Intelligent Electricity Load Forecasting Method using ARIMA-LSTM-Random Forest

Oleksandr AZIUKOVSKYI[1], Volodymyr HNATUSHENKO[2] Vita KASHTAN[3], Alla POLYANSKA[4], Andrzej JAMRÓZ[5]

[1] PhD, Assoc. Prof., Professor of Electric Drive Department, Faculty of Electrical Engineering, Dnipro University of Technology, Dmytra Yavornytskoho Ave 19, Dnipro, Ukraine, https://orcid.org/0000-0003-1901-4333, aziukovskiy_O@nmu.org.ua

[2] Dr.Sc., Prof., Head of Information Technology and Computer Engineering Department, Faculty of Information Technologies, Dnipro University of Technology, Dmytra Yavornytskoho Ave 19, Dnipro, Ukraine, https://orcid.org/0000-0003-3140-3788, hnatushenko.v.v@nmu.one

[3] Ph.D., Assoc. Prof., Associate Professor of Information Technology and Computer Engineering Department, Faculty of Information Technologies, Dnipro University of Technology, Dmytra Yavornytskoho Ave 19, Dnipro, Ukraine, https://orcid.org/0000-0002-0395-5895, kashtan.v.yu@nmu.one

[4] Dr.Sc., Prof. of Management and Administration Department, Institute of Economics and Management, Ivano-Frankivska National Technikal University of Oil and Gas, Karpatska, 15, Ukraine; AGH University of Krakow, Faculty of Management, al. Mickiewicza 30, 30-059 Kraków, Poland, https://orcid.org/0000-0001-5169-1866 , email: polyanska@agh.edu.pl

[5] Dr inż., AGH University of Krakow, Faculty of Management, al. Mickiewicza 30, 30-059 Kraków, Poland, https://orcid.org/0000-0002-1847-8630 , email: ajamroz@agh.edu.pl

http://doi.org/10.29227/IM-2025-01-18

Submission date: 26-05-2025 | Review date: 06-06-2025

Abstract

The instability of energy systems caused by internal economic factors and external challenges, including geopolitical conflicts, significantly complicates the process of planning and managing energy resources. An essential tool for implementing energy-saving measures is introducing modern computer technologies, including artificial intelligence systems, in the energy sector. Intelligent technologies make it possible to use methods for predicting electrical load, including artificial intelligence algorithms. This paper proposes a combined ARIMA-LSTM-Random Forest model for forecasting electric load. The combination of the approaches allows considering both linear and nonlinear dependencies in the data, which is critical to ensure the accuracy of forecasts. Using data for the previous seven days provides enough information to identify seasonal trends and fluctuations, which makes this a promising prospect for medium-term forecasting in energy monitoring tasks. Thus, combining the ARIMA, LSTM, and Random Forest methods achieves high accuracy in forecasting electricity consumption. The proposed approach is an optimal solution since it combines the advantages of each model and compensates for their shortcomings. The proposed ARIMA-LSTM-Random Forest method significantly improved the results: MSE = 0.27, RMSE = 0.23, MAPE = 0.35%. The method minimized absolute and relative errors, confirming its advantage for this forecasting task. The results are promising for practical application in the load management of electric networks.

1. Introduction

The management of resources contributes to economic stability, reduces the artificial impact on the environment, and supports sustainable development. The instability of energy systems caused by internal economic factors and external challenges, including geopolitical conflicts, significantly complicates the process of planning and managing energy resources [1]. Introducing modern technologies and increasing energy efficiency will help reduce dependence on imported energy and strengthen energy sustainability. An essential tool for implementing energy-saving measures is introducing modern computer technologies, including artificial intelligence systems [2, 3], in the energy sector.

Intelligent technologies make it possible to use methods for predicting electrical load, including artificial intelligence (AI) algorithms. Traditional methods, such as time series analysis and regression analysis, demonstrate low accuracy in cases of stable data, but their effectiveness is significantly reduced when analyzing complex nonlinear dependencies. Instead, AI algorithms have significant potential to solve these problems, providing higher forecast accuracy.

Effective forecasting of electricity load is one of the key tasks of the modern energy sector. Accurate predictions allow energy providers to balance supply and demand more efficiently, avoiding both shortages and excesses [4]. Intelligent IT systems make it possible to predict electricity demand, which helps reduce overproduction, cut greenhouse gas emissions, and use resources more efficiently [4, 5]. Machine learning algorithms, as part of these intelligent systems, can detect patterns of inefficient energy consumption and suggest

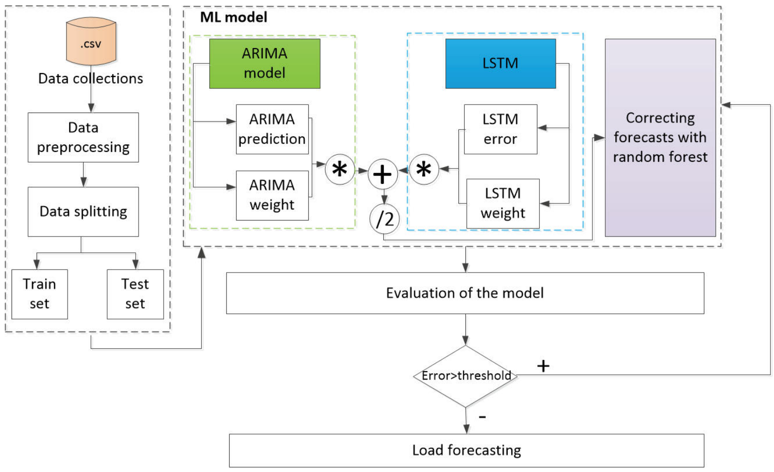

Fig. 1. Flowchart of intelligent electricity load forecasting method

Rys. 1. Schemat blokowy inteligentnej metody prognozowania obciążenia energią elektryczną

optimal strategies to minimize waste [6]. These technologies contribute not only to reducing operational costs but also to achieving environmental sustainability goals. As a result, energy systems become more resilient, adaptive, and aligned with global decarbonization efforts.

An essential task of every country is to optimize the use of energy resources, reduce environmental impact, and introduce renewable energy sources [7]. As global energy demand continues to rise, efficient resource management becomes increasingly important for ensuring energy security and economic stability [8]. Reducing environmental impact involves not only lowering emissions but also minimizing waste and preserving natural ecosystems [9]. The transition to renewable energy sources, such as solar, wind, and hydroelectric power, plays a crucial role in achieving long-term sustainability. By investing in cleaner technologies and smarter energy systems, countries can build more resilient and environmentally responsible energy infrastructures.

The transition to renewable energy sources, such as solar, wind, and hydroelectric power, plays a crucial role in achieving long-term sustainability. These sources offer a cleaner alternative to fossil fuels, significantly reducing greenhouse gas emissions and air pollution in the mining sector [8, 10]. Shifting to renewables also helps decrease dependence on imported energy, enhancing national energy security [11]. By investing in cleaner technologies and smarter energy systems, countries can build more resilient and environmentally responsible energy infrastructures. Modern energy systems integrate digital tools, such as smart grids and real-time monitoring, to optimize performance and reliability [10 - 12]. This transformation supports not only environmental goals but also economic development and job creation in the green energy sector.

Intelligent systems can process large amounts of data, assess risks, and provide optimal recommendations to improve energy companies' efficiency. Insufficiently accurate forecasting can lead to negative consequences, from localized power outages to significant economic losses due to excessive or insufficient electricity production. The integration of intelligent forecasting systems contributes to the efficient planning of electricity generation and distribution, which helps to reduce the manufactured impact, cut costs for energy companies, and ensure the stable operation of power grids.

2. Related works

Exponential smoothing [13] and the autoregressive moving average (ARIMA) model [14] are examples of classical forecasting methods based on time series analysis. These methods apply the concepts of mathematical statistics and stochastic process theory to assess system development trends. The main disadvantage of the above techniques is their limitation in forecasting based on multiple variables. These methods demonstrate high accuracy for time series with low fluctuation, where the absence of sharp changes allows for reliable forecasts. However, in the case of multiple factors affecting the predicted variable, such as weather conditions or date changes, these methods cannot effectively account for interactions between multiple variables. Therefore, they are used in complex scenarios where several interrelated variables must be analyzed simultaneously for accurate forecasting.

In case it is necessary to take into account multiple factors affecting the load, it is advisable to apply multivariate forecasting using machine learning methods, such as support vector machines (SVMs) [15], random forests [16], and XGBoost [17, 18]. These models can automatically detect more complex data structures and estimate non-linear relationships between various factors and load. However, their effectiveness could be improved by their heavy reliance on high-quality and similar datasets. It may cause instability when building and updating models and reduce the accuracy of forecasts, which only sometimes meets the requirements of modern power systems.

Due to their properties of associative memory, parallel information distribution, self-learning, and the ability to approximate complex functions, Neural networks can effectively analyze nonlinear relationships between load data and other variables. One of the simplest and most basic neural network architectures is the multilayer perceptron (MLP) [19, 20], which consists of fully connected neurons. In such a network, load data and related variables are passed through the network, and the parameters are optimized using the Back Propagation (BP) method [21], which allows for more accurate predictions, approaching actual values.

In a paper [22], the authors show that the proposed model based on long short-term memory (LSTM) outperforms other forecasting methods in terms of accuracy compared to alternative approaches. The study results indicate that this model

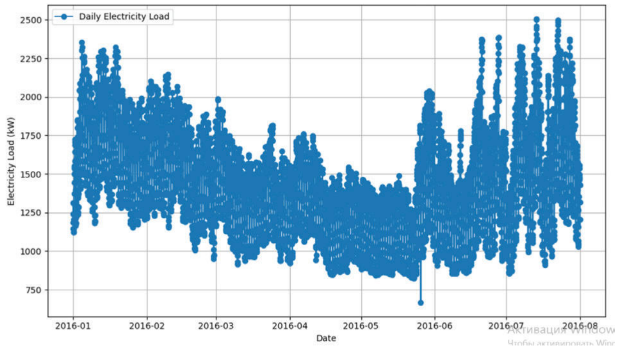

Fig. 2. Daily Electricity Load (kW) in 2016 year

Rys. 2. Dzienne zapotrzebowanie na energię elektryczną (kW) w roku 2016

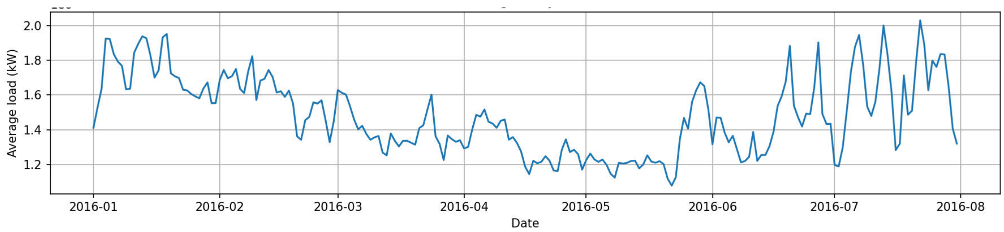

Fig. 3. Average daily load

Rys. 3. Średnie obciążenie dobowe

Fig. 4. Average hourly load

Rys. 4. Średnie obciążenie godzinowe

is an effective tool for electric power companies and energy management systems focused on optimizing operations and reducing costs. Of note is a new methodology for forecasting ultra-short-term electricity demand [23], which integrates the LSTM neural network with other machine learning methods to create a hybrid ensemble model. This approach improves forecast accuracy by combining the strengths of different algorithms. The paper by [24] improves the efficiency of the LSTM model for predicting electric load by selecting features and applying genetic algorithms. The authors of [25] presented a parallel genetic LSTM model enhanced with an evolutionary genetic algorithm (EPGA) that demonstrates high performance for short-term load forecasting (STLF) in the Jimma city electricity distribution system.

The article proposes a method of intelligent electric load forecasting using a hybrid model that combines ARIMA, LSTM, and Random Forest algorithms. This integrated approach leverages the strengths of each individual method to improve overall forecasting accuracy. By combining statistical analysis, deep learning, and ensemble techniques, the model can capture both linear and nonlinear patterns in energy consumption data. As a result, it enables effective short-term and long-term electric load forecasting, which is essential for efficient energy planning and management.

3. Methods and Methodology for Intelligent electricity load forecasting

The proposed intelligent electricity load forecasting framework integrates three complementary techniques, ARIMA (AutoRegressive Integrated Moving Average), LSTM (Long Short-Term Memory networks), and Random Forest, to enhance prediction accuracy and robustness for both short-term and long-term forecasting [26]. This hybrid approach is designed to leverage the strengths of each method: ARIMA is effective in modeling linear temporal relationships and seasonality, LSTM captures complex nonlinear time-series patterns and long-term dependencies, while Random Forest serves as an ensemble-based regressor that refines the final output by learning from residuals and feature interactions. Together, these methods enable the system to forecast electricity load with greater precision than any individual model alone [3, 5, 27].

The methodology begins with data collection and preprocessing. Historical electricity load data, often combined with external factors such as temperature, humidity, and time-based indicators (e.g., hour, day of the week, and holidays), are prepared through a series of steps including cleaning, normalization, and feature engineering [26, 28]. The ARIMA model is applied first to identify and model linear trends in the time series. Once the ARIMA model provides its forecast,

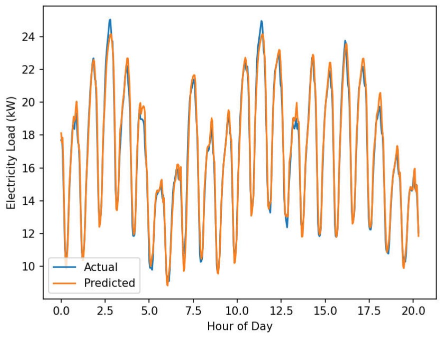

Fig. 5. Electricity load chart for the last twenty hours

Rys. 5. Wykres obciążenia energią elektryczną w ciągu ostatnich dwudziestu godzin

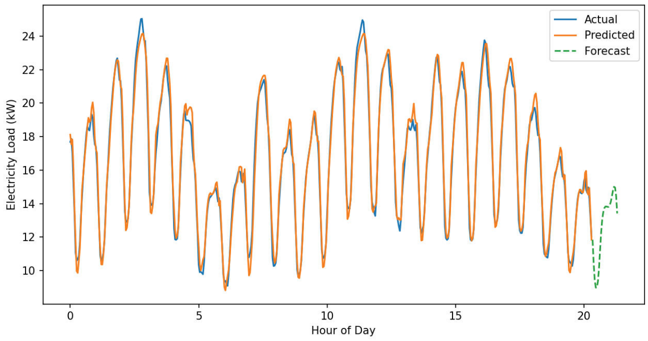

Fig. 6. Electricity Load Forecast

Rys. 6. Prognoza obciążenia energią elektryczną

Fig. 7. Electricity Load Forecast using machine learning methods and the proposed method

Rys. 7. Prognoza obciążenia energią elektryczną z wykorzystaniem metod uczenia maszynowego i proponowanej metody

the residuals, representing unexplained nonlinear behavior, are passed to the LSTM network [26, 27]. The LSTM is trained on sequences of past load values and features to capture long-range and nonlinear dependencies. Hyperparameters are fine-tuned to achieve optimal performance.

In the final stage, the outputs from both ARIMA and LSTM models, along with additional engineered features, are fed into a Random Forest regressor [26, 29]. This ensemble model acts as a meta-learner that further improves prediction accuracy by reducing variance and handling complex feature interactions [26-30]. The model's performance is evaluated using standard metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Comparative analysis with standalone models demonstrates the superior accuracy and robustness of the proposed hybrid method, making it a valuable tool for intelligent energy forecasting and resource management.

Fig. 1 shows a flowchart of the proposed intelligent electric load forecasting method using ARIMA-LSTM-Random Forest. The diagram illustrates the sequential steps involved in data preprocessing, model training, and prediction generation. It visually highlights how ARIMA handles linear trends, LSTM captures nonlinear time dependencies, and Random Forest refines the final output. This structured approach ensures a clear understanding of how the hybrid model integrates multiple techniques to achieve accurate and reliable forecasting.

The study uses open data [31]. The data are presented in .csv format and contain indicators of electric load in the Toronto regional power grid for 2016 (Fig. 2). They reflect electricity consumption in kilowatts and serve as the basis for time series analysis and forecasting.

Before applying machine learning methods, the dataset was pre-processed and prepared, including filling in missing values and standardizing the data structure [32]. This stage

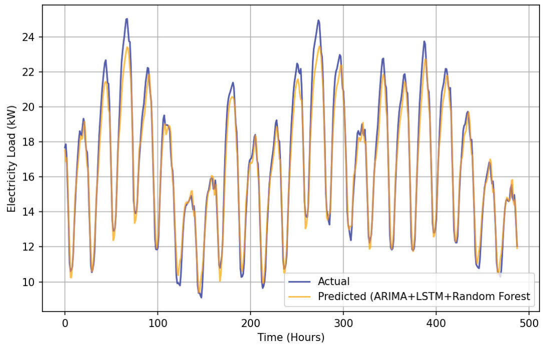

Fig. 8. Experiment 1 Load forecast for the next hour (seven days)

Rys. 8. Eksperyment 1 Prognoza obciążenia na następną godzinę (siedem dni)

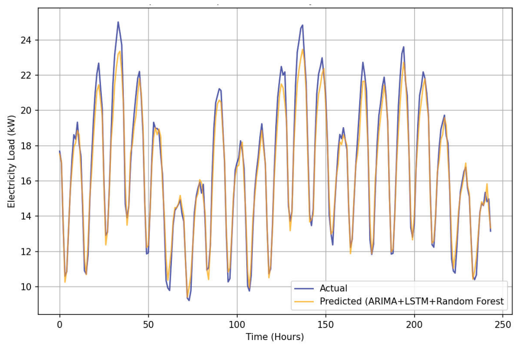

Fig. 9. Experiment 2 Forecast for the next month (seven days)

Rys. 9. Eksperyment 2 Prognoza na najbliższy miesiąc (siedem dni)

Fig. 10. Experiment 3 Load forecast for the next hour (two days)

Rys. 10. Eksperyment 3 Prognoza obciążenia na następną godzinę (dwa dni)

is essential to ensure the high quality of the input data, minimize systematic errors in forecasting, and improve the accuracy of the results.

Cleaning included the removal of missing values (NaN) and erroneous data that could affect the modeling results. The presence of gaps or abnormal values in the time series was determined by visualizing and statistically evaluating them.

All features were normalized to a single scale (-1, 1) to eliminate the influence of different units of measurement on the modeling results. Normalization helps to improve model convergence during training and prevents uncontrolled growth of gradients, which can lead to numerical instability:

(1)where F is the input data, F is the

ˆ

normalized data, F_{max} and F_{min} are the maximum and minimum values of the feature,

After preprocessing, the dataset includes a normalized time series with all characteristics. This data is input for model training and further electric load forecasting. Prepared data provides high-quality input for computations, which is essential for forecast accuracy.

After pre-processing, the data was divided into training and test sets. The training set contained 80% of all data and was used to train machine learning (ML) models. The remaining 20% was the test set, which was used to evaluate the performance of the models. This ratio provides enough data for training while leaving an independent sample for validation, preventing the models' overtraining.

The automatic integrated model of autoregression (ARIMA) is used to model time series [33] and consists of two steps: ARIMA Prediction and ARIMA Weight. It allows for the improvement of forecast accuracy by using weighting coefficients. The ARIMA Prediction step uses a predicted time series with three parameters: p – autoregression order, d – de-

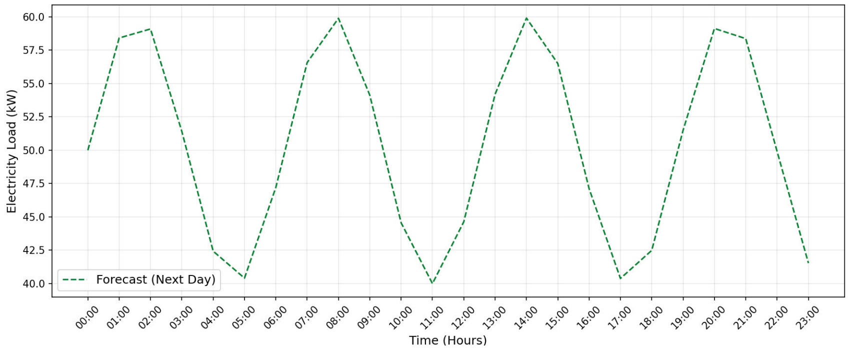

Fig. 11. Experiment 4 Load forecast for the next hour (one day)

Rys. 11. Eksperyment 4 Prognoza obciążenia na następną godzinę (jeden dzień)

Tab. 1. Assessment of Forecasting Accuracy

Tab. 1. Ocena dokładności prognozowania

Method	MSE	RMSE	MAPE
LSTM Forecast	0 38	0 62	8 37%
ARIMA STM	0 63	0 79	1 27%
ARIMA LSTM Random Forest	0.27	0.23	0.35%

gree of differentiation to eliminate the trend, and q – order of the moving average model. The ARIMA (p, D, q) mathematical formula can be described as follows [34]:

(2)

where L is the lag operator, φ_i are the parameters of the autoregressive part of the model, θ_i are the parameters of the moving average part, ε_t are the error terms.

The second step involves calculating the weights used to correct the baseline forecast. The value of the weights is calculated for each forecasted value to ensure that the data features (e.g., seasonal or trend components) are taken into account more accurately. The weighting factor is determined by calculating the Mean Absolute Error (MAE) for each period [35]:

(3)where Y_t denotes the original time

series, while $\hat{Y}_t$ represents the predicted time series generated by the model. The low values of these three indicators suggest minimal deviation between the actual data and the forecasted values

The final forecast is calculated as the product of the base forecast $\hat{Y}_{t,ARIMA}$ and the weighting factor w_t:

(4)

The LSTM method predicted the electric load based on the date, time, and kilowatt values. During the training of the LSTM method, the number of epochs varied from 100 to 400, which indicates the speed of model learning. Five different model variants were created with the number of neurons in the hidden layer: 200, 250, 300, 350, and 400 neurons. Each model was trained with the number of epochs 100, 200, 300, and 400, resulting in 20 different combinations of param-

e
t
e

LSTM error and LSTM weight. The final forecast is calculated as the product of LSTM error and LSTM weight, which increases the forecast accuracy. The final forecast combines the results of ARIMA and LSTM. For this purpose, a weighted average of forecasts was used, which allows for a more stable and accurate result by combining the advantages of each algorithm:

(5) where Y_{ARIMA} is an

ARIMA forecast, Y_{LSTM} is an LSTM forecast, w_{ARIMA} and w_{LSTM} are weights for each model determined based on their accuracy.

We applied the Random Forest method to improve forecast accuracy, which uses ARIMA and LSTM errors with the final forecast. Forecast accuracy metrics are used to assess the quality of the model: Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) [35]. These metrics help to determine how well the model performs and whether additional optimization is required.

After the model is trained and reaches the desired level of accuracy, the trained model parameters are saved. It ensures that the model can be used for future forecasts without re-training. Saving the model parameters is crucial for practical application in real-world conditions, particularly for operational estimates over time.

After evaluating the accuracy of the forecasts using metrics such as MAE and RMSE, we compare the error with a certain threshold. If the error exceeds this threshold, the forecast is passed to the ML model block for further correction and refinement. Otherwise, if the error is less than the threshold, it is passed to the load forecasting stage, which allows unnecessary corrections to be avoided, and a more accurate forecast for further load management in the networks is applied. This approach ensures precise and efficient use of the method in real-time, minimizing the impact of significant errors on the final forecast.

The final value of Y_{final}, obtained after combining all stages

o
f

t
h
e

electric load in the networks. This value is the basis for further load management operations and resource optimization in power grids.

4. Experimental analysis and discussion

This section presents the results of the study of changes in electricity load by average daily and average hourly values and assesses the efficiency of the forecast model. The analysis aims to identify characteristic patterns in electricity consumption, particularly seasonal and daily trends, and evaluate the accuracy of the forecasting model to ensure reliable energy management.

Figure 3 shows the average daily load from January to August 2016. There is significant variability in the load over the study period. The load is consistently high in the winter months (January-February), with peak values exceeding 2.0 kW. In the spring (March-May), a gradual decrease in the average daily load is recorded, reaching minimum values in April-May (~1.2 million kW). In the summer months (June-July), there is a sharp increase in the average load, indicating a seasonal rise in electricity demand.

Figure 4 shows the average hourly load during the day. There is an explicit daily load dynamics with a minimum at night (about 1.1 kW) and a gradual increase from 6:00 am to reach a maximum value (~1.7 kW) between 14:00 and 18:00. After the peak, the load gradually decreases until late in the evening.

Figure 5 shows a graph of the electric load for the last 20 hours, reflecting changes in the Actual and Predicted electricity consumption values. The graph demonstrates the cyclical nature of the load, with periodic peaks and troughs throughout the day. Several pronounced peaks in the electric load reach more than 24 kW and repeat with a certain regularity. The highest load occurs between 3-4 hours and 11-12 hours, possibly due to the characteristic periods of consumer activity.

The predicted load values correspond to the actual values, with minor deviations at specific points in time. The deviations are particularly noticeable during sharp changes in load when actual values show a rapid increase or decrease. Nevertheless, the general trend of Actual and Predicted values is almost identical, indicating the forecast model's high accuracy.

Periodic fluctuations in the load are visible, indicating a possible seasonal or daily dependence on electricity consumption. Forecasting this data type allows for more efficient resource allocation planning and regulation of power grids to avoid overloads during peak hours. The model's accuracy provides reliable results that ensure informed management decision-making in real-time.

Figure 6 shows a graph of the electric load for the last 20 hours, reflecting changes in the Actual and Predicted electricity consumption values. The graph demonstrates the cyclical nature of the load, with periodic peaks and troughs throughout the day. Several pronounced peaks in the electric load reach more than 24 kW and repeat with a certain regularity. The highest load occurs between 3–4 hours and 11–12 hours, possibly due to the characteristic periods of consumer activity.

The predicted load values correspond precisely to the actual values, with minor deviations at specific points in time. The deviations are particularly noticeable during sharp changes in load when actual values show a rapid increase or decrease. Nevertheless, both graphs' general trend and shape are almost identical, indicating the forecast model's high accuracy.

Periodic fluctuations in the load are visible, indicating a possible seasonal or daily dependence on electricity consumption. Forecasting this data type allows for more efficient resource allocation planning and regulation of power grids to avoid overloads during peak hours. The model's accuracy provides reliable results that ensure informed management decision-making in real-time.

Figure 7 shows the results of forecasting electricity consumption (kW) daily by hour using machine learning methods and the proposed method. The actual data serves as a benchmark for evaluating the forecasting accuracy of different models. The LSTM model can reproduce the general trend of electricity consumption but has some deviations in peak load values. The combined ARIMA-LSTM approach improves forecast accuracy by synergizing the capabilities of the models: ARIMA effectively processes linear time series trends, while LSTM considers their nonlinear components. Subsequent adjustment of forecasts using Random Forest helps to minimize residual errors by processing unaccounted-for trends in the data. The proposed ARIMA-LSTM-Random Forest method provides the highest correspondence to actual values, especially in periods with significant peak loads, confirming its high efficiency in short-term electricity consumption forecasting.

Figure 8 shows the results of forecasting the electric load for the next hour over 7 days using the ARIMA-LSTM-Random Forest method. The actual values of the electric load demonstrate a pronounced periodicity with daily fluctuations, where the maximums are observed during peak hours, and the minimums are observed during off-peak periods. Predicted values obtained using the ARIMA-LSTM-Random Forest method accurately reproduce trends and cyclical changes in load. At the same time, forecast errors are observed at sharp changes or anomalies, indicating the method's limitations when processing nonlinear fluctuations. The application of the ARIMA-LSTM-Random Forest method shows efficiency for forecasting periodic time series but requires optimization to improve accuracy in cases of significant load fluctuations.

Fig. 9 shows the results of experiment 2, in which the forecast of the electric load for the next month was made based on time series for the previous 7 days. The proposed ARIMA-LSTM-Random Forest method demonstrates a high correlation between actual and predicted values over time. The forecast reflects seasonal load fluctuations, mainly repeating cyclic trends with peaks and troughs. The method effectively considers the periodicity and data dynamics inherent in time series. The predicted values are very close to the actual values at the peak points of the electric load (about 24 kW). In periods of minimum load (about 10-12 kW), the method demonstrates a high level of correspondence between actual and predicted values, but there are slight deviations. Over the entire time interval (more than 250 hours), it is clear that the model remains stable in its forecasting. It indicates that combining ARIMA for modeling linear dependencies and LSTM for accounting for nonlinearities allows for a balance between short-term and long-term forecasting.

Figure 10 shows the results of experiment 3, which demonstrates the prediction of the electric load for the next hour based on the time series for the previous two days. The graph compares the actual load values with the predicted values obtained using the ARIMA-LSTM-Random Forest method. It predicts peak values and load decreases at different times of the day. Slight deviations between the actual and predicted values are noticeable in the peak load zones (around 17 kW). It indicates that the model does a good job of predicting critical moments, although some errors may be caused by complex nonlinear dependencies in the data. The ARIMA-LSTM-Random Forest method efficiently predicts electricity load, especially for short-term time series. Using two days of data allows the model to consider the main trends and periodicity in the load, which is essential for short-term forecasting tasks.

Figure 11 shows the electric load forecast for the next hour for one day using the ARIMA-LSTM-Random Forest method. The horizontal axis (X) indicates the time in hours, starting from midnight (00:00) and ending with the last hour of the day (23:00). The vertical axis (Y) shows the value of the electrical load in kilowatts (kW) in the range from 40 to 60. Analysis of the curve shows a pronounced periodicity, which indicates the cyclical nature of electricity consumption during the day. Peak load values of approximately 60 kW are observed in the morning (around 02:00), afternoon (14:00), and evening (22:00). In contrast, minimum values of around 40 kW are typical for the periods after 04:00, in the afternoon (16:00) and at night (18:00). This forecast reflects the behavioral characteristics of electricity consumption, indicating peak periods that may be associated with people's daily routine activities, such as the use of household appliances or the operation of industrial facilities. The graph data can help plan energy resources and optimize their daily distribution.

Table 1 shows the results of the forecasting quality assessment of the three methods: LSTM Forecast, the combined ARIMA-LSTM model, and the improved ARIMA-LSTM-Random Forest model.

The metrics Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE) are used to quantify the accuracy of the forecasts. The LSTM-based model showed average forecasting performance, with MSE = 0.38 and RMSE = 0.62. The MAPE value of 8.37% indicates a relatively high average relative error. It demonstrates that the LSTM, despite its power in modeling time series, could have achieved optimal results for the available data. Combining ARIMA with LSTM improved the mean absolute error (MAPE = 1.27%) compared to the pure LSTM, but the MSE = 0.63 and RMSE = 0.79 are higher. It indicates that the combination of models reduced the relative error (MAPE), but the total squared error increased. The proposed ARIMA-LSTM-Random Forest method significantly improved the results: MSE = 0.27 – the lowest value among all models. RMSE = 0.23 – significantly reduced root mean square error. MAPE = 0.35% – the lowest average relative error, indicating high forecasting accuracy. These results demonstrate the effectiveness of ensemble approaches, where Random Forest adaptively corrects the forecasts of the combined ARIMA-LSTM model. The method minimized absolute and relative errors, confirming its advantage for this forecasting task.

5. Conclusions

The experimental study results confirm the proposed approach's effectiveness for predicting electric load based on the combined ARIMA-LSTM-Random Forest model. The analysis of the forecasting quality using the MSE, RMSE, and MAPE metrics shows a significant improvement in accuracy compared to individual LSTM and ARIMA-LSTM methods. In particular, the proposed model provided the lowest values of the mean square error (MSE = 0.27), root mean square error (RMSE = 0.23), and mean absolute percentage error (MAPE = 0.35%), which is evidence of high accuracy of forecasts.

The combination of approaches allows for considering linear dependencies (ARIMA model) and nonlinear data characteristics (LSTM). At the same time, using Random Forest ensures the adaptive correction of forecasts and the minimization of errors. The use of data for the previous 7 days made it possible to consider seasonal trends and fluctuations, which is critical for medium-term forecasting tasks in the energy sector.

Thus, the proposed ARIMA-LSTM-Random Forest model demonstrates high accuracy and adaptability, making it promising for practical use in energy monitoring and managing power grids. Combining the advantages of each model as part of an ensemble approach allows for compensating for their shortcomings, ensuring the reliability of forecasts and the possibility of effective resource planning.

Literatura – References

1. Ye, H., Yang, B., Han, Y., & Chen, N. (2022). State-Of-The-Art Solar Energy Forecasting Approaches: Critical Potentials and Challenges. Frontiers in Energy Research, 10, 268. https://doi.org/10.3389/fenrg.2022.875790.
2. Kashtan, V., Hnatushenko, V. (2023). Deep Learning Technology for Automatic Burned Area Extraction Using Satellite High Spatial Resolution Images. In: Babichev, S., Lytvynenko, V. (eds) Lecture Notes in Data Engineering, Computational Intelligence, and Decision Making. ISDMCI 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 149. Springer, Cham. https://doi.org/10.1007/978-3-031-16203-9_37.
3. Kashtan, V., Hnatushenko, V. (2023). Automated building damage detection on digital imagery using machine learning. Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, № 6. P.134-140. https://doi.org/10.33271/nvngu/2023-6/134.
4. Kyrylenko O. (2022). Measures and Ways of Transforming Ukraine's Energy Sector into an Intelligent Environmentally Friendly System. Visnyk Natsionalnoi Akademii Nauk Ukrainy. No 3. PP. 18-23. https://doi.org/10.15407/visn2022.03.018.
5. Kyrylenko, O., Shcherba, A., Kucheriava, I. (2021). Intellectual technologies for monitoring of technical state of up-to-date high-voltage cable power lines. Tekhnichna Elektrodynamika. 2021. 29-40. https://doi.org/10.15407/techned2021.06.029 .
6. Psyuk, V., & Polyanska, A. (2024). The usege of artificial intelligence in the activities of mining enterprises. E3S Web of Conferences, 526, 01016. https://doi.org/10.1051/e3sconf/202452601016
7. Polyanska, A., Pazynich, Y., Petinova, O., Nesterova, O., Mykytiuk, N., & Bodnar, G. (2024). Formation of a Culture of Frugal Energy Consumption in the Context of Social Security, pp. 60-87. The Journal of the International Committee for the History of Technology, 29(2), 60–87. https://doi.org/10.11590/icon.2024.2.03
8. Pavlychenko, A., Sala, D., Pyzalski, M., Dybrin, S., Antoniuk, O., & Dychkovskyi, R. (2025). Utilizing Fuel and Energy Sector Waste as Thermal Insulation Materials for Technical Buildings. Energies, 18(9), 2339. https://doi.org/10.3390/en18092339
9. Lewicka, D. (2020). Employee institutional trust as an antecedent of diverse dimensions of organisational commitment. Argumenta Oeconomica, 2019(1), 321–340. https://doi.org/10.15611/aoe.2020.1.13
10. Dychkovskyi, R.O., Tymoshenko, Y.V., Astafiev, D.O. (2014). Method of analytical investigation of wall advance speed and forms of line face influence on stress-strain state of a rock massif. Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, (1), 11–16
11. Pivniak, H., Aziukovskyi, O., Papaika, Yu., Lutsenko, I., Neuberger, N. (2022). Problems of development of innovative power supply systems of Ukraine in the context of European integration. Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, (5), 89-103. https://doi.org/10.33271/nvngu/2022-5/089.
12. Sytnyk, R., Hnatushenko, V., Hnatushenko, V. (2022). Decentralized Information System for Supply Chain Management Using Blockchain. IntelITSIS'2022: 3nd International Workshop on Intelligent Information Technologies and Systems of Information Security, Khmelnytskyi, Ukraine. http://ceur-ws.org/Vol-3156/paper45.pdf .
13. Smyl, S. (2020). A hybrid method of exponential smoothing and recurrent neural networks for time series forecasting. Int. J. Forecast. 36, 75–85. https://doi.org/10.1016/j.ijforecast.2019.03.017.
14. Kim, Y., Son, H.-g., and Kim, S. (2019). Short term electricity load forecasting for institutional buildings. Energy Rep. 5, 1270–1280. https://doi.org/10.1016/j.egyr.2019.08.086.
15. Aasim, Singh, S. N., Mohapatra, A (2021). Data driven day-ahead electrical load forecasting through repeated wavelet transform assisted SVM model. Appl. Sof Comput. 111, 107730. https://doi.org/10.1016/j.asoc.2021.107730.
16. Li, Y., Jia, Y., Li, L., Hao, J., and Zhang, X. (2020). Short term power load forecasting based on a stochastic forest algorithm. power Syst. Prot. control 48, 117–124. https://doi.org/10.1109/ICVRIS51417.2020.00202.
17. Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., et al. (2021a). Short-term load forecasting of industrial customers based on SVMD and XGBOOST. Int. J. Electr. Power and Energy Syst. 129, 106830. https://doi.org/10.1016/j.ijepes.2021.106830.
18. Dai, Y., Zhou, Q., Leng, M., Yang, X., and Wang, Y. (2022). Improving the bi-LSTM model with xgboost and attention mechanism: A combined approach for short-term power load prediction. Appl. Sof Comput. 130, 109632. https://doi.org/10.1016/j.asoc.2022.109632.
19. Mohammed, N. A., Al-Bazi, A. (2022). An adaptive backpropagation algorithm for long-term electricity load forecasting. Neural Comput. Appl. 34, 477–491. doi:10.1007/s00521-021-06384-x
20. Pavićević, M., Popović, T. (2022). Forecasting day-ahead electricity metrics with artificial neural networks. Sensors 22, 1051. https://doi.org/10.3390/s22031051.
21. Qiao, W. (2018). Intelligent pre-estimation method study on short-period power load. Jiangsu Electr. J., 13–17. https://doi.org/10.3390/en13226105.
22. Kong, W, Dong, Z,, Jia, Y, Hill, D,, Xu, Y., Zhang Y. (2017). Short-term residential load forecasting based on LSTM recurrent neural network. IEEE transactions on smart grid 2017; 10(1): 841–851.

23. Tan, M., Yuan, S., Li, S., Su, Y., Li, H., He, F. (2019). Ultra-short-term industrial power demand forecasting using LSTM based hybrid ensemble learning. IEEE transactions on power systems 2019; 35(4): 2937–2948.
24. Bouktif, S., Fiaz, A., Ouni, A., Serhani, M. (2018). Optimal deep learning lstm model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches. Energies; 11(7): 1636.
25. Tsegaye, S., Sanjeevikumar, P., Tjernberg, L., Fante, K. (2024) Short term load forecasting of electrical power distribution system using enhanced deep neural networks (DNNs). IEEE Access.
26. Mehra, A., Spreeuwers, L., & Strisciuglio, N. (2021). Deepfake Detection using Capsule Networks and Long Short-Term Memory Networks. Proceedings of the 16th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, 407–414. https://doi.org/10.5220/0010289004070414
27. Wencheng, L., & Zhizhong, M. (2024). Short-Term Photovoltaic Power Forecasting Using Enhanced Bidirectional Long Short-Term Memory Networks. 2024 36th Chinese Control and Decision Conference (CCDC), 3308–3312. https://doi.org/10.1109/ccdc62350.2024.10587916
28. Dychkovskyi, R., Falshtynskyi, V., Ruskykh, V., Cabana, E., & Kosobokov, O. (2018). A modern vision of simulation modelling in mining and near mining activity. E3S Web of Conferences, 60, 00014. https://doi.org/10.1051/e3sconf/20186000014
29. Graw, J. H., Wood, W. T., & Phrampus, B. J. (2021). Predicting Global Marine Sediment Density Using the Random Forest Regressor Machine Learning Algorithm. Journal of Geophysical Research: Solid Earth, 126(1). https://doi.org/10.1029/2020jb020135
30. Pazynich, Y., Kolb, A., Korcyl, A., Buketov, V., & Petinova, O. (2024). Mathematical model and characteristics of dynamic modes for managing the asynchronous motors at voltage asymmetry. Polityka Energetyczna – Energy Policy Journal, 27(4), 39–58. https://doi.org/10.33223/epj/191779
31. Toronto Hydro website. https://www.torontohydro.com/SITES/ELECTRICSYSTEM/BUSINESS/YOURBILLOVERVIEW/NETSYSTEMLOADSHAPE/Pages/default.aspx
32. Kholodniak, Y., Havrylenko, Y., Halko, S., Hnatushenko, V., Suprun, O., Volina, T., Shchur, T. (2023). Improvement of the algorithm for setting the characteristics of interpolation monotone curve. Informatyka, Automatyka, Pomiary W Gospodarce I Ochronie Środowiska, 13(4), 44–50. https://doi.org/10.35784/iapgos.5392 .
33. Ruiyan, Z., Xingchen Z. (2024). Short-term power load forecasting based on ARIMA-LSTM. J. Phys.: Conf. Ser. 2803 012002. https://doi.org/10.1088/1742-6596/2803/1/012002.
34. Dongyan, F., Hai, S., Jun, Y., Kai, Z., Xia, Y., Zhixue, S. (2021). Well production forecasting based on ARIMA-LSTM model considering manual operations, Energy, Volume 220, 119708, https://doi.org/10.1016/j.energy.2020.119708.
35. Kasemset, C., Phuruan, K.t, Opassuwan, T. (2023). Shallot Price Forecasting Models: Comparison among Various Techniques. Production Engineering Archives. 29. 348-355. https://doi.org/10.30657/pea.2023.29.40.
36. Bareth, R., Yadav, A., Gupta, S., Pazoki, M. (2024). Daily average load demand forecasting using LSTM model based on historical load trends. IET Gener. Transm. Distrib. 18, 952–962. https://doi.org/10.1049/gtd2.13132
37. Sun, F., Huo, Y., Fu, L., Liu, H., Wang, X., Ma, Y. (2023). Load-forecasting method for IES based on LSTM and dynamic similar days with multi-features. Global Energy Interconnect. 6(3), 285–296. https://doi.org/10.1016/j.gloei.2023.06.003.

Inteligentna metoda prognozowania obciążenia energią elektryczną z wykorzystaniem ARIMA-LSTM-Random Forest

Niestabilność systemów energetycznych, spowodowana wewnętrznymi czynnikami ekonomicznymi i wyzwaniami zewnętrznymi, w tym konfliktami geopolitycznymi, znacząco komplikuje proces planowania i zarządzania zasobami energetycznymi. Niezbędnym narzędziem wdrażania działań na rzecz oszczędności energii jest wprowadzenie do sektora energetycznego nowoczesnych technologii komputerowych, w tym systemów sztucznej inteligencji. Technologie inteligentne umożliwiają wykorzystanie metod prognozowania obciążenia elektrycznego, w tym algorytmów sztucznej inteligencji. W niniejszym artykule zaproponowano połączony model ARIMA-LSTM-Random Forest do prognozowania obciążenia elektrycznego. Połączenie tych podejść pozwala na uwzględnienie zarównoliniowych, jak i nieliniowych zależności w danych, co jest kluczowe dla zapewnienia dokładności prognoz. Wykorzystanie danychz ostatnich siedmiu dni dostarcza wystarczających informacji do identyfikacji trendów i wahań sezonowych, co czyni to obiecu-jącą perspektywą dla prognozowania średnioterminowego w zadaniach monitorowania energii. Zatem połączenie metod ARIMA,LSTM i Random Forest pozwala osiągnąć wysoką dokładność prognozowania zużycia energii elektrycznej. Proponowane podejściejest rozwiązaniem optymalnym, ponieważ łączy zalety każdego modelu i kompensuje ich wady. Zaproponowana metoda ARIMA--LSTM-Random Forest znacząco poprawiła wyniki: MSE = 0,27, RMSE = 0,23, MAPE = 0,35%. Metoda zminimalizowała błędybezwzględne i względne, co potwierdza jej przewagę w tym zadaniu prognostycznym. Wyniki są obiecujące pod kątem praktycznychzastosowań w zarządzaniu obciążeniem sieci elektroenergetycznych.

Słowa kluczowe: *prognozowanie obciążenia energią elektryczną, metoda inteligentna, ARIMA, LSTM, las losowy*