sustainable energy practices. In this study, previously observed, evenly spaced energy consumption data are analysed using recurrent neural networks based on Gated Recurrent Unit (GRU) and Long Short-Term Memory (LSTM) architectures to extract important insights, features, and remarkable patterns. First, the study examines the influence of meteorological features on energy consumption. The most significant meteorological features are determined by computing the MIC and Pearson's correlation coefficient. The selected features are then combined with historical energy consumption data to feed the neural network. Second, to improve and optimise the performance of the proposed models, two technical indicators - the daily energy usage average and the simple moving average - are considered. The following are some instances of comparisons in terms of prediction accuracy: (1) The MAPE of the proposed model is 2.47, whereas that of the current model is 4.03. (2) The MAPE of the existing model is 25.83, whereas the proposed solution is 18.68. (3) The MAPE of the suggested model is 24.8, while the MAPE of the current model is 26.6. (4) The MAPE of the present model is 4.77, whereas the suggested approach's is 4.42.

Received: February 07, 2024 Revised: April 20, 2024 Accepted: July 04, 2024

Keywords: Time Series; Energy Forecasting; MIC; GRU; LSTM

1. Introduction

In the present economic scenario, time series data prediction plays a critical role in risk management, resource allocation, strategic planning, and informed decision-making. Advancements in technology provide several opportunities and promise to capture useful insights about the future from the past time series data. By examining historical data and calculating new, unknown future values, time-series forecasting seeks to provide valuable information and future insights. Predictive and statistical methods got a significant role in preparing and analysing the data [1]. In order to anticipate energy consumption, stock market movements, weather, aerospace, and defence, among other real-time applications, researchers use sophisticated data mining techniques, deep learning combined with recurrent neural networks. This study's goal is to look at energy usage data to better accurately forecast future power demand by employing Recurrent Neural Networks (RNN). Exploration of power consumption data for forecasting future power demand is vital in ensuring that, electricity load and price estimates at the corporate level have become critical inputs to the decision-making systems of energy firms. Over or under contracting, followed by selling or purchasing power in the balancing market, can often result in significant financial losses and in the worst-case scenario insolvency. Electric utilities are the most vulnerable here since they cannot pass on their expenses to retail customers frequently. It is necessary to have long-term (months ahead) projections for power system planning, mid-term (weeks ahead) forecasts for fuel supply requirements and maintenance, and short-term (hours ahead) forecasts for daily energy system operations, according to [2]. Applying the best models for projecting future trends based on past observed patterns is the goal of the well-researched discipline of time series forecasting. The time series forecasting paradigm includes the forecasting of electricity loads [3]. In today's technological era, an accurate load forecast is critical for

implementing the notion of smart grids and smart buildings. In essence, effective energy forecasting enables stakeholders in the energy sector to make intelligent choices, optimise operations, save costs, and contribute to an energy system that is more reliable and sustainable. One of the biggest challenges facing the power industry is developing high-accuracy forecasting models for electrical load demand. Over the years, many machine learning and statistical methods have been suggested and used for effective power load forecasting [4]. The impact of temperature and other meteorological factors on load demand is substantial. To improve prediction accuracy, all factors impacting load demand such as past loads and influential meteorological data should be considered as inputs to the forecast model. By employing technical indicators like the simple moving average and the day-ahead 24-hour energy consumption average, this work also aims to increase the day-ahead load forecasting model's accuracy and generalizability. To assess how effective the models are, common error calculations, including (a) root mean square error (RMSE), (b) mean absolute error (MAE), and (c) mean absolute percentage error (MAPE), are utilised as per [5]. One of the study's two primary contributions assesses forecast accuracy; examples of comparisons of prediction accuracy include the following: (1) the proposed model's MAPE is 2.47, whereas the existing model's is 4.03. (2) The proposed model has a MAPE of 18.68, while the current model's is 25.83. (3) The proposed model has a MAPE of 24.8, while the present model's is 26.6. (4) The proposed model has a MAPE of 4.42, compared to the existing model's 4.77. The identification of the most relevant meteorological data as well as the selection of technical indicators from a broad range of possibilities, including the arithmetic mean (AM) and simple moving average (SMA), that impact energy demand projections, constitute another noteworthy addition to this study.

The following is the structure of the remaining portions of the paper: Section 2 provide literature review; Section 3 covers methodology; Section 4 presents proposed models; Section 5 contains results and discussion; Section 6 compares the performance of the suggested models with the current models; and Section 7 provides an overview of the paper's conclusion and future research directions.

2. Literature review

Time series forecasting is essential in many technical and scientific domains, such as finance, energy, ecology, and the environment. A set of scalar or vector observations collected throughout time is called a time series [6]. A model-building process that establishes a link between the input and output data may be used to conceptualise time series prediction. Following the development of the model, future values may be predicted using both historical and current data [7]. A modern power plant has to have an uninterrupted supply of electricity available for the load system. This calls for a precise and error-free estimate of the load needs both now and in the future. Researchers and scientists have attempted to establish an optional but extremely successful technique to achieve this goal: energy forecasting, which estimates the demand for future power use. Energy forecasting directs a number of choices, including how to allocate fuel, dispatch, off-line network, production, and other operations in response to changes in consumption [8].

One technique for managing supply and demand is load forecasting. In this extremely challenging project, it is crucial to look at several influencing direct and indirect components. Weather-related factors have an impact on energy demand forecasts. In load forecasting, complex, stochastic algorithms are employed. Therefore, in order to calculate the load for the current hour, the following factors are considered: demographic information, weather, the total number of devices in the forecasting region, economic data, the quantity and type of customers, the load from the previous hour, the load from the previous day, the load from the previous week at the exact same hour on the same day with the same denomination, and so on as per [9]. Short-term forecasting is a method for estimating energy from only a few minutes to a couple of days in advance. It is crucial to a number of grid activities that call for dispatch and reliability evaluations [10]. Additionally, it makes energy usage less likely to be underestimated or overestimated, which significantly increases grid reliability [11]. In a year, medium-term nature forecasting is used for periods of time ranging from just a couple of days to many months in advance [12]. It helps with fuel distribution, adequacy evaluation, and smart grid system maintenance. Additionally, by assessing the energy system's financial characteristics, it aids in risk management as in [13]. Long-term forecasts cover time spans that vary from months to years. Extended projections are necessary for long-term planning of both output and load increases as per [14].

Load prediction is one of the many significant sequence-to-sequence transitions involving sequential inputs and corresponding sequential results that occur in real time. RNN can recycle the responses of previous or subsequent neurons in an arrangement to better predict the next sequence. Networks based on LSTM are a sort of a RNN that stores and outputs data using a specific type of memory cell. This type of memory cell, as described in [15], is made to retain information for a considerable amount of time by the use of a number of "gates" that control the flow of information into and out of the cell. The gate mechanisms in the LSTM network are activated by sigmoid functions of activation, which produce values ranging from 0 to 1. The gates enable the network's cells to selectively remember or delete information based on the current values associated with the input and the earlier state of the cell. GRUs, on the other hand, are a more basic type of LSTM. They use one "update gate" to control the flow of data into the memory cell, as opposed to the three gates seen in LSTMs. As a result, GRUs are simpler to train and run than LSTMs, but their ability to manage and retrieve highly long-lasting dependencies may be limited [16]. LSTMs and GRUs are two kinds of recurrent neural networks employed in a variety of applications such as time series data processing and processing data in sequence such as text,

voice, or video. There is no best instance of an RNN for all jobs, and the decision between LSTMs and GRUs will be determined by the task's specific needs. The gradient descent optimisation approach is used to train LSTM and GRU network models. These models learn from the input training data, and the loss function assesses how precise the forecasting ability is for each successive iteration when parameters are adjusted. Developing a model and then iteratively adjusting the parameters to reduce the prediction error are the goals of training. There is a forward move and a backward move in the gradient descent method. Forward propagation occurs when input data flows through the network according to an algorithm that determines the value of each neuron in the layer below it. This algorithm iterates until an output or forecast is obtained. We then compute the difference between the target variable and forecast variables as described by a loss function, such as RMSE [17, 18]. Simply put, LSTM and GRU designs are immune to the optimisation constraints imposed by disappearing or bursting slopes, which are beyond the capabilities of SRNs. Using a modified window slide approach with a data overlay, the subsequent values of the time series data are extracted from the preceding sequence. To forecast the 24-hour energy demand for the next day, a window of a particular size will be swept over the daily 24-hour historical usage of energy data. It is possible to train models that connect the 24-hour energy consumption of one day to that of the following day. The recommended models are built on top of deep learning techniques. Strong algorithms and remarkable effectiveness characterise recurrent neural networks, which are used to develop prediction models by identifying patterns in data [19]. Below is an explanation of the methodology employed.

3. Methodology

To match the scenario with the existing historical data, we have to make certain guesses about the future. We are taking this measure in the hopes that the pattern observed in past information will be maintained in the event of future unforeseen exogenous events [20]. Based on these assumptions, we develop several types of computational frameworks for irregular regression analysis using GRU and LSTM architectures. Time-series forecasting uses patterns found in historical data to predict future events. Hourly peak load historical data is gathered day-by-day from various geographic regions and characterised as discrete numerical data. Predicting whether the previous trend of the time series data will continue is done using a quantitative technique. For some data, meteorological features are available. In such cases, meteorological features are combined with the energy consumption data as their input to the models for more precisely predicting the hourly peak loads for the following day.

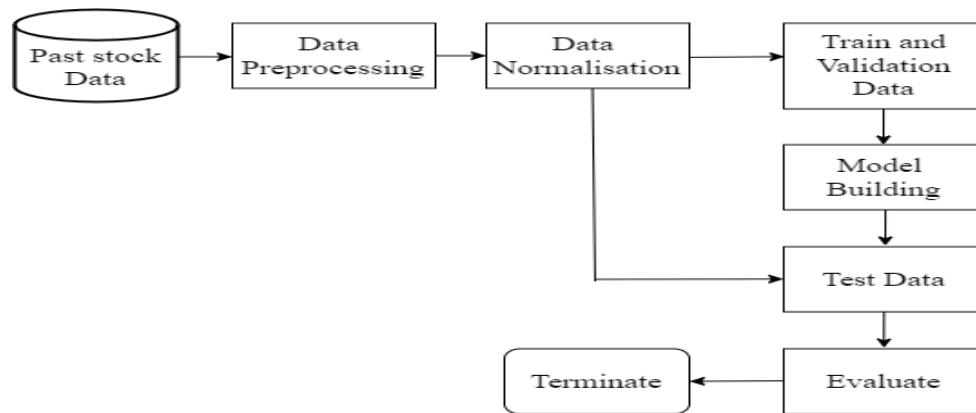


Figure 1. Flow Diagram

To quantify linear correlations, the coefficient of Pearson's correlation, as explained in [21, 22], is calculated in order to eliminate redundant input data and choose features from a range of meteorological components that are strongly linked with energy consumption. As in [1, 23], the Maximal Information Coefficient (MIC) is employed. The maximum information coefficient (MIC) is a helpful tool for determining and evaluating the strength of associations between variables when linear correlations are unable to sufficiently represent the diverse nature of the data. As a consequence, fewer parameters are required in total for model training. Figure 1 provides a detailed flow diagram of the decision flows. The analysis comes next in the process.

3.1. LSTM-based modelling for energy forecasts

With LSTMs, the optimisation restrictions of bursting and disappearing gradients may be circumvented as per [24]. Earlier but long-prevailing information can survive through a sequential network called an LSTM, which remembers prior data and makes use of it in conjunction with the present input to carry out future forecasting objectives. The LSTM's short-term memory is capable of preserving thousands of time steps. The LSTM can remember long-term dependence because it possesses memory that is both short- and long-term. Therefore, the LSTM is a better choice for prediction when using time-series information with dependencies that extend over time. The LSTM architecture is widely used in two fields: deep learning and artificial neural networks. In addition

to many gating strategies, the LSTM features a cell phase. The secret is in its cell comprehension, a clever gate-controlled knowledge transport channel that transmits pertinent data. The forget gate layer facilitates the process of locating and eliminating irrelevant data from the cell state. Together, a *tanh* layer and an input gate layer generate new candidate values and choose which fresh data will be used to update the cell. The LSTM's output is accessible via its output layer. The several steps of the processes listed in Equations (1) through (6) are explained in depth as provided in [25, 26].

$$f_t = \sigma(W[f].[h_{t-1}, x_t] + b[f]) \quad (1) \quad (1)$$

$$i_t = \sigma(W[i].[h_{t-1}, x_t] + b[i]) \quad (2) \quad (2)$$

$$C_t = \tanh(W[c].[h_{t-1}, x_t] + b[c]) \quad (3) \quad (3)$$

$$C_t = f_t * C_{t-1} + i_t * C_t \quad (4) \quad (4)$$

$$o_t = \sigma(W[o].[h_{t-1}, x_t] + b[o]) \quad (5) \quad (5)$$

$$h_t = o_t * \tanh(C_t) \quad (6) \quad (6)$$

Where: f Forget gate, W Weight, h Output of LSTM, x Input, b Bias, i Input gate, C Cell state, C_t new candidate cell scenario, h Output of LSTM, o Output gate, \tanh Nonlinear activation function, t Time, σ Sigmoid activation function.

3.2. Modelling energy forecasts using GRU

The GRU, a simplified version of the LSTM, was developed by Cho et al. (2014). The GRU algorithm merges the input and forget gates into a single update gate. By integrating the hidden state with the cell state, it makes the system simpler. The update gate must keep in reserve the quantity of prior data that must be carried over to the next state. This is a very prudent decision since the model may consider relevant information from the past and lessen the likelihood that the gradient would vanish as supplied. In equations (7) through (10) the explanations for several process phases given as reported in [27].

$$z_t = \sigma(W[z].[h_{t-1}, x_t]) \quad (7)$$

$$r_t = \sigma(W[r].[h_{t-1}, x_t]) \quad (8)$$

$$\tilde{h}_t = \tanh(W[h][r_t * h_{t-1}, x_t]) \quad (9)$$

$$h_t = (1 - z_t) * h_{t-1} + (z_t * \tilde{h}_t) \quad (10)$$

Where: z Block Input, r Reset gate, \tilde{h} Current state, z Update gate.

4. Proposed models

The suggested models were developed leveraging GRU and LSTM architectures using historical data in order to predict the energy consumption for the next 24 hours. Input characteristics and the selection of suitable weather data those affect model performance, and the best outcomes depend on selecting the proper weather data as in [2]. Different performance metrics including (i) MAPE, (ii) MEA, and (iii) RMSE are employed to assess the models' performance. Models are created to find a deeper pattern and generalise previous data as in [28]. The finest neural network forecasting model has to go through a number of processes utilising historical time series data. Pre-processing, normalisation, and purification of data are necessary before it is supplied to the network. Following normalisation, the collected data needs to be divided into three groups: training, validation, and testing. Models are constructed using the training data set, and their hyper-parameters are subsequently adjusted using the validation data set. The hyper-parameters of a neural network model include hidden layer count, neurons per layer, activation functions, training procedures, optimisation techniques, learning rate, etc. Last but not least, the model's effectiveness is evaluated using a test data set. Two categories of the suggested single-layer networks can be distinguished based on the number of output nodes. Those with many output nodes that predict a range of hourly loads fall into the first group. In other words, they can anticipate the 24 hourly loads for the next day using their 24 nodes. Those with a single output node are in the second category; their load is estimated for the upcoming hour. Figure 2 provides an illustration of the models.

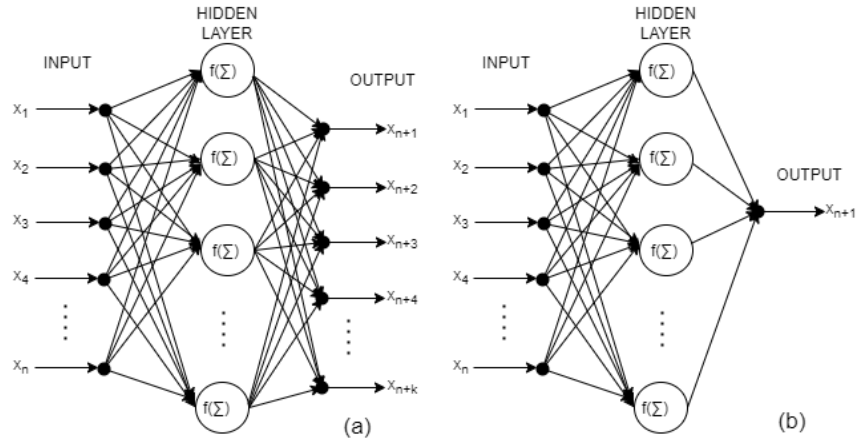


Figure 2. Neural network (a) multiple output, (b) single output

4.1. Dataset

The following two datasets were used: (i) the publicly available dataset from the 2016 Electrical Engineering Mathematical Modelling Competition, as produced in [2, 29]. These collection includes information on the amount of electricity consumed between January 1, 2009 and December 31, 2014, as well as information on temperature, humidity, and rainfall between January 1, 2012 and December 31, 2014. (ii) The state grid Chongqing Electric Power Company in China is the source of the real-world power load data collection [2, 29]. The data collection includes the amount of power consumed between January 1 and December 31, 2017. Data is standardised, null values are eliminated, and pre-processing is done to increase the predictability of the models.

4.2. Methodical preparation of the data and feature selection

An essential step in the deep learning pipeline is data preparation. Outliers, noise, missing values, and consistency problems are common with raw data. As a result, data preparation is very important since the functionality and efficacy of deep learning algorithms are directly impacted by the quality and condition of the data. Data preparation is essential in deep learning for a variety of reasons including data quality and consistency, normalisation and scaling, feature engineering, handling missing data, dimensionality reduction, reducing processing needs, benchmarking, and repeatable findings. Some of the data processing methods employed in this study are listed below.

4.2.1. Feature selection

In the historical dataset that was used to create the model, hourly energy use and meteorological data were examined. According to the research, meteorological properties influence how much energy will be used in the future. Using the Pearson's correlation coefficient for linear relationships of load with meteorological factors and MIC for non-linear correlations, the relevant meteorological elements for the dataset are selected.

4.2.2. Pearson's correlation coefficient

The pattern and magnitude of the linear connection among two variables that are continuous are determined by a measure of statistics called Pearson's correlation coefficient, or simply "r." It is in the range of -1 and 1. When two variables have a perfect positive linear connection - that is, when one rises, the other rises predictably - and the correlation coefficient is 1, a correlation coefficient of -1 indicates a complete negative linear link between two variables, meaning that as one rises, the other falls off predictably. In the absence of a linear connection between the variables, a correlation coefficient of 0 is used. The characteristics having the highest correlation to the target are chosen in this procedure to create the model. The chosen characteristics also shouldn't be substantially associated with one another in order to prevent duplicate processing [21, 22]. The following is a description of the Pearson correlation methodology, which is one of the most helpful approaches for this:

$$r_{x,y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (11)$$

Where the two variables' individual data points are denoted by x_i and y_i . The means, or averages, of the two variables are \bar{x} and \bar{y} , respectively. The Pearson correlation coefficient, or $r_{x,y}$, is a statistical measure that may be used to assess the pattern and magnitude of linear correlations among variables.

4.2.3. Maximal information coefficient (MIC)

A measure used to assess the degree of statistical dependence between two variables is the Maximal Information Coefficient (MIC). It was created as a method of identifying and measuring nonlinear correlations between variables that may not have been picked up by more conventional correlation measurements like the Pearson's correlation coefficient. In 2011, Reshef et al. published a work titled "Detecting Novel Associations in Large Data Sets" that introduced the MIC [1, 23].

$$\text{Let } X = \{x_i, i = 1, 2, 3, \dots, n\}, Y = \{y_i, i = 1, 2, 3, \dots, n\}$$

be the two different factors. $I(X, Y)$ is the mutual knowledge between X and Y. It is calculated as follow:

$$I(X, Y) = \sum_{X \in x} \sum_{Y \in y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (12)$$

Here, the joint probability density function of variables X and Y is represented as $p(x,y)$, while the marginal probability density functions of variables X and Y are indicated by $p(x)$ and $p(y)$, respectively, as provided in [23].

Assuming that there is a finite ordered pair set $D\{(x_i, y_i), i = 1, 2, \dots, n\}$, the grid of $x * y$ is created by the scatter diagram made up of x_i and y_i in the set D . Next, each grid's mutual information is separately determined. Since there are several ways to partition the $x * y$ grid, the mutual information value of the grid is selected as the maximum value in a variety of ways and is recorded as $MI(X, Y)$. The maximal Mutual Information $MI(X, Y)$ is then normalised to get the MIC, as provided in [23].

$$MIC(X, Y) = \max_{|X||Y| < T} \frac{MI(X, Y)}{\log(\min|X|, |Y|)}, \quad (13)$$

Where: T, which rises with the amount of samples in the data set, is the upper limit of $x * y$ for the grid partition. Setting the value of T to $n^{0.6}$, where n is the number of samples as determined in [23], yields the best results.

4.5. Model architecture

Figure 2 displays the recommended deep learning network models that make use of the GRU and LSTM frameworks. Both models for predicting energy consumption for the upcoming 24 hours and for predicting energy consumption for the upcoming hour are created. The results show that the proposed GRU and LSTM models are superior to the current models in terms of prediction accuracy. GRU networks outperform LSTM networks in terms of prediction accuracy among the suggested models.

Setting hyperparameters evenly across models allows for an impartial comparison. The settings of the models are (i) number of hidden layers one, (ii) 24 neurons per layer, (iii) Adam optimiser, and (iv) 50 epochs. The models were created with Python 3.10 with TensorFlow, Keras, NumPy, Pandas, Sklearn, and other libraries.

5. Results and discussions

5.1. Phase-1: Day ahead 24-hour energy demand forecast

Publicly available data from the 2016 Electrical Engineering Mathematical Modelling Competition is used in this work, as in [29]. In the first phase models are created to precisely predict the 24-hour peak demand of energy consumption for the following day using historical data between 1 January 2008 and 31 December 2014. The impact of climatic variables on energy consumption is examined, and in order to improve forecast accuracy, datasets including these variables—such as the highest and lowest temperatures as well as the average weekly temperature—are employed. Figures 3 and 4 illustrate measurements of the Pearson's correlation coefficient and MIC for load and meteorological data, respectively. Figures 5 and 6 display the performance graphs of the recommended GRU and LSTM models, respectively. Table 1 displays several error estimations of the models' while forecasting the next-day, 24-hour power requirements.

Table 1: Day ahead 24 hours forecast error

MODEL	MAPE	MAE	RMSE(MW)
GRU	2.473	196.341	303.534
LSTM	2.563	208.270	349.988

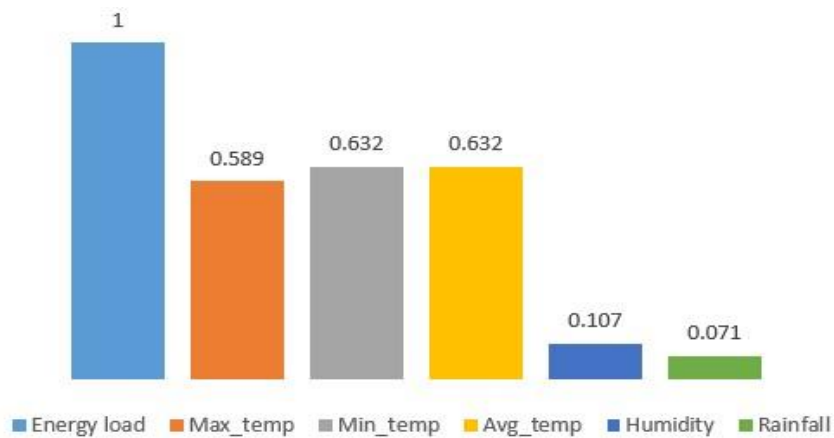


Figure 3. Pearson's correlation coefficient public data set

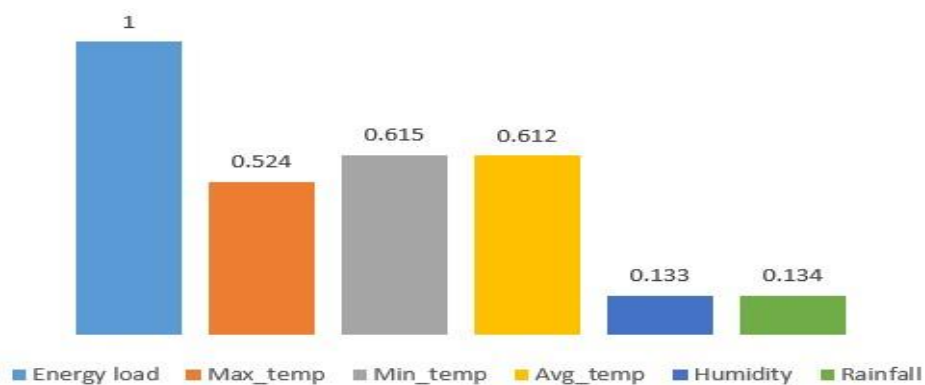


Figure 4. MIC - load & meteorological factors public data set

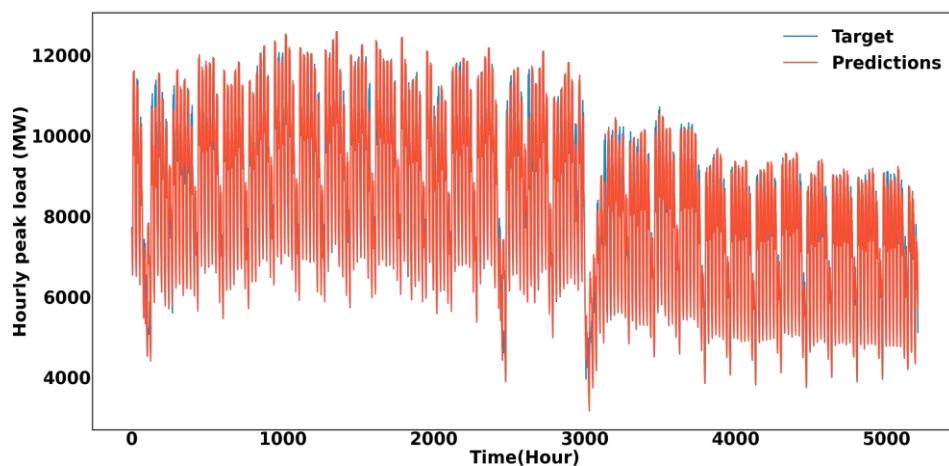


Figure 5. GRU: Day-ahead 24 hours prediction public dataset

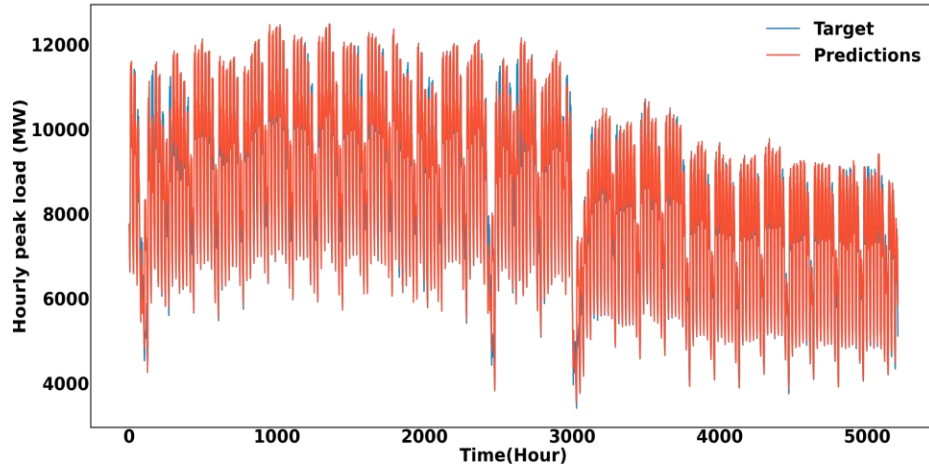


Figure 6. LSTM: Day-ahead 24 hours prediction public dataset

For a closer look, day-ahead 24-hour load forecast for a models are taken and shown in Fig. 7. The results show that the predictions made by the suggested models are very close to the actual values.



Figure 7. Day-ahead 24 hours prediction on 3rd June, 2014 public dataset

5.2. Phase-2: Predicting day-by-day peak energy consumption

The second phase involves developing models for predicting day-by-day peak energy consumption. So that power grid operators can get ready to meet the predicted peak load demand for the day. Table 2 displays the suggested models' performance metrics. Figures 8 and 9 provide the performance graphs for the GRU and LSTM structures of networks, respectively.

Table 2: Day ahead peak hour forecast error

MODEL	MAPE	MAE	RMSE(MW)
GRU	2.199	208.435	335.046
LSTM	2.777	271.973	417.584

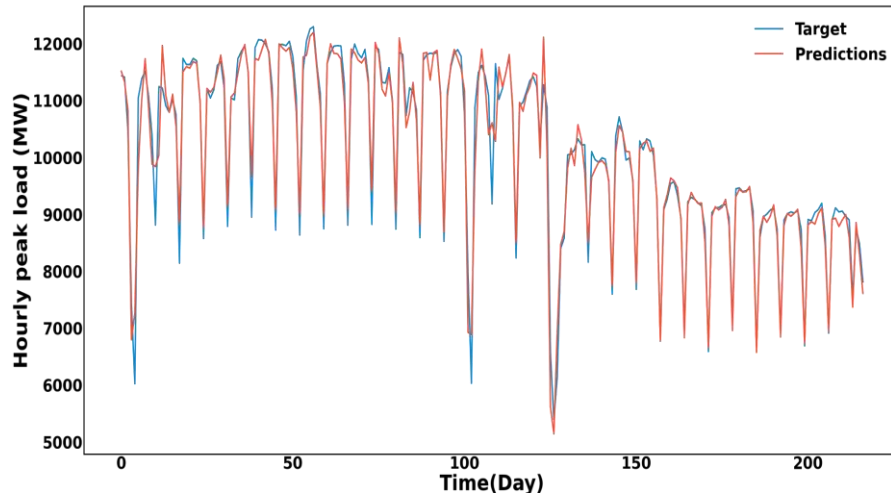


Figure 8. GRU: Day-ahead peak hour prediction public dataset

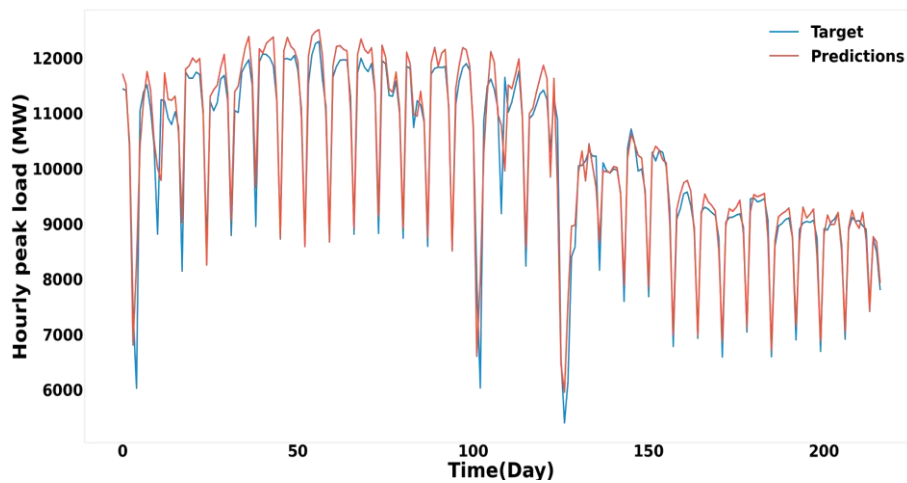


Figure 9. LSTM: Day-ahead peak hour prediction public dataset

5.3. Phase-3: Predicting energy usage for the following hour

In the third experiment, the peak energy usage for the following hour is predicted using the data from the preceding hours so that more precise forecasts may be produced using the most current data. Table 3 displays the performance ratings for the models that are suggested. Figures 10 and 11 present the performance graphs for the GRU and LSTM network topologies respectively.

Table 3: Next hour's forecast

MODEL	MAPE	MAE	RMSE(MW)
GRU	3.713	0.021	0.036
LSTM	4.204	0.024	0.038

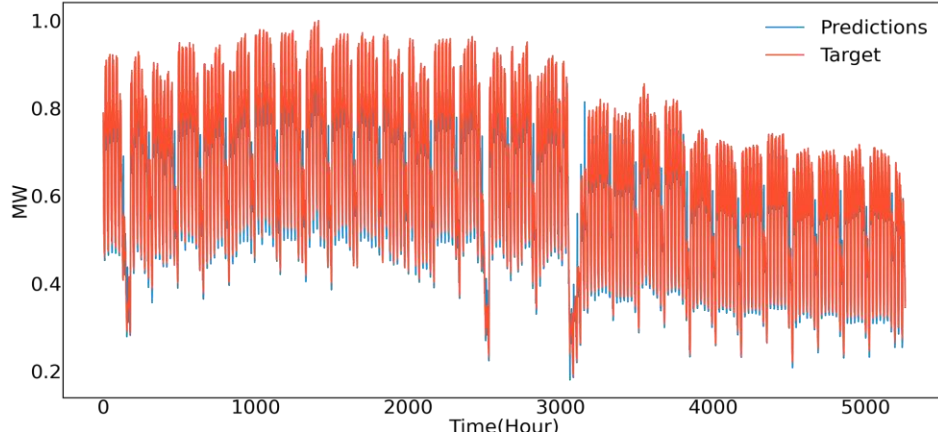


Figure 10. GRU: Next hour's prediction public dataset

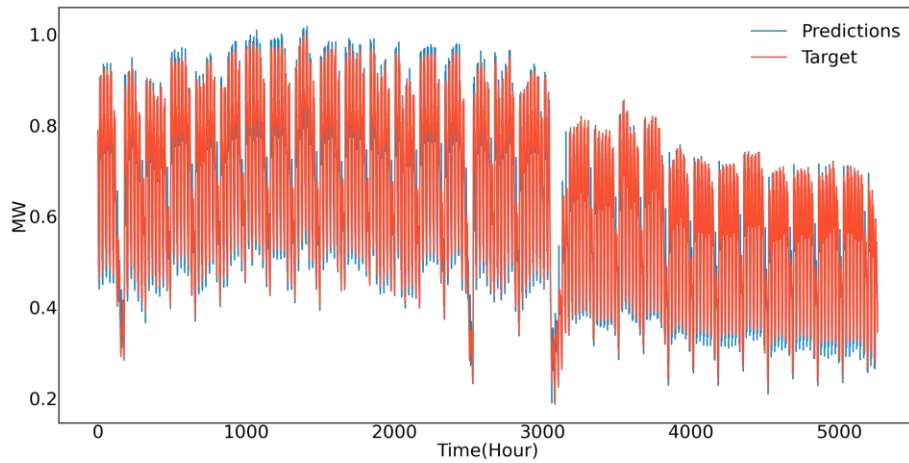


Figure 11. LSTM: Next hour's prediction public dataset

6. Performance comparison proposed and existing models

A comparison is made between the proposed models and the current ones. Publicly available dataset from electrical engineering mathematical modelling competition in 2016 [29] is used for this experiment. The outcomes, see Table 4, demonstrate that the suggested models outperform the current models in terms of electric load predictions.

Table 4: Performance comparison day ahead 24 hour's prediction using public dataset

Proposed Models		
MODEL	MAPE	RMSE(MW)
GRU	2.473	303.534
LSTM	2.563	349.988
Existing Models [2]		
TCN-AN	4.03	357.26
LSTM	4.93	506.61
CNN-LSTM	4.81	483.60

7. Conclusion and future scope

Based on historical data, different neural network models with GRU and LSTM architectures have been developed to forecast the energy demand for the next 24 hours. The impact of meteorological factors on energy consumption was studied, and the related factors were taken into account while developing the models. In addition to meteorological indicators, technical indicators such as average and simple moving average are also examined, with the outcomes being integrated into more advanced models. The study shows that data analytics plays a major role in developing neural network models for future forecasting. When accomplishment is measured and compared, the suggested GRU and LSTM frameworks are superior to the current models in terms of prediction accuracy. It is interesting to notice that, in most cases, the GRU networks fared better at predicting outcomes than the LSTM networks. Additionally, depending on the past hourly data, models are created to predict how much energy will be used in the upcoming hour. Future research might improve the models' ability to accurately anticipate energy loads by discovering and adding additional meteorological parameters as well as pertinent technical indications.

References

- [1] Ahmad O. Aseeri, "Effective RNN-Based Forecasting Methodology Design for Improving Short-Term Power Load Forecasts: Application to Large-Scale Power-Grid Time Series", *Journal of Computational Science* Volume 68, April 2023, 101984.
- [2] Xianlun Tang , Hongxu Chen, Wenhao Xiang , Jingming Yang, Mi Zou, "Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network", *Electric Power Systems Research* 205 (2022) 107761.
- [3] Seyed Mohammad Jafar Jalali, Student Member, IEEE, Sajad Ahmadian, Abbas Khosravi, Senior Member, IEEE, Miadreza Shafie-khah, Senior Member, IEEE, Saeid Nahavandi, Fellow, IEEE, and Joao P.S. Catalao, Senior Member, IEEE, "A Novel Evolutionary-based Deep Convolutional Neural Network Model for Intelligent Load Forecasting", 12 March 2021, *IEEE Transactions on Industrial Informatics* (Volume: 17, Issue: 12, December 2021).
- [4] Muhammad Qamar Raza, Abbas Khosravi, "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings", *Renewable and Sustainable Energy Reviews*, Volume 50, October 2015, Pages 1352-1372.
- [5] E. T. Sivadasan, N. Mohana Sundaram, R.Santhosh. "Stock market forecasting using deep learning with long short-term memory and gated recurrent unit", *Soft Computing*, 2024
- [6] Hao, Jin, "Input selection using mutual information - applications to time series prediction", 2004 - 2005, Master's thesis, Helsinki University of technology, Department of Computer and Information Science.
- [7] Milos B. Stojanovic, Milos M. Bozic, and Milena M. Stankovic, "MidTerm Load Forecasting Using Recursive Time Series Prediction Strategy With Support Vector Machines", *FACTA UNIVERSITATIS (NIS), SER.: ELEC. ENERG.* vol. 23, no. 3, Decemer 2010, 287-298.
- [8] J.P.S. Catalao, S.J.P.S. Mariano, V.M.F. Mendes, L.A.F.M. Ferreira, "Short-term electricity prices forecasting in a competitive market: A neural network approach", *Electric Power Systems Research* 77 (2007) 1297–1304 , ELSEVIER.
- [9] Henrique Steinhert Hippert, Carlos Eduardo Pedreira, and Reinaldo Castro Souza, "Neural Networks for Short-Term Load Forecasting: A Review and Evaluation", *IEEE TRANSACTIONS ON POWER SYSTEMS*, VOL. 16, NO. 1, FEBRUARY 2001.
- [10] C. S. Manigandaa,V. D. Ambeth Kumar,G. Ragunath,R. Venkatesan,N. Senthil Kumar. (2023). De-Noising and Segmentation of Medical Images using Neutrophilic Sets. *Journal of Fusion: Practice and Applications*, 11 (2), 111-123.
- [11] Shu Fan, Senior Member, IEEE, and Rob J Hyndman, "Short-term load forecasting based on a semi-parametric additive model", 18 August 2011 , *IEEE Transactions on Power Systems* (Volume: 27, Issue: 1, February 2012).
- [12] Nima Amjady, Senior Member, IEEE, and Ali Daraeepour, Member, IEEE, "Midterm Demand Prediction of Electrical Power Systems Using a New Hybrid Forecast Technique", *IEEE TRANSACTIONS ON POWER SYSTEMS*, VOL. 26, NO. 2, MAY 2011.
- [13] S. Hemamalini ,V. D. Ambeth Kumar ,R. Venkatesan,S. Malathi. (2023). Relevance Mapping based CNN model with OSR-FCA Technique for Multi-label DR Classification. *Journal of Fusion: Practice and Applications*, 11 (2), 90-110.

- [14] Wei-ting XU, Yun-ling Wang, Ting Li, Quan Tang, Jin-fang Zhang, Li Shen, Mi Zhu, "Long Term Intelligent Load Forecasting Method Considering the Expectation of Power Market Transaction", 2017 29th Chinese Control And Decision Conference (CCDC), Date Added to IEEE Xplore: 17 July 2017, ISBN:978-1-5090-4658-4.
- [15] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997
- [16] Sherubha, "Graph Based Event Measurement for Analyzing Distributed Anomalies in Sensor Networks", *Sādhanā*(Springer), 45:212, <https://doi.org/10.1007/s12046-020-01451-w>
- [17] Piyush K. Pareek, Pixel Level Image Fusion in Moving objection Detection and Tracking with Machine Learning "Fusion: Practice and Applications, Volume 2 , Issue 1 , PP: 42-60, 2020
- [18] Shivam Grover, Kshitij Sidana, Vanita Jain, "Egocentric Performance Capture: A Review", *Fusion: Practice and Applications, Volume 2, Issue 2 , PP: 64-73, 2020.*
- [19] Abdel Nasser H. Zaied, Mahmoud Ismail and Salwa El- Sayed, A Survey on Meta-heuristic Algorithms for Global Optimization Problems, *Journal of Intelligent Systems and Internet of Things*, Volume 1 , Issue 1 , PP: 48-60, 2020
- [20] Mahmoud H. Alnamoly, Ahmed M. Alzohairy, Ibrahim M. El-Henawy, "A survey on gel images analysis software tools, *Journal of Intelligent Systems and Internet of Things*, Volume 1 , Issue 1 , PP: 40-47, 2021.
- [21] Kumar, V.D.A., Sharmila, S., Kumar, A. et al. (2023). A novel solution for finding postpartum haemorrhage using fuzzy neural techniques. *Neural Comput & Applic.* 35(33), 23683–23696
- [22] Guanzhi Li, Aining Zhang, Qizhi Zhang, Di Wu, and Choujun Zhan, Member, IEEE. "Pearson Correlation Coefficient-Based Performance Enhancement of Broad Learning System for Stock Price Prediction", *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II: EXPRESS BRIEFS*, VOL. 69, NO. 5, MAY 2022.
- [23] Sathya Preiya, V., and V. D. Ambeth Kumar. (2023). Deep Learning-Based Classification and Feature Extraction for Predicting Pathogenesis of Foot Ulcers in Patients with Diabetes. *Diagnostics* 13(12), 1983.
- [24] Balakrishnan, Chitra, and V. D. Ambeth Kumar. (2023). IoT-Enabled Classification of Echocardiogram Images for Cardiovascular Disease Risk Prediction with Pre-Trained Recurrent Convolutional Neural Networks. *Diagnostics* 13(4), 775.
- [25] Hemamalini, Selvamani, and Visvam Devadoss Ambeth Kumar. (2022). Outlier Based Skimpy Regularization Fuzzy Clustering Algorithm for Diabetic Retinopathy Image Segmentation. *Symmetry*, 14(12), 2512.
- [26] R. Dey, F. M. Salem, "Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks", *Neural and Evolutionary Computing*, arXiv:1701.05923.
- [27] M. Hiransha, E.A. Gopalakrishnan, V.K. Menon, K.P. Soman, "NSE Stock Market Prediction using Deep-Learning Models", *Procedia computer science*, vol. 132, pp. 1351–1362, 2018.
- [28] Yi Qiao; Kaiwen Xu; Kai Zhou, "Research on time series based on improved LSTM", 2023 IEEE 3rd International Conference on Power, Electronics and Computer Applications (ICPECA), 29 March 2023.