

This paper presents an approach for using a long short-term memory (LSTM)-based recurrent neural network with various configurations to construct a forecasting model for electrical load prediction of a 110 kV substation.

The issues of unbalances arising in energy management systems due to discrepancies between generated and consumed energy can lead to power outages and blackouts. With the introduction of the most accurate forecasts, the task of maintaining electrical reliability for grid operators can be greatly simplified.

Through an investigation into 81 different parameter combinations the research revealed the optimal setup for an LSTM model in the task of electrical load forecasting. This configuration comprised 15 neurons, a batch size of 16, and employed the Adamax optimization algorithm. Applying this specific setup yielded a mean squared error (MSE) of 5.584 MW² and an R² value of 0.99. High R² values and low MSE values indicate that the LSTM model accurately captures changes in electricity consumption with minimal deviation between predicted and actual consumption values. Selection of appropriate parameters significantly enhances the performance of the predictive model and resulted in a reduction of the MSE error from 12.706 to 5.584 MW². The optimized configuration of the LSTM model, tailored through extensive experimentation, enhances its predictive capabilities.

The proposed LSTM model holds practical utility for integrating into systems for monitoring and forecasting mode reliability of electrical networks, particularly in the Western energy hub of the Republic of Kazakhstan. Its accuracy and reliability make it valuable for energy resource management and infrastructure planning

Keywords: short-term load forecasting, recurrent neural networks, long short-term memory-based load forecasting

UDC 621.311

DOI: 10.15587/1729-4061.2024.299192

CONSTRUCTION OF A RECURRENT NEURAL NETWORK-BASED ELECTRICAL LOAD FORECASTING MODEL FOR A 110 kV SUBSTATION: A CASE STUDY IN THE WESTERN REGION OF THE REPUBLIC OF KAZAKHSTAN

Yerlan Kenessov

Corresponding author

PhD Student*

E-mail: ye.kenessov@aes.kz

Karmel Tokhtibakiev

Candidate of Technical Sciences, Senior Lecturer*

Almaz Saukhimov

PhD

Technical Expert of Project Development and Digital Transformation
KazNIPI Energoprom JSC

Abylai Khan ave., 58A, Almaty, Republic of Kazakhstan, 050004

Daniil Vassilyev

Master's Student*

Alexandr Gunin

PhD Student*

Azamat Ilyasov

PhD Student*

*Department of Electric Power Systems

Almaty University of Power Engineering and Telecommunications
named after Gumarbek Daukeyev

Baytursynuli str., 126/1, Almaty, Republic of Kazakhstan, 050013

Received date 16.01.2024 **How to Cite:** Kenessov, Y., Tokhtibakiev, K., Saukhimov, A., Vassilyev, D., Gunin, A., Ilyasov, A. (2024). Construction of a recurrent neural network-based electrical load forecasting model for a 110 kV substation: a case study in the western region of the Republic of Kazakhstan. Accepted date 26.03.2024 Eastern-European Journal of Enterprise Technologies, 2 (8 (128)), 6–15. <https://doi.org/10.15587/1729-4061.2024.299192> Published date 30.04.2024

1. Introduction

The energy system of the Republic of Kazakhstan, much like many other nations, is a complex network comprising multiple interconnected regions. The regions are interconnected to ensure the stability and reliability of the energy system. Parallel operation of interconnected systems provides for mutual regulation of unplanned power flows caused by an unbalanced schedule of production and consumption of electric power. At the maximum load of the UES of the Republic of Kazakhstan in the amount of 14335 MW, deviations from the production-demand schedule amounted to 16.7 %, which was a critical value for managing normal

and post-accident modes, one of the reasons is the insufficient controllability of generation, which should track changes in consumption [1]. The challenges faced in ensuring stability and reliability within this system are not unique to the Republic of Kazakhstan alone but are shared by numerous countries worldwide. Therefore, the development of effective solutions for managing energy systems, such as accurate electrical load forecasting, is of global significance.

From July 1, 2023, the functioning of the balancing electric energy market has been transferred from a simulation mode (operating for more than 15 years) to a real-time mode in Kazakhstan. The balancing electricity market (BEM) is a system of relationships between market entities and the

settlement center of the balancing market, arising as a result of the physical settlement of electricity imbalances in the unified electric power system of the Republic of Kazakhstan by the system operator and associated with the purchase and sale of balancing electricity and negative imbalances [2]. BEM entities are required to enter into agreements with the Settlement Center for the purchase and sale of balancing electricity and negative imbalances, as well as a connection agreement. In this regard, there is a need for accurate forecasting of electricity consumption.

One of the key features to improve the efficiency of power system management is electrical load forecasting, which plays an essential role in the control of balance between production and consumption. The value of the predicted consumption scale is an important indicator in the planning of electrical modes for the whole system, group of systems, individual electrical energy consumers, and particular nodes in the electrical system. The forecasting task is based on complex mathematical or empirical (intuitive) methods for searching for patterns in the considered time interval. Electrical load forecasting is generally a univariate time series forecasting problem, which is more complex than the corresponding multivariate time series forecasting problem. Compared with linear, stationary and seasonal time series, short-term time series of electrical load in power systems are nonlinear, non-stationary and non-seasonal, where non-seasonal means no obvious periodicity in time. It is difficult to accurately forecast such time series over time.

Electrical load forecasting is a crucial aspect of managing the balance between electricity production and consumption. Its importance extends beyond the borders of the Republic of Kazakhstan, as it directly impacts the planning and operation of electrical systems in diverse geographical locations. Moreover, the challenges associated with forecasting electrical loads – such as nonlinearity, non-stationarity, and lack of obvious periodicity are common across various power systems globally. Owing to the operation of the electricity balancing markets and the necessity to mitigate adverse imbalances for electrical energy consumers, alongside the complexities involved in employing conventional mathematical forecasting methods due to their linear correlation between observed and future time series, researches on the construction of more efficacious neural network-based forecasting models becomes relevant.

2. Literature review and problem statement

As the requirement for electrical load forecasting emerged, many different forecasting methods have been applied for electrical load predictions such as Time Series Analysis, Regression Analysis, Artificial Neural Networks, Support Vector Machines (SVM), Hybrid models, etc.

Time series analysis involves analyzing historical load data to identify patterns, trends, and seasonality. It forms the basis for many load forecasting techniques. Common approaches within time series analysis include:

- moving Averages: Calculating the average load over a specific period, like days or weeks, to smooth out short-term fluctuations;
- exponential Smoothing: Assigning exponentially decreasing weights to past observations to emphasize recent data while minimizing the impact of older data.

- autoregressive integrated moving average (ARIMA) and the autoregressive moving average (ARMA) models seek to capture the temporal dependencies within a time series data by considering past values, past forecast errors, or both. These methods were first introduced in 1970 by two statisticians, George Box and Gwilym Jenkins [3].

A general class of methods includes statistical forecasting models, including autoregressive (AR), moving average (MA), autoregressive integrated moving average (ARIMA) models. Among them, ARIMA stands out as one of the most widely used time series forecasting methods.

The paper [4] presents the utilization of two forecasting methods: Auto Regressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN). It compares the performance of both methods using Mean Absolute Percentage Error (MAPE). The results suggest that both ANN and ARIMA have the potential to predict consumption but ANN better copes with non-linear data. However, the issue of constructing an optimal model for the most accurate load forecast has not yet been resolved. The paper [5] proposes autoregressive moving average (ARMA) model using including non-Gaussian process considerations the concept of cumulant and bispectrum embedded into the ARMA model in order to facilitate the Gaussian and non-Gaussian process. Nevertheless, using this approach may increase the time and resources required for model fitting and forecasting, particularly for large datasets.

The paper [6] proposes a study of different techniques: multiple linear regression (MLR), random forests (RF), artificial neural networks (ANNs), and automatic regression integrated moving average (ARIMA) to determine the load demand for the next month in the whole country and different municipal areas in Dubai, as well as to assist a utility company in future system scaling by adding new power stations for high-demand regions. The findings indicated that both ANN and RF attained a high level of accuracy, approximately 97 %. Nonetheless, given the utilization of a test dataset, it proves challenging to definitively assert the suitability of these models for practical application.

In [7], the authors introduced the application of SARI-MAX with standard statistical measures. They endeavored to segregate the impacts of these events and forecast the static and dynamic components of system demand separately. The resulting demand forecast accuracy proved to be superior compared to directly applying standard methods.

However, these methods operate on the basis of a linear relationship between observed and future time series, which makes them less effective for time series that exhibit pronounced nonlinear characteristics. In general, fluctuations in electrical load consumption turn out to be a complex non-stationary random process. They differ in seasonal changes in temperature and duration (longitude of the day) in the context of the year, the technological mode of operation of enterprises, the mode of work, and rest of the population. This circumstance can be excluded by using intelligent methods such as artificial neural networks (ANNs).

Artificial Neural Networks (ANNs) are a type of machine learning model inspired by the human brain. They are used for complex load forecasting tasks by learning intricate relationships between input variables and historical load data. There are many research papers that deal with neural networks because the use of conventional models leads to unsatisfactory solutions due to the high complexity of variables relationships and the extent of computation power

requirements. The paper [8] provides a comprehensive review of methodologies and applications using recent developments in ANN, ML, and DL for forecasting in microgrids, with the goal of providing a systematic analysis. Among the published studies, some were successful due to geographical location, workload, season, holidays, etc., but others failed to show good results. Possible explanations for this could include insufficient data for generating accurate forecasts, inadequate noise removal during data cleaning, and the selection of inappropriate model parameters. Detailed analysis shows that ANNs cannot work effectively with time series that have variable intervals between samples, which is often typical for electricity consumption data. However, some adaptations and extensions of support vector machines (SVM) have been proposed to handle time series data with irregular intervals. For instance, Support Vector Regression (SVR) has been developed to address time series forecasting tasks using SVM-based approaches.

SVM is a machine learning technique used for classification and regression tasks. It works by finding a hyperplane that best separates data points in high-dimensional space, making it suitable for capturing nonlinear relationships in load forecasting. The use of SVM is presented in [9] where authors examined six metaheuristic optimization algorithms, culminating in the identification of the most effective algorithm for optimizing SVR model parameters. The integration of the SVR-AEO hybrid model facilitated the attainment of notably favorable forecast outcomes.

In [10], the authors employ support vector regression (SVR) to address nonlinear electrical load forecasting issues under conditions of limited information. They propose SVR modeling utilizing the Gaussian kernel function along with historical electrical load data as input and target data. However, SVM techniques may not be efficient when dealing with large data sets due to high computational complexity, especially when using kernel functions.

Most of traditional methods have their limitations when dealing with complex load patterns and nonlinear relationships between load and influencing factors. Therefore, there is a growing need for advanced forecasting techniques that can capture these intricate dynamics and provide reliable predictions. However, recurrent neural network-based electrical load forecasting offers several advantages over conventional methods. These advantages stem from the neural network's ability to capture complex patterns and relationships in data, making it a powerful tool for accurate load prediction. The use of neural network models for the problem of predicting electrical loads has the following advantages:

Handling complex dependencies: neural networks (NNs) are able to automatically detect and take into account complex dependencies and interactions between various factors that affect energy consumption.

Adaptability to change: neural networks can learn from historical data and adapt to changes in conditions that can affect electricity consumption, such as seasonal changes, holidays, changes in consumer behavior, etc. This makes them more flexible than some traditional methods.

Big Data Processing: neural networks are capable of processing large amounts of data, which can be critical for accurately predicting electricity consumption given many factors such as weather, economic factors, demographics, and others.

Better generalizing ability: a well-tuned neural network is able to generalize its knowledge and predictions to new situations that have not been considered in the historical

data. This allows more accurate forecasting of electricity consumption in various scenarios.

Automation: neural networks can be configured to automatically update and retrain as new data is received. This simplifies the process of updating forecasting models and allows them to adapt to changing conditions.

It is important to understand that the effectiveness of NNs for predicting the consumption of electrical loads can depend on many factors, such as the choice of network architecture, the volume and quality of data, training and validation processes, and the context of application. All this allows us to assert that it is expedient to conduct a study on the challenge of accurate short-term electrical load forecasting using an RNN-based model.

3. The aim and objectives of the study

The primary aim of this study is to develop a robust and precise electrical load forecasting model using neural networks. The expected outcome of the work is a comprehensive validation of the NN-based load forecasting approach, showcasing its ability to provide accurate and reliable predictions for short-term load management:

- to design an NN-based architecture suitable for load forecasting, and experiment with different configurations to find the optimal architecture;
- to evaluate model's accuracy using appropriate error metrics.

4. Materials and methods

4.1. Object and hypothesis of the study

The object of the study is electrical loads. The main hypothesis of the study is that recurrent neural networks (in particular, LSTM) produce fairly accurate predictions and are suitable for predicting electrical loads. It is assumed that the use of different hyperparameters is necessary because the prediction accuracy without their selection for a particular case may not be high enough. Simplifications have been adopted in the work: for the forecasting task, only date, time and electrical load are used as initial data, whereas in real conditions it is necessary to take into account external factors, such as weather conditions, holidays and weekends, etc.

4.2. Methodology

For the task of electrical load forecasting, selecting the appropriate type of forecasting model depends on the characteristics of the data, the complexity of load patterns, and the forecasting horizon. As mentioned earlier, neural network-based forecasting models offer several advantages over conventional methods. There are many methodologies that are being explored in recent years such as [11–15], which have shown promise in much better forecasting results as compared to traditional methods.

The most commonly used neural network types for electrical load forecasting tasks are: Feedforward Neural Networks (FNNs), Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs).

Feedforward Neural Networks (FNNs), often referred to as Multi-Layer Perceptrons (MLPs), constitute a pivotal class of artificial neural networks that occupy a central role in contemporary machine learning and deep learning re-

search. These neural architectures are characterized by their sequential and stratified structure, affording them the capacity to elucidate intricate data relationships. FNNs can be effectively applied to electrical load forecasting tasks [16, 17].

Using Convolutional Neural Networks (CNNs) for electrical load forecasting is an innovative approach that has shown promise in capturing spatial and temporal patterns within load data. To use CNNs, this data can be converted into a two-dimensional grid, where one axis represents time and the other axis represents different load zones or features. For example, rows could represent hours of the day, while columns represent different regions or load characteristics. To capture temporal dependencies explicitly, a variant known as Temporal Convolutional Networks (TCNs) can be employed. TCNs utilize causal convolutions, ensuring that information flows only from past to future, making them particularly suited for time-series forecasting tasks [18–20].

Recurrent neural network (convolutional neural network) – a special neural network, the purpose of which is to predict the next observation in a series. The idea behind RNN is the desire to extract useful information from a series of observations to make predictions. Accordingly, earlier observations must be remembered. In the RNN model, the inner layer serves to store information from previous observations of the series. The RNN simplified structure is presented in Fig. 1.

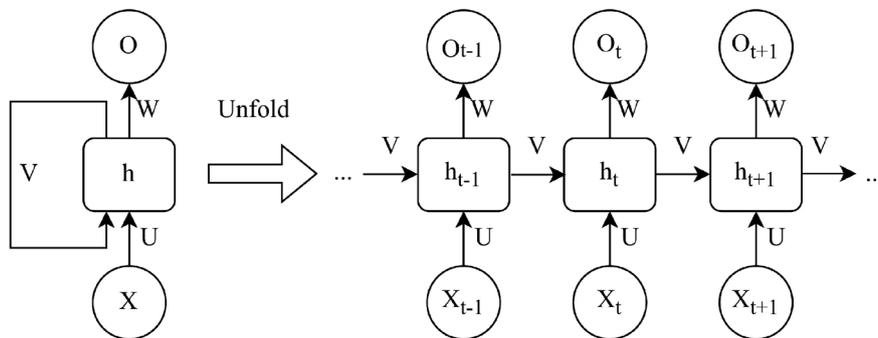


Fig. 1. Compressed (left) and unfolded (right) RNN structure

The main problem is the memorization of a small number of previous observations, which is not suitable for long periods. Given the sequential and temporal nature of load data, recurrent neural networks (RNNs) and their advanced variants are often well-suited for the electrical load forecasting task. Among these, the Long Short-Term Memory (LSTM) network stands out due to its ability to capture long-range dependencies and handle temporal patterns.

LSTM networks are a type of recurrent neural network (RNN) architecture designed to effectively capture and model sequential data, making them well-suited for tasks involving time series analysis, natural language processing, and various other sequential data applications. LSTMs were introduced to address some of the limitations of traditional RNNs, particularly their struggle to capture long-term dependencies in data. LSTMs offer a robust framework for modeling temporal dependencies, non-linearity, and multivariate relationships.

4. 3. Long short-term memory network-based electrical load forecasting

Long short-term memory networks, a special kind of RNN, are much promising having the capabilities to learn

long-term dependencies. The LSTM model [21] is a powerful recurrent neural system specially designed to overcome the exploding/vanishing gradient problems that typically arise when learning long-term dependencies, even when the minimal time lags are very long [22].

LSTM recurrent neural system specially designed to overcome the exploding/vanishing gradient problems. Unlike conventional feed-forward neural networks, LSTM networks have feedback. Such networks are capable of processing not only individual single data (for example, images), but also entire sequences of data (for example, audio recordings of speech or video recordings). Therefore, LSTM networks are capable of solving such problems as handwriting recognition, speech recognition, anomaly detection in large data flows (network traffic, banking transactions). LSTM neural networks are well-suited for classifying, processing, and making time-series-based forecasts, where interrelated phenomena can occur over an indefinite time span. This time lag leads to difficulties in using classical neural networks in solving these problems due to gradient decay, while LSTM networks are insensitive to the magnitude of the time lag.

Electricity data are in the nature of constantly repeating data (for example, a daily schedule), and their shape (peaks) changes only under the influence of external factors. Nevertheless, they are quite stationary and the influence of external factors (disconnection of consumers, holidays, etc.)

may well be taken into account when building a predictive model. Based on this, the LSTM model was chosen as the most suitable for solving the problem in this case study. An additional factor for choosing an LSTM model, compared, for example, with linear regression, is also the specific nature of the change in electricity consumption in winter and summer periods. If in the winter period, we can observe 2 peak values during the day, then in the summer there is only 1 peak value (in the middle of the day).

There are 3 types of layers in an LSTM network:

1. Layer of “forgetting” (Forget gate) – the output is a number from 0 to 1, where 1 indicates the need for complete memorization, and 0 completely erases from memory.
2. The “Memory gate” layer selects which data to store. First of all, values are selected using the sigmoid layer, which are then stored.
3. The output layer (Output gate) selects information from each “cell” in which the memorization is made.

The LSTM algorithm falls under the realm of supervised learning techniques. Within this framework, data is input into the Recurrent Neural Network (RNN) consisting of factors influencing the prediction of values for preceding hourly periods ($t-1$, $t-2$... $t-n$). These factors encompass temporal information, ambient temperature, and load. Additionally, a parameter is introduced, which plays a pivotal role in the prediction process. LSTMs exhibit the capability to selectively retain or remove information from the cell state, a process orchestrated by components referred to as filters.

Filters serve as regulatory structures that facilitate the filtration of information based on specific conditions. Each filter is comprised of a sigmoidal neural network layer coupled with a pointwise multiplication operation. The sigmoid

layer yields values ranging from zero to one, signifying the extent to which individual information blocks should progress through the network. A value of zero implies the retention of all information, whereas a value of one implies the exclusion of all information. Notably, the LSTM architecture features three such filters, collectively responsible for safeguarding and governing the cell state.

LSTMs are trained within the ambit of a supervised learning paradigm, wherein the network is trained to predict future values or sequences predicated on historical data. The iterative adjustment of the network’s weights and biases is achieved through Backpropagation Through Time (BPTT).

In our specific context, selecting the right values of hyperparameters can significantly influence the performance and quality of neural network training. This usually requires experimentation and tuning for a specific task and data set. For our data set, the following hyperparameters were used:

- optimizer: this is an algorithm used to update the weights (and possibly other parameters) of a neural network in order to minimize the loss function. Different optimizers have different properties and can produce different results when training a model. Some of the more popular optimizers that are often used in LSTM training include Adam, one of the most widely used optimizers, which usually provides a

good combination of learning speed and stability. Adamax, a variant of Adam that slightly modifies the updating of weights to account for infinite norms of gradients [23] and RMSprop, an optimizer that is also well suited for training LSTMs and is specifically adapted to work with recurrent neural networks [24, 25];

- number of neurons: this is the number of neurons in each layer of the neural network. This parameter determines the complexity of the model and its ability to learn complex functions;

- batch size: this is the number of samples fed to the model in one training pass (or step). It affects the learning speed and stability of weight updating.

Prior to input into the RNN, all data is subjected to a normalization procedure. Subsequently, the dataset is partitioned into training and testing subsets in the ratio of 70/30. The training subset is harnessed for the training of the RNN, while the test subset serves as a means to evaluate the accuracy of the trained model. It is noteworthy that owing to the unavailability or limited accessibility of certain specific data types, current load forecasting relies solely on historical data.

The load forecasting algorithm rooted in LSTM modeling is visually presented in Fig. 2.

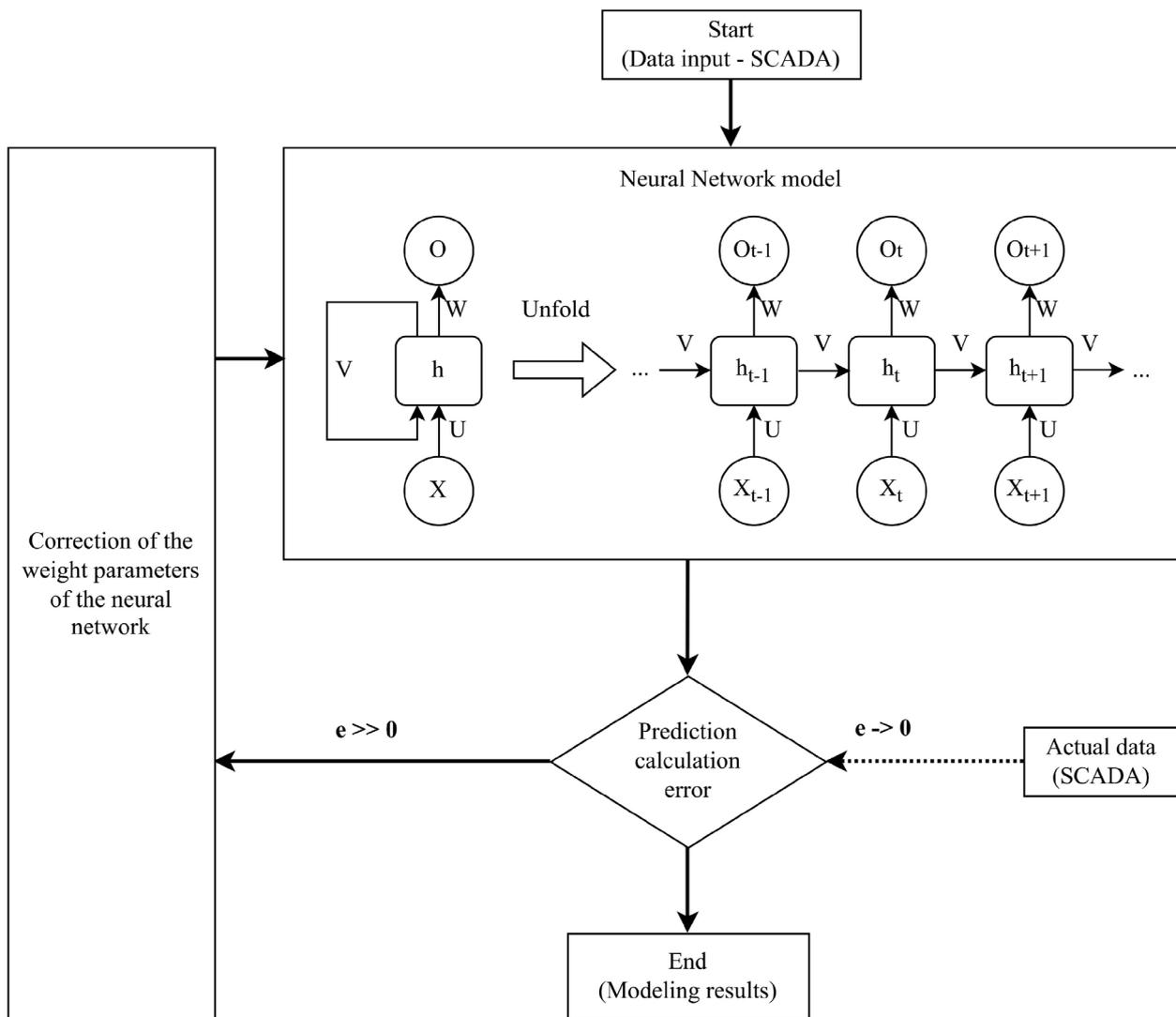


Fig. 2. Load forecasting algorithm on long short-term memory network-based model

The selection of weights for this model is an iterative process, a portion of the raw data that was previously separated from the training set is applied to test the accuracy of the model during training. In the future, it is necessary to perform regular checks on new data, if the accuracy decreases it is necessary to perform training again.

4. 4. Data collection and preprocessing

The model's focus will be on accurate predictions, leveraging historical load data from a 110 kV substation. Load forecasting was made for the 110 kV substation located in the western region of the Republic of Kazakhstan. Testing was carried out on the SCADA data archive for the period from 01/01/2022 to 01/09/2023.

Data preparation is the process of identifying and correcting errors, outliers, and inconsistencies in data in order to improve their quality, often classified as an integral part of data mining. For this task, the main statistical methods are used, such as correlation analysis and linear regression. The calibrated data is a data normalization of the range of values of independent variables or features of the data, which allows to adequately train the neural network.

Data archives were pre-processed, inconsistencies, errors, outliers and distortions in SCADA data were identified. At this point, it is important to check for missing data for each hour. The data processing algorithm allows you to replace zero or Not a Number (NaN) values with adjacent data. By filtering the data, errors were removed. Model training should generate data for the next 24, 48 and 96 hours based on historical and current data from the SCADA system (collection of information from power system objects).

5. Results of model configuration and training

5. 1. Model training, configuration and performance validation

The quality of the algorithms is checked by methods known in machine learning: MSE , $RMSE$, R^2 . To check the adequacy of the model, the mean square error (MSE) is calculated, all individual regression residuals are squared, summed, the sum is divided by the total number of errors:

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

The square root of this value is also calculated, denoted as $RMSE$ (Root Mean Square Error) to check discontinuous data in the event

that zeros appear in the data, it shows an adequate estimate:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}. \quad (2)$$

R^2 is a measure of the goodness of fit of a model [24]. In regression, the R^2 coefficient of determination is a statistical measure of how well the regression predictions approximate the real data points. An R^2 of 1 indicates that the regression predictions perfectly fit the data:

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}. \quad (3)$$

The result of the simulation is the day ahead hourly outgoing electrical load of the 110 kV substation. Following data preparation and normalization, we explored 81 distinct configurations of LSTM models, varying in initial dataset, number of neurons in the hidden layer, batch size, and optimization algorithm. Tables 1–3 present the results of testing these configurations using data from the preceding period of 24, 48, and 96 hours.

From the obtained tables we can conclude that the most accurate prediction is observed when using the Adam optimizer with the number of neurons equal to 15 and batch size equal to 16.

Table 1

Results of experiments using data from the preceding 24-hour period

Preceding time	Number of neurons	Batch size	Optimizer	MSE , MW ²	$RMSE$, MW	R^2
24	5	16	ADAMAX	6.641	2.201	0.966
24	5	32	ADAMAX	7.651	2.397	0.963
24	5	64	ADAMAX	9.127	2.673	0.939
24	10	16	ADAMAX	5.865	2.028	0.980
24	10	32	ADAMAX	7.402	2.418	0.967
24	10	64	ADAMAX	7.671	2.449	0.949
24	15	16	ADAMAX	6.341	2.087	0.979
24	15	32	ADAMAX	7.061	2.282	0.970
24	15	64	ADAMAX	9.096	2.693	0.883
24	5	16	ADAM	5.863	1.987	0.985
24	5	32	ADAM	7.375	2.283	0.981
24	5	64	ADAM	8.183	2.484	0.957
24	10	16	ADAM	5.661	1.904	0.989
24	10	32	ADAM	6.265	2.068	0.974
24	10	64	ADAM	7.166	2.263	0.962
24	15	16	ADAM	5.584	1.936	0.990
24	15	32	ADAM	6.397	2.078	0.970
24	15	64	ADAM	10.335	2.914	0.947
24	5	16	RMSprop	6.492	2.181	0.980
24	5	32	RMSprop	8.078	2.384	0.966
24	5	64	RMSprop	7.937	2.493	0.929
24	10	16	RMSprop	6.410	2.104	0.983
24	10	32	RMSprop	6.045	2.039	0.982
24	10	64	RMSprop	6.506	2.138	0.969
24	15	16	RMSprop	5.899	1.995	0.986
24	15	32	RMSprop	6.078	2.020	0.973
24	15	64	RMSprop	6.036	2.046	0.976

Table 2

Results of experiments using data from the preceding 48-hour period

Preceding time	Number of neurons	Batch size	Optimizer	MSE, MW ²	RMSE, MW	R ²
48	5	16	ADAMAX	6.905	2.145	0.983
48	5	32	ADAMAX	8.262	2.459	0.962
48	5	64	ADAMAX	10.497	2.787	0.940
48	10	16	ADAMAX	6.428	2.090	0.984
48	10	32	ADAMAX	7.011	2.239	0.971
48	10	64	ADAMAX	9.894	2.777	0.906
48	15	16	ADAMAX	6.305	2.065	0.982
48	15	32	ADAMAX	7.942	2.440	0.964
48	15	64	ADAMAX	9.924	2.847	0.930
48	5	16	ADAM	6.929	2.176	0.979
48	5	32	ADAM	6.523	2.124	0.976
48	5	64	ADAM	9.007	2.567	0.951
48	10	16	ADAM	6.751	2.189	0.977
48	10	32	ADAM	6.873	2.205	0.973
48	10	64	ADAM	11.256	3.062	0.870
48	15	16	ADAM	5.881	2.002	0.979
48	15	32	ADAM	6.059	2.041	0.981
48	15	64	ADAM	7.392	2.305	0.976
48	5	16	RMSprop	6.907	2.175	0.975
48	5	32	RMSprop	9.009	2.577	0.956
48	5	64	RMSprop	8.137	2.461	0.964
48	10	16	RMSprop	7.044	2.191	0.983
48	10	32	RMSprop	6.089	2.059	0.982
48	10	64	RMSprop	7.271	2.293	0.975
48	15	16	RMSprop	6.360	2.128	0.968
48	15	32	RMSprop	6.138	2.038	0.981
48	15	64	RMSprop	7.363	2.300	0.960

Table 3

Results of experiments using data from the preceding 96-hour period

Preceding time	Number of neurons	Batch size	Optimizer	MSE, MW ²	RMSE, MW	R ²
96	5	16	ADAMAX	9.161	2.562	0.961
96	5	32	ADAMAX	10.995	2.899	0.937
96	5	64	ADAMAX	15.998	3.643	0.839
96	10	16	ADAMAX	7.359	2.231	0.977
96	10	32	ADAMAX	8.171	2.425	0.967
96	10	64	ADAMAX	9.477	2.670	0.953
96	15	16	ADAMAX	8.085	2.355	0.974
96	15	32	ADAMAX	8.461	2.497	0.939
96	15	64	ADAMAX	9.852	2.779	0.924
96	5	16	ADAM	7.959	2.353	0.984
96	5	32	ADAM	7.772	2.310	0.983
96	5	64	ADAM	14.175	3.487	0.921
96	10	16	ADAM	7.203	2.210	0.987
96	10	32	ADAM	8.484	2.424	0.970
96	10	64	ADAM	9.562	2.594	0.946
96	15	16	ADAM	7.301	2.187	0.984
96	15	32	ADAM	9.586	2.671	0.976
96	15	64	ADAM	12.706	3.097	0.957
96	5	16	RMSprop	8.042	2.340	0.973
96	5	32	RMSprop	9.216	2.555	0.957
96	5	64	RMSprop	9.435	2.660	0.958
96	10	16	RMSprop	7.527	2.290	0.979
96	10	32	RMSprop	8.401	2.488	0.956
96	10	64	RMSprop	9.331	2.567	0.950
96	15	16	RMSprop	9.115	2.543	0.963
96	15	32	RMSprop	7.458	2.234	0.983
96	15	64	RMSprop	8.758	2.532	0.965

5. 2. Models accuracy validation

By employing the suggested model featuring the three most optimal hyperparameter configurations, predictive outcomes were achieved. The results were compared with the real electrical load data for the same day, as illustrated in Fig. 3.

Due to the presence of slight deviation from the actual load data, the MAE error parameter was calculated, MAE stands for Mean Absolute Error, which is a commonly used metric for evaluating the accuracy of forecasts or predictive models. It measures the average absolute difference between the predicted values and the actual observed values. In order to show the dynamics of the training process of the LSTM model, Fig. 4 illustrates a graph of MAE for each of 3 predictions depending on the number of epochs, where epoch is one complete pass through the entire training dataset.

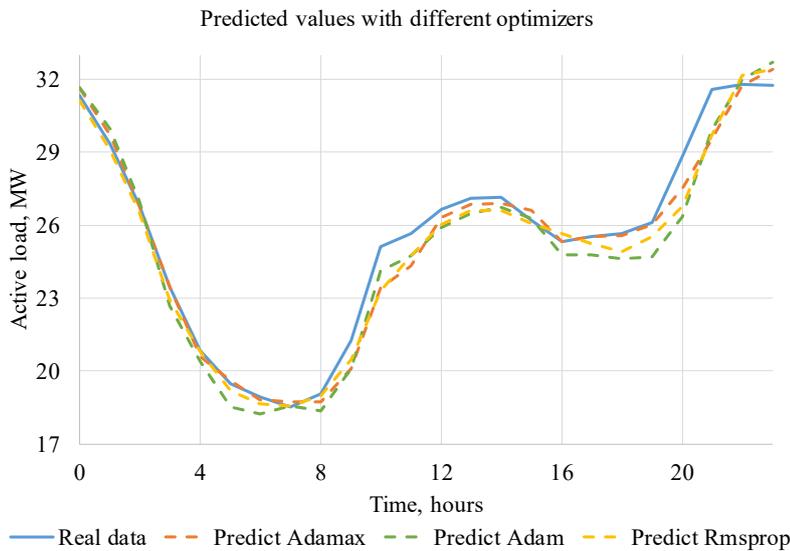


Fig. 3. Actual and forecasted electrical load values for 110 kV substation

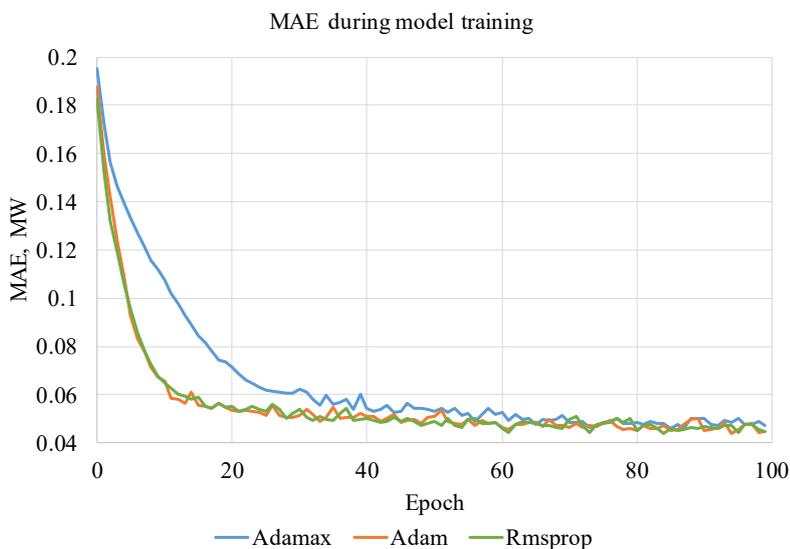


Fig. 4. Dependence of mean absolute error on the number of model epochs

During each epoch, the model is trained on the entire dataset, and the weights of the model are updated based on the error or loss calculated from the training data. As the model is exposed to more data or undergoes more training iterations, it learns from the discrepancies between its predictions and the actual outcomes. It adjusts its internal parameters (weights

and biases in the case of neural networks) to minimize these discrepancies and make more accurate predictions. The training process is terminated when there is no further significant reduction in loss. Too many epochs can lead to model overfitting, in which case forecasts on new data are highly inaccurate.

6. Discussion of the results of the constructed long short-term memory-based model

Initially, 81 different LSTM model configurations were investigated, the model was categorized into three groups with a preceding time of 24, 48 and 96 hours, documented in Tables 1–3. This classification was undertaken to facilitate the management of a substantial array of setting parameters. The most accurate prediction was made by a model with a lag of 24 hours, hidden number of neurons equal to 15, batch size equal to 16 and the “Adam” optimization algorithm, with *MSE*, *RMSE* and *R²* values of 5.584 MW², 1.936 MW and 0.99 respectively. This exceptional level of accuracy highlights the significant potential and robustness of LSTM models in electrical load forecasting.

The ability of LSTM networks to capture intricate patterns and dependencies within electrical load data is crucial for achieving such high levels of accuracy. This precision in forecasting is particularly valuable in applications where even minor deviations can have significant consequences, such as energy grid management and cost optimization.

In [26], the LSTM-based model also demonstrated good results with the best predicted week having an error of 1.65%. Compared to existing methods, the proposed LSTM-based approach demonstrates superior predictive capabilities. While traditional forecasting techniques may struggle to capture nonlinear relationships and dynamic patterns in electrical load data, LSTM models excel in this regard. The study provides a robust validation of the effectiveness of LSTM networks in electrical load forecasting, offering actionable insights for real-world decision-making.

A decrease in the values of mean squared error (1) and root mean squared error (2) indicates superior model performance, reflecting smaller discrepancies between predicted and actual values. R-squared (3) represents the proportion of the variance in the dependent variable that is explained by the independent variables (features) in the model, R-squared values range from 0 to 1, with 1 indicating a perfect fit and 0 indicating no relationship between the independent and dependent variables.

Despite its accuracy, the LSTM-based approach has certain limitations, such as the model focusing on short-term electrical load forecasting, limiting the generalizability of the results to longer forecasting horizons. Moreover, the accuracy of the forecasts is contingent upon the availability and quality of input data, including historical load data and relevant external factors such as temperature and humidity.

Future research can focus on addressing the following challenges and advancing the field of tuning hyperparameters and exploring hybrid forecasting techniques can further enhance the accuracy and robustness of LSTM-based forecasts. Integrating supplementary data sources, such as weather forecasts and socioeconomic indicators, can improve the model's predictive capabilities. Advancements in methodology, including ensemble techniques and model interpretability, can contribute to the continued progress of electrical load forecasting research.

7. Conclusions

1. The recurrent neural network-based model was trained using various optimizers, and its performance was validated on the test set. Experimentation was conducted on various configurations to design a neural network-based architecture suitable for load forecasting, employing a basic LSTM (Long Short-Term Memory) model preceded by an initial selection process for model parameters. 81 distinct configurations of LSTM models, varying in initial dataset, number of neurons in hidden layer, batch size, and optimization algorithm were explored. For the available input data, the most appropriate amount of input data is 24 and 48 hours. The best results were obtained with 24 hours as input data, number of neurons in the hidden layer equal to 15, batch size equal to 16 and applying ADAM optimization algorithm. It should be realized that depending on the input data, the optimal hyperparameters of the model will be different, so a separate selection of these parameters must be made for each substation. Also, the optimal hyperparameters will change over time, so if the accuracy of the model decreases, they must be re-selected.

2. The model's accuracy was evaluated on a test set, which was not used during training, using *MSE*, *RMSE*

and R^2 metrics. The best-performing configuration among the 81 considered parameter cases demonstrated promising results with an *MSE* of 5.584 MW², *RMSE* of 1.936 MW and an R^2 value of 0.99.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, whether financial, personal, authorship or otherwise, that could affect the study and its results presented in this work.

Financing

The studies given in this paper were conducted under the scientific and technical projects under a grant of the Ministry of Education and Science of the Republic of Kazakhstan for 2022–2024 years in the priority area “Power Engineering and Machinery” on the topic AP14870626 “Intelligent system of monitoring and forecasting the reliability of electrical networks of oil and gas complexes, including off-grid renewable energy”.

Data availability

The manuscript has no associated data.

Use of artificial intelligence

The authors have used artificial intelligence technologies within acceptable limits to provide their own verified data, which is described in the research methodology section.

References

- Ilyasov, A. Z., Saukhimov, A. A., Didorenko, Y. V., Landman, A. K., Janzakov, Y. B. (2022). Creation of a automation generation control system in the unified power system of Kazakhstan. *Bulletin of the Almaty University of Power Engineering and Telecommunications*, 1, 6–17. https://doi.org/10.51775/2790-0886_2022_56_1_6
- O rabote balansiruyushchego rynka elektricheskoy energii. Kazahstanskaya kompaniya po upravleniyu elektricheskimi setyami. Available at: <https://www.kegoc.kz/ru/electric-power/etn/>
- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. (2008). *Time Series Analysis*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc. <https://doi.org/10.1002/9781118619193>
- Tarmanini, C., Sarma, N., Gezezin, C., Ozgonenel, O. (2023). Short term load forecasting based on ARIMA and ANN approaches. *Energy Reports*, 9, 550–557. <https://doi.org/10.1016/j.egyr.2023.01.060>
- Huang, S.-J., Shih, K.-R. (2003). Short-term load forecasting via ARMA model identification including non-gaussian process considerations. *IEEE Transactions on Power Systems*, 18 (2), 673–679. <https://doi.org/10.1109/tpwrs.2003.811010>
- Sayed, H. A., William, A., Said, A. M. (2023). Smart Electricity Meter Load Prediction in Dubai Using MLR, ANN, RF, and ARIMA. *Electronics*, 12 (2), 389. <https://doi.org/10.3390/electronics12020389>
- Moghran, I., Rahman, S. (1989). Analysis and evaluation of five short-term load forecasting techniques. *IEEE Transactions on Power Systems*, 4 (4), 1484–1491. <https://doi.org/10.1109/59.41700>
- Barakat, E. H., Qayyum, M. A., Hamed, M. N., Al Rashed, S. A. (1990). Short-term peak demand forecasting in fast developing utility with inherit dynamic load characteristics. I. Application of classical time-series methods. II. Improved modelling of system dynamic load characteristics. *IEEE Transactions on Power Systems*, 5 (3), 813–824. <https://doi.org/10.1109/59.65910>
- Wazirali, R., Yaghoubi, E., Abujazar, M. S. S., Ahmad, R., Vakili, A. H. (2023). State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques. *Electric Power Systems Research*, 225, 109792. <https://doi.org/10.1016/j.epsr.2023.109792>
- Cai, W., Wen, X., Li, C., Shao, J., Xu, J. (2023). Predicting the energy consumption in buildings using the optimized support vector regression model. *Energy*, 273, 127188. <https://doi.org/10.1016/j.energy.2023.127188>

11. Yusuf, L. O. M., Handoko, S., Setiawan, I. (2023). Forecasting Short-Term Electrical Loads Using Support Vector Regression with Gaussian Kernel Functions. *Advances in Physics Research*, 131–142. https://doi.org/10.2991/978-94-6463-228-6_16
12. Choi, W., Lee, S. (2023). Performance evaluation of deep learning architectures for load and temperature forecasting under dataset size constraints and seasonality. *Energy and Buildings*, 288, 113027. <https://doi.org/10.1016/j.enbuild.2023.113027>
13. Aseeri, A. O. (2023). 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*, 68, 101984. <https://doi.org/10.1016/j.jocs.2023.101984>
14. Jayashankara, M., Shah, P., Sharma, A., Chanak, P., Singh, S. K. (2023). A Novel Approach for Short-Term Energy Forecasting in Smart Buildings. *IEEE Sensors Journal*, 23 (5), 5307–5314. <https://doi.org/10.1109/jsen.2023.3237876>
15. Singh, S., Tripathi, M. M. (2022). Short-Term Electricity Demand Forecast Using Deep RNN and Stacked LSTM. *Advances in Manufacturing Technology and Management*, 578–588. https://doi.org/10.1007/978-981-16-9523-0_64
16. Haque, A., Rahman, S. (2022). Short-term electrical load forecasting through heuristic configuration of regularized deep neural network. *Applied Soft Computing*, 122, 108877. <https://doi.org/10.1016/j.asoc.2022.108877>
17. Alhmoud, L., Abu Khurma, R., Al-Zoubi, A. M., Aljarah, I. (2021). A Real-Time Electrical Load Forecasting in Jordan Using an Enhanced Evolutionary Feedforward Neural Network. *Sensors*, 21 (18), 6240. <https://doi.org/10.3390/s21186240>
18. Meng, Y., Yun, S., Zhao, Z., Guo, J., Li, X., Ye, D. et al. (2023). Short-term electricity load forecasting based on a novel data preprocessing system and data reconstruction strategy. *Journal of Building Engineering*, 77, 107432. <https://doi.org/10.1016/j.jobe.2023.107432>
19. Jalali, S. M. J., Ahmadian, S., Khosravi, A., Shafie-khah, M., Nahavandi, S., Catalao, J. P. S. (2021). A Novel Evolutionary-Based Deep Convolutional Neural Network Model for Intelligent Load Forecasting. *IEEE Transactions on Industrial Informatics*, 17 (12), 8243–8253. <https://doi.org/10.1109/tii.2021.3065718>
20. Yin, L., Xie, J. (2021). Multi-temporal-spatial-scale temporal convolution network for short-term load forecasting of power systems. *Applied Energy*, 283, 116328. <https://doi.org/10.1016/j.apenergy.2020.116328>
21. Tang, X., Chen, H., Xiang, W., Yang, J., Zou, M. (2022). Short-Term Load Forecasting Using Channel and Temporal Attention Based Temporal Convolutional Network. *Electric Power Systems Research*, 205, 107761. <https://doi.org/10.1016/j.epsr.2021.107761>
22. Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9 (8), 1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
23. Hochreiter, S., Schmidhuber, J. (1996). LSTM can solve hard long time lag problems. In *Advances in neural information processing systems*. Vol. 9. MIT Press, 473–479. Available at: https://proceedings.neurips.cc/paper_files/paper/1996/file/a4d2f0d23dcc84ce983ff9157f8b7f88-Paper.pdf
24. Kingma, D. P., Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv*. <https://doi.org/10.48550/arXiv.1412.6980>
25. Zou, F., Shen, L., Jie, Z., Zhang, W., Liu, W. (2019). A Sufficient Condition for Convergences of Adam and RMSProp. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). <https://doi.org/10.1109/cvpr.2019.01138>
26. Caicedo-Vivas, J. S., Alfonso-Morales, W. (2023). Short-Term Load Forecasting Using an LSTM Neural Network for a Grid Operator. *Energies*, 16 (23), 7878. <https://doi.org/10.3390/en16237878>