INNOVATIVE MECHANICAL ENGINEERING **ISSN 2812-9229 (Online)**
University of Niš, Faculty of Mechanical Engineering
VOL. 3, NO 3, 2024, PP. 31-47

Original scientific paper *

MACHINE LEARNING APPROACH FOR ENERGY CONSUMPTION ESTIMATION IN DISTRICT HEATING SYSTEMS

Ivan Ćirić[1], Milica Tasić[1], Marko Ignjatović[1], Dušan Stojiljković[1], Milica Ćirić[2]

[1]Faculty of Mechanical Engineering in Niš, University of Niš, St. Aleksandra Medvedeva 14, 18000 Niš, Serbia

[2]Faculty of Civil Engineering and Architecture in Niš, University of Niš, St. Aleksandra Medvedeva 14, 18000 Niš, Serbia

Abstract. *This paper explores the application of various machine learning methods for time series estimation in district heating systems. The focus is on predicting heat load using supervised machine learning techniques, such as artificial neural networks and the Random Forest algorithm. Input parameters are derived from the SCADA system, including outdoor air temperature, water temperature, and pressure in the primary and secondary circuits. The development of advanced predictive models enables a better analysis of energy consumption patterns, which is crucial for improving the efficiency of district heating systems. The paper examines different machine learning approaches to identify the models that best meet the requirements for prediction in real operating conditions. The expected contribution of this study lies in establishing a foundation for further automation and optimization of district heating systems, potentially leading to reduced operational costs and energy consumption, as well as enhancing the sustainability of energy systems.*

Key words: *Machine Learning, District Heating Systems, Artificial Neural Networks, Random Forest Algorithm, SCADA Systems*

1. INTRODUCTION

The development of artificial intelligence and machine learning (ML) has brought significant changes to many industrial sectors, including the energy sector. The application of ML techniques primarily enables more accurate predictions, followed by the

*Received: December 19, 2024 / Accepted December 27, 2024.
Corresponding author: Milica Tasić
Institution: Faculty of Mechanical Engineering University of Niš
E-mail: milica.tasic@outlook.com

optimization and automation of complex processes. In recent years, the application of various machine learning models has become a key component in improving energy management systems, allowing intelligent real-time adjustments of operational parameters [1,2]. Some of these methods have proven to be particularly effective in systems that rely on large amounts of data, as they can analyze complex relationships between variables and provide highly accurate predictions. Although the optimization of energy systems, such as district heating, has traditionally relied on static models and manual adjustments, it is believed that the application of machine learning can significantly improve this field [2]. Therefore, in this study, ANNs were used to predict the thermal load in the district heating system at the Faculty of Mechanical Engineering in Niš, as they are capable of recognizing patterns in data that are often too complex for traditional methods [1,2]. In addition to artificial neural networks, Random Forest analysis was also used to predict the same energy indicator as another machine learning approach [1]. Random Forest analysis, like neural networks, can recognize complex patterns in the data but has the advantage of being less prone to overfitting and more resistant to parameter adjustments. The existence of the SCADA system at the heating plant of the Faculty of Mechanical Engineering in Niš enables continuous monitoring of the district heating system's operation, as well as data collection on energy performance. This data provides a basis for applying machine learning methods to predict any energy indicator. However, current operational decisions are made manually, and the application of one of these methods provides the opportunity to transition to automated, optimized management strategies. Their integration into control processes can not only improve system efficiency but also reduce operational costs and fossil fuel consumption. In this way, the system not only becomes more transparent but also enables data-driven decision-making that is easier for operators and relevant decision-makers to understand. In the long term, this approach can lead to improvements in the overall heating system efficiency, enhanced sustainability, and optimized resource use within the energy sector.

2. METHODOLOGY

Research in the field of energy consumption estimation (e.g., thermal, electrical, cooling) in buildings, as well as in the prediction of heat demand in district heating systems (DHS), can be categorized into two approaches: the knowledge-based approach (also known as "forward," classical, or expert rules- based) and the data-driven approach (inverse) [3]. The knowledge-based approach relies on physical models and mathematical equations that describe system behavior, enabling predictions based on known principles. On the other hand, the data-driven approach utilizes historical data and applies techniques such as regression models and machine learning algorithms to forecast consumption. While the knowledge-based approach requires deep understanding of the physical characteristics of the system, the data-driven approach focuses on identifying patterns in large datasets, which also means that the system or method can improve its accuracy over time as new data becomes available [4].

Using data from the heating system of the Faculty of Mechanical Engineering in Niš, this research focuses on the data-driven approach, allowing for the analysis of past energy consumption behaviors and external conditions to develop models that can more accurately predict future thermal energy needs. Artificial Neural Networks (ANN) are central to the data-driven approach and are widely applied in energy consumption prediction. Fahlman and Lebiere in [5] introduced the Cascade-Correlation Neural Network (CCNN). CCNN functions by constructing itself during training, adding new hidden units one by one, which reduces the need for manual tuning and improves convergence.

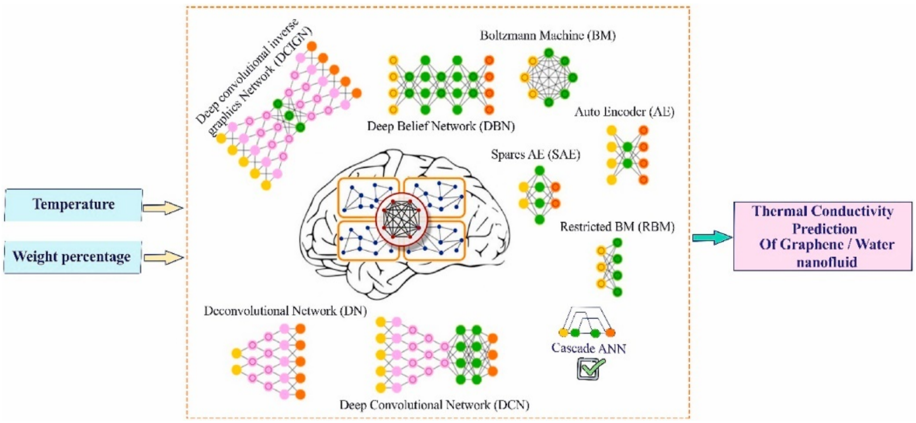


Fig. 1. Example of Cascade Neural Network Architecture [6]

In the work of Adya and Collopy [7], the efficiency of neural networks in business forecasting was evaluated, comparing different architectures and their predictive accuracy. This paper also highlighted the role of cascade and recurrent structures in time series tasks. Zhang et al. [8] proposed an adaptive cascade neural network for human action recognition based on skeletal data using two adaptive neural networks known as VA-RNN and VA-CNN. These networks are based on convolutional neural networks and recurrent neural networks with long short-term memory. Wu and Pan introduced a modular convolutional neural network (CNN) model in [9] that addresses issues such as overfitting of model parameters and slow convergence in large-scale models. The proposed model integrates multiple modules in parallel and includes a control unit to optimize the model structure and reduce computation costs, achieving high accuracy in image recognition tasks across multiple datasets. Chandra and Zhang [10] presented a cooperative coevolutionary approach for time series estimation using Elman recurrent neural networks (RNN), demonstrating the flexibility of combining evolutionary computation with recurrent structures for long-term forecasting. Elman [11] was the first to introduce the basic structure of the RNN, which can learn patterns using feedback loops to retain information over time.

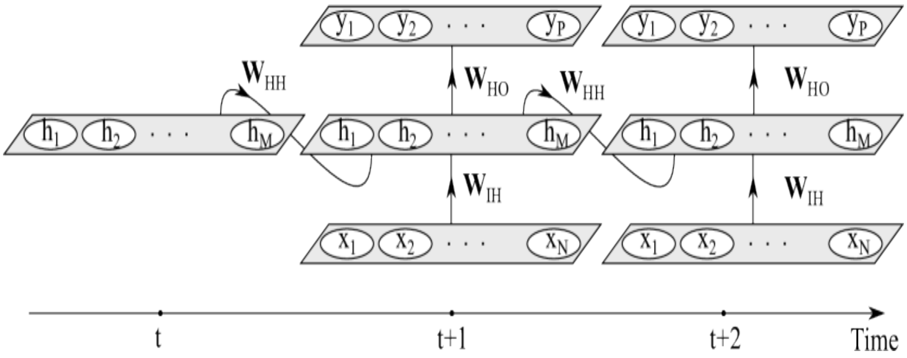


Fig. 2: Example of Recurrent Neural Network [12]

Hochreiter and Schmidhuber solved the vanishing gradient problem by introducing Long Short-Term Memory (LSTM) networks in [13], which use mechanisms for selectively retaining and forgetting information. Graves and Schmidhuber [14] further refined LSTM networks by creating Bidirectional LSTMs, which process data both forwards and backwards, allowing superior performance in classification tasks by capturing both past and future context in speech data. Vaswani et al. eliminated the recurrent nature of RNNs by relying on attention mechanisms in [15], significantly improving efficiency in machine translation and laying the foundation for future models such as BERT and GPT. Lipton [16] reviewed various RNN variations, including LSTM and GRU, which have been used in applications such as language modeling and time series forecasting. Paper [17] introduced a two-stage attention-based RNN for time series forecasting, successfully integrating recurrent structures with attention mechanisms. Lai et al. [18] introduced a Recurrent Convolutional Neural Network for traffic forecasting, which integrates both convolutional and recurrent layers.

In addition to artificial neural networks, another less exploited machine learning algorithm was applied in this research. As noted by various empirical studies [19-23], the random forest algorithm has proven to be a serious competitor to state-of-the-art machine learning methods such as boosting, according to [24], and support vector machines, according to [25]. These methods are fast, easy to implement, produce highly accurate predictions, and can handle a very large number of input variables without overfitting. In a series of papers and technical reports, Breiman demonstrated that significant improvements in classification and regression accuracy can be achieved using ensembles of trees, where each tree in the ensemble is formed according to a random parameter. Final predictions are obtained by aggregating across the ensemble. Since the basic elements of the ensemble are tree-structured predictors, and each of these trees is constructed using random variations, these procedures are called "random forests" [26].

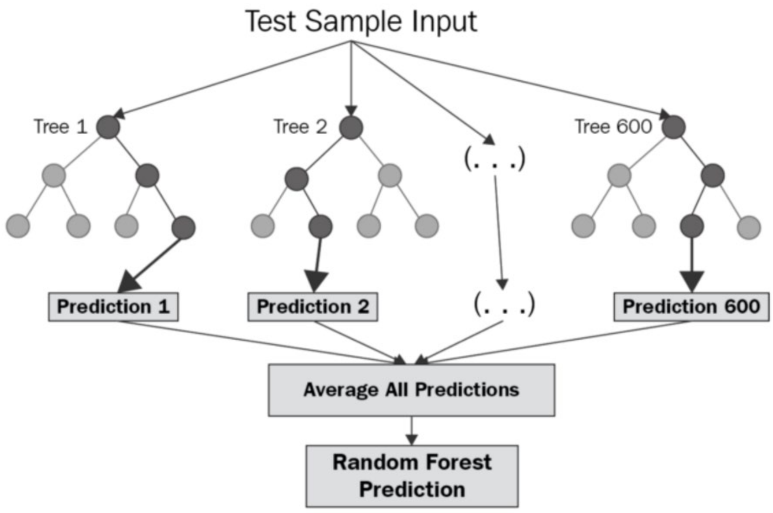


Fig. 3: Example of Random Forest Algorithm [26]

In Breiman's approach, each tree in the collection is formed by randomly selecting a small group of input coordinates (which are further referred to as features or variables) at each node, based on which the split is performed. Then, the best split is calculated based on these features within the training dataset. The tree is built using the CART methodology [27] to its maximum size, without pruning. This scheme of random subspace selection is combined with bagging [28-30], where each time a new individual tree is grown, the training dataset is resampled with replacement. The performance of Cascade-Correlation Neural Networks (CCNN), Recurrent Neural Networks (RNN), and even the random forest algorithm is evaluated using metrics that quantitatively assess model accuracy, particularly in time series estimation or sequential data classification tasks. Although there are various measures of prediction performance quality, the most important is the accuracy that can be achieved using the training data. However, a suitable accuracy measure for a given problem is not universally accepted by scientists and practitioners. The accuracy measure is often defined through prediction error, which represents the difference between the actual (desired) and predicted values. There are several such prediction accuracy measures found in the literature, each with its own advantages and limitations.

This study will provide an overview of the performance of CCNN, RNN, and the random forest algorithm in the context of time series and classification tasks, using a dataset of 1200 energy indicators obtained from the SCADA system of the Faculty of Mechanical Engineering heating plant in Niš, employing the following standard evaluation metrics:

1. Mean Squared Error (MSE)

$$\mathrm{MSE} = \frac{1}{n} * \sum_{i=1}^{n} (yi - \widehat{yi})^2 \tag{1}$$

where yi is the actual value, and $\widehat{yi}$ is the predicted value. MSE is a popular metric in time series predictions because it places greater emphasis on larger errors. It is used to evaluate the performance of both CCNN and RNN models in tasks such as time series and business forecasting [31].

2. Mean Absolute Error (MAE)

$$\mathrm{MAE} = \frac{1}{n} * \sum_{i=1}^{n} |yi - \hat{y}i| \tag{2}$$

MAE is useful in cases where larger errors are less significant, as it avoids squaring the deviations and better reflects the average errors in predictions. This metric is often used in time series [32].

3. Root Mean Squared Error (RMSE)

$$\mathrm{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (yi - \hat{y}i)^2} \tag{3}$$

RMSE returns errors in the same units as the original values, making it useful for interpreting results in time series and predictions [33].

4. Coefficient of Determination (R^2)

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (Xi - Yi)^2}{\sum_{i=1}^{m} (\overline{Y} - Yi)^2} \tag{4}$$

where Yi is the actual (observed) values of the dependent variable, Xi is the predicted values of the dependent variable (from the model) and $\overline{Y}$ is the average value of the actual dependent variable. An R^2 value closer to 1 indicates better model performance. This metric is particularly useful in analyzing complex predictions in time series [34].

5. Mean Absolute Percentage Error (MAPE)

$$\mathrm{MAPE} = \frac{1}{n} \sum \left| \frac{et}{yt} \right| * 100 \tag{5}$$

2.1 Control and data acquisition system in district heating

The SCADA system continuously monitors the temperature, flow, and pressure of water in both the primary and secondary circuits. Based on this data, along with the outside air temperature, the operation of boilers, pumps, and valves is automatically adjusted to optimize energy consumption and maintain a stable supply water temperature. The total demand for district heating, or thermal load, corresponds to the difference between the supply and return water temperatures measured at the plant, adjusted for distribution losses. The automated control of the district heating system at the Faculty of Mechanical Engineering in Niš is based on a parametric function of the outside air temperature offset, often referred to as the control curve with two setpoints [35].

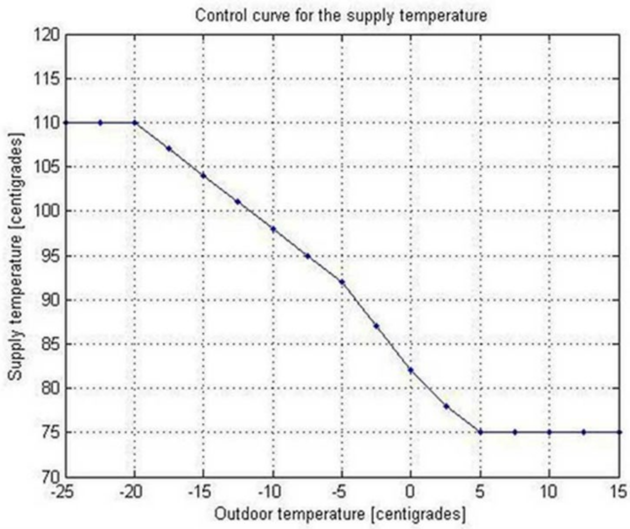


Fig. 4: Example of control curve

The logic of the control curve is implemented in the PLC controller, which calculates the water temperature and flow at the corresponding setpoints and transmits the obtained values to the PID controller. The function of the PID controller is to maintain the output parameters, temperature, and flow, close to the target value determined by the control curve. The PID controller uses proportional, integral, and differential gains to minimize the error between the actual and set values, thereby ensuring stable and precise regulation.

The mathematical model of the PID controller in the general case is given by:

$$\mathrm{u(t)} = K_{\mathrm{p}}\mathrm{e(t)} + Ki\int_0^{\mathrm{t}} \mathrm{e}(\tau)\mathrm{d}\tau + Kd\,\frac{\mathrm{de(t)}}{\mathrm{dt}} \tag{6}$$

The parameters *Ki, Kd* and *Kp* represent the integral, differential, and proportional gains, respectively. Control in the heat substations of the district heating system at the Faculty of Mechanical Engineering in Niš is carried out through local control valves and sensors, which monitor outdoor temperature conditions and water temperature in the primary and secondary circuits. The measured data is then recorded in the SCADA system and organized into a time series, enabling further analysis and time series estimation for system modeling. The measured data in the time series represent a sequence of observations organized according to the time at which the measurements were taken [36, 37].

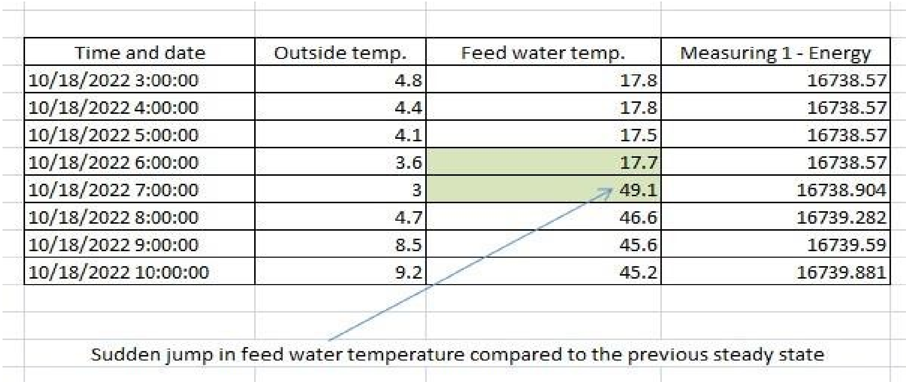
For time series prediction, two machine learning approaches were used: Cascade Neural Networks (CCNN) and Recurrent Neural Networks (RNN), along with the Random Forest algorithm.

The inputs were:

1. Time and date (format: date/time)
2. Outdoor air temperature (format: number)
3. Supply water temperature (format: number)

The output being predicted was the consumed thermal energy (format: number).

All the data used for processing in the Neural Networks (NN) was normalized beforehand. Due to the presence of extreme values in the data, standard normalization methods did not yield sufficiently good results, so robust normalization was applied for data transformation, which uses the median and interquartile range (IQR) instead of the mean and standard deviation.



Time and date	Outside temp.	Feed water temp.	Measuring 1 - Energy
10/18/2022 3:00:00	4.8	17.8	16738.57
10/18/2022 4:00:00	4.4	17.8	16738.57
10/18/2022 5:00:00	4.1	17.5	16738.57
10/18/2022 6:00:00	3.6	17.7	16738.57
10/18/2022 7:00:00	3	49.1	16738.904
10/18/2022 8:00:00	4.7	46.6	16739.282
10/18/2022 9:00:00	8.5	45.6	16739.59
10/18/2022 10:00:00	9.2	45.2	16739.881

Sudden jump in feed water temperature compared to the previous steady state

Fig. 5: Data set from SCADA system

Extreme values in the data were observed in the supply water temperature, and they occurred when the supply water temperature increased due to a decrease in outdoor air temperature, leading to an increased demand for more heat in end-user buildings, as well as when the boilers were started in the early morning hours, usually at 5 AM, 6 AM, or 7 AM, compared to the night period when the boilers were in standby mode. This data transformation process ensured that extreme values did not overly affect the data distribution, which improved the performance of machine learning algorithms. The neural network models were trained with 80% of the data, while 20% was used for testing. The neural networks used in the study were the cascade network and the Elman recurrent network, and they were trained using the Levenberg-Marquardt algorithm on the same dataset. There was no need for prior data preparation for the application of Random Forest analysis, as this method produced very accurate predictions using the data in its original form. The number of decision tree models was 100, and the final prediction was calculated as the average value of all previous predictions. The prediction process used data randomization and variable randomization, which are automatic parts of the Random Forest algorithm, as well as cross-validation, which was chosen to evaluate model performance to ensure robustness and generalization of the model to unseen data. These two concepts were not used in the neural network algorithms. Given that the district heating system at the Faculty of Mechanical Engineering in Niš is connected to 12 heat substations, of which 5 substations are in the Nikola Tesla residential area, for the purposes of this research, data was used for the period from October 15 to October 31, 2022, in hourly intervals, only from the heat substations in this neighborhood. The goal was to make general assumptions about which neural network model (NNM) predicted the actual input data with the least error, viewed from all perspectives considered in this research, and to compare the obtained results with the results of the analysis using the Random Forest method.

Since the analysis using the Random Forest method yielded better prediction results in terms of the time required for prediction, simplicity of parameter tuning, and robustness to noise in the data, the algorithm was applied to predict the same output, consumed thermal energy, using three additional combinations of different input parameters:

Combination 1

The inputs were:

1. Outdoor air temperature (format: number)
2. Return water pressure in the secondary circuit (format: number)
3. Reference temperature (format: number)

Combination 2

The inputs were:

1. Supply water temperature (format: number)
2. Supply water pressure in the primary circuit (format: number)
3. Power at the source (format: number)

Combination 3

The inputs were:

1. Water pressure in the primary circuit (format: number)
2. Return water pressure in the secondary circuit (format: number)
3. Outdoor air temperature (format: number)

3. RESULTS AND DISCUSSION

The energy indicators used as inputs in the neural network (NN) and Random Forest analysis were selected based on the previously calculated correlation between all available data from the SCADA system and the output indicator that needed to be predicted, which is the consumed thermal energy.

The inputs were:

1. Time and date (format: date/time)
2. Outdoor air temperature (format: number)
3. Supply water temperature (format: number)

The output being predicted was the consumed thermal energy (format: number).

The correlation between the variables time and date (Var1), outside temperature (Var2), feed water temperature (Var3), and the dependent variable consumed thermal energy (Var4) was calculated to determine the strength and direction of the linear relationships between these variables. The graph displays the data dispersion for each variable in relation to consumed thermal energy, along with the corresponding correlation coefficients:

1. **Time and date (in seconds) vs consumed thermal energy**: The correlation coefficient between these variables is 0.0029778, indicating an almost non-existent linear correlation. This result suggests that changes in the time and date variable have no significant impact on the values of consumed thermal energy. The scatter of points on the graph confirms that there is no clear linear relationship between these variables. This means that the time and date variable does not significantly contribute to the

prediction of consumed thermal energy through linear analysis.

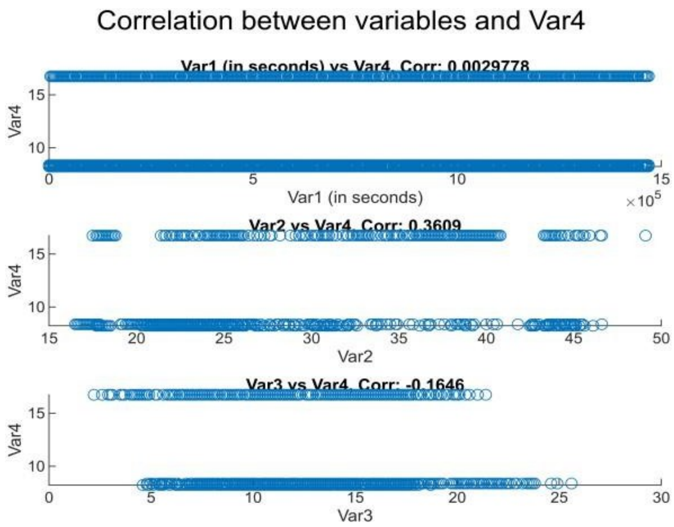


Fig. 6: Visualization of the correlations between input parameters and the output

2. **Outside temperature vs consumed thermal energy**: The correlation coefficient is 0.3609, indicating a weak to moderate positive correlation. Although there is a tendency for the value of consumed thermal energy to increase as outside temperature increases, the relationship between these variables is not strong enough to be considered decisive for prediction. This result implies that outside temperature has only a partial influence on consumed thermal energy, although there is a positive relationship.
3. **Feed water temperature vs consumed thermal energy**: The correlation coefficient is -0.1646, indicating a weak negative correlation. This means that as feed water temperature increases, the value of consumed thermal energy tends to decrease slightly, although the relationship is not strong. Negative correlation suggests an inverse dependency, but given the weak correlation value, this relationship has a limited impact on the prediction of consumed thermal energy.

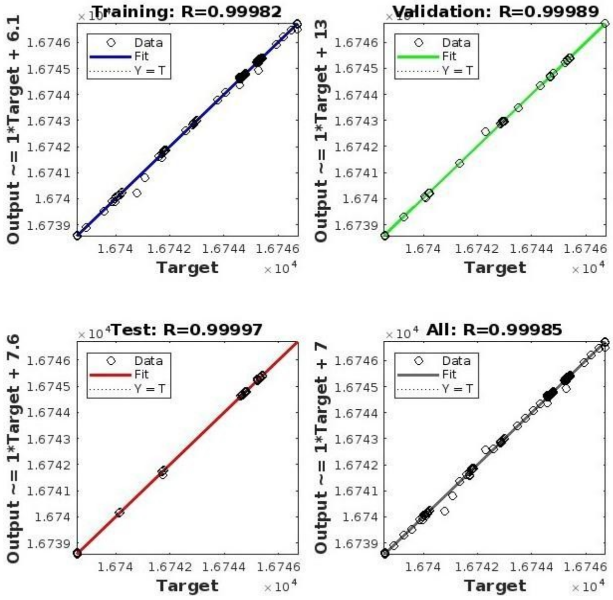


Fig. 7: Graphical representation of the prediction results of CNN

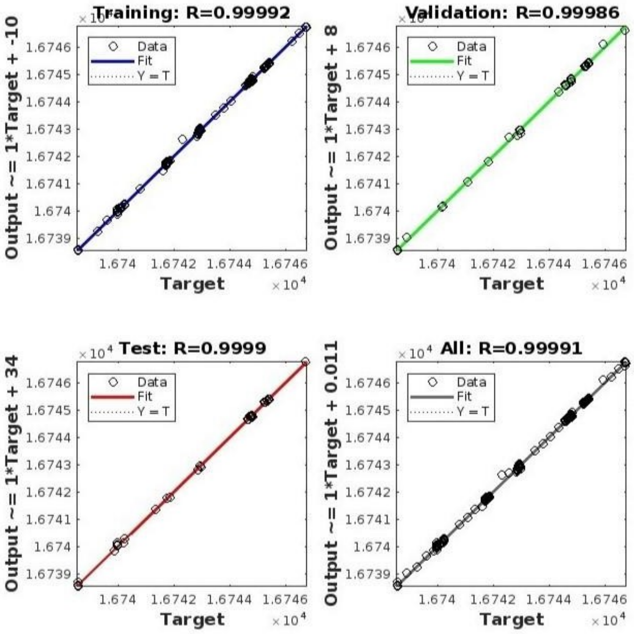


Fig. 8: Graphical representation of the prediction results of RNN

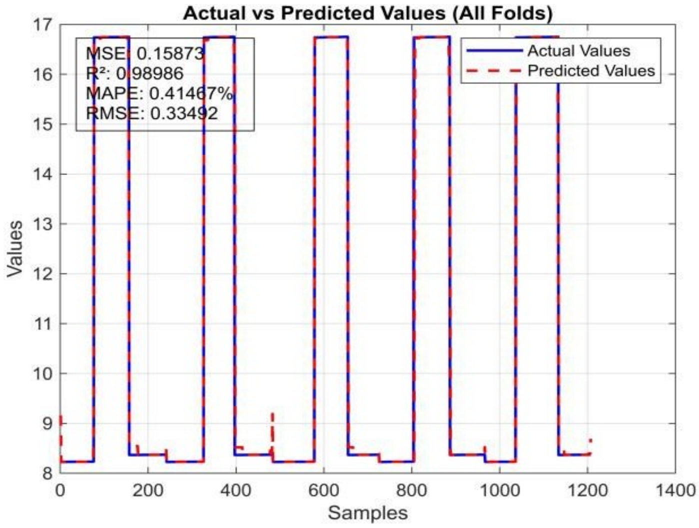


Fig. 9: Graphical representation of the prediction results of Random Forest

Table 1: Prediction results using ANN and Random Forest analysis

Machine learning approches	Number of epochs	Correlation coefficient R	Determination coefficient R^2	MSE	RMSE
CNN	113	0.99982	0.9964	0.001751	0.04185
RNN	277	0.9992	0.9997	0.0025635	0.05063
RFA	/	0.9926	0.9853	0.1668	0.33054

Table 1 and Figs. 7, 8 and 9 show the prediction results using the mentioned machine learning methods. All models show an extremely high correlation between predicted and actual values, but the Cascade Neural Network performs slightly better than the other two models with an R value of 0.99982, indicating a near-perfect linear relationship. The Recurrent Neural Network follows closely, while the Random Forest has a slightly lower correlation. If the coefficient of determination (R^2) is considered, the Recurrent Neural Network outperforms the others with the highest R^2, which shows how well the model explains the variance in the data. The Cascade Neural Network is very close, while Random Forest is still quite strong but lags behind the neural networks. Lower MSE indicates better performance. The Cascade Neural Network has the lowest MSE, meaning its predictions are closest to the actual values on average. The Recurrent Neural Network follows closely behind, but the Random Forest has a significantly higher MSE, indicating larger prediction errors. RMSE gives an intuitive sense of the prediction error in the same units as the data. Again, the Cascade Neural Network has the smallest error, followed by the

Recurrent Neural Network. The Random Forest shows the largest error, highlighting that it struggles more with precise predictions.

After that, the prediction of consumed thermal energy was performed using the Random Forest method on three different sets of input energy indicators:

Combination 1

The inputs were:

1. Outdoor air temperature (format: number)
2. Return water pressure in the secondary circuit (format: number)
3. Reference temperature (format: number)

Combination 2

The inputs were:

1. Supply water temperature (format: number)
2. Supply water pressure in the primary circuit (format: number)
3. Power at the source (format: number)

Combination 3

The inputs were:

1. Water pressure in the primary circuit (format: number)
2. Return water pressure in the secondary circuit (format: number)
3. Outdoor air temperature (format: number)

Figs. 10, 11 and 12 show the visualization of the correlations between the input combinations and the output, based on which the combinations were previously selected as relevant inputs.

Subsequently, Fig. 13 shows the prediction results of consumed thermal energy using the Random Forest method. The number of trees during training increased to 120 compared to the previous 100, which resulted in slightly more accurate prediction results.

Table 2: Prediction results using Random Forest analysis on three different combinations of input parameters

Random Forest method	Determination coefficient R^2	MSE	RMSE	MAPE (%)
Combination 1	0.98925	0.17193	0.3652	0.43378
Combination 2	0.98995	0.16345	0.33799	0.44362
Combination 3	0.98897	0.17653	0.38764	0.42768

Combination 2 performed best overall based on R^2, MSE, and RMSE, making it the most accurate combination, combination 1 had the best MAPE, indicating that it is good in minimizing percentage errors, and combination 3 was the least favorable option, particularly because of its higher RMSE, but its MAPE is better than combination 2.

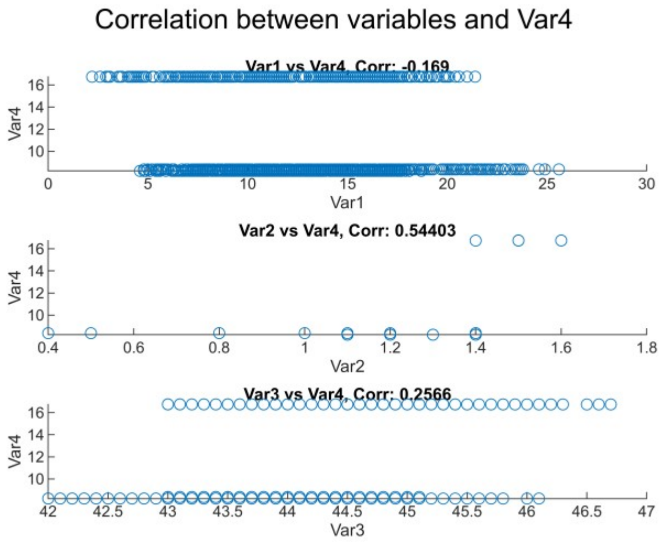


Fig. 10: Visualization of the correlations between input parameters and the output for combination 1

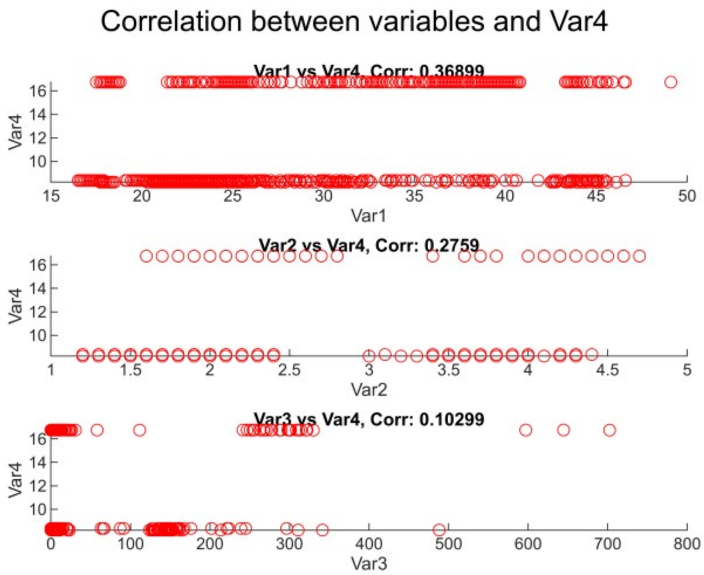


Fig. 11: Visualization of the correlations between input parameters and the output for combination 2

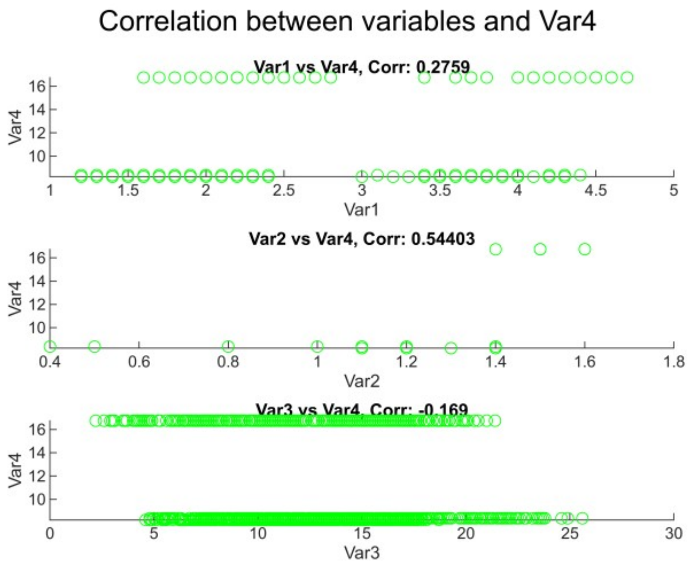


Fig. 12: Visualization of the correlations between input parameters and the output for combination 3

Combination 1 Combination 2 Combination 3

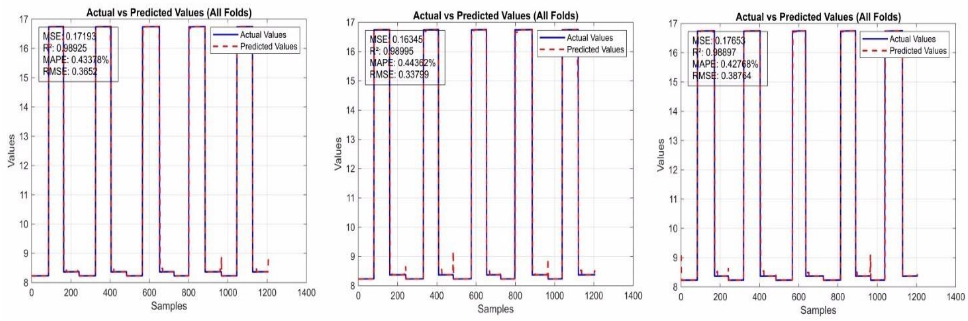


Fig. 13: Graphical representation of the prediction results and Random Forest analysis on three different combinations of input parameters

4. CONCLUSION

Based on the obtained results, the research demonstrated how different machine learning methods can contribute to prediction of heat energy consumption in district heating systems. By using data collected from the SCADA system of the Faculty of Mechanical Engineering heating plant in Niš, predictions were made using different machine learning methods: CCNN, RNN, and the Random Forest algorithm. The aim of the research was to compare the performance of these algorithms in the context of time series and classification tasks.

The results showed that the Random Forest algorithm offers the most stable accuracy and robustness, especially in combination with input parameters such as water temperature and pressure in the primary circuit, making it highly effective for time series estimation. This method was more resistant to overfitting and provided consistent results, particularly in reducing the Mean Squared Error (MSE) and Root Mean Squared Error (RMSE). Although it had a slightly higher MAPE compared to other models, Random Forest demonstrated great resilience to noise in the data, making it especially suitable for systems with high variability in input data. On the other hand, the Cascade-Correlation Neural Network (CCNN) model achieved the lowest MSE and RMSE, demonstrating a high level of accuracy in predicting heat energy. This model was particularly efficient in recognizing complex patterns in the data, but it required more precise parameter tuning and was more prone to overfitting. The Recurrent Neural Network (RNN) also achieved notable results, with a high coefficient of determination R^2, indicating its ability to recognize time patterns in data series, but it was less accurate compared to the CCNN and Random Forest algorithms.

Acknowledgement: *This research was financially supported by the Ministry of Science, Technological Development and Innovation of the Republic of Serbia under the grant 451-03-65/2024-03/200095.*

5. REFERENCES

1. Kudela L., Chýlek R., and Pospíšil J., "Efficient Integration of Machine Learning into District Heating Predictive Models," Energies, vol. 13, no. 23, pp. 6381, 2020.
2. Shesho I., "Optimization Model for Improvement of District Heating System by Integration of Cogeneration," Thermal Science, vol. 25, no. 1, pp. 307-320, 2021.
3. Tasić Milica, Ćirić Ivan, Ignjatović Marko, Stojiljković Dušan, " Comparison of Various Machine Learning Methods for Heat Load Prediction in Heat Load System" , XVIIth International SAUM Conference on Systems, Automatic Control and Measuremetns in Niš, Serbia, 2024., doi: 10.46793/SAUM24.169T
4. Oh K., Kim E-J., and Park C-Y., "A Physical Model-Based Data-Driven Approach to Overcome Data Scarcity and Predict Building Energy Consumption," in Sustainability, vol. 14, no. 15, 2022, pp. 9464.
5. S.E. Fahlman and C. Lebiere, "The cascade-correlation learning architecture," in Advances in Neu- ral Information Processing Systems (NIPS), vol. 2, 1990, pp. 524-532.
6. Elcioglu, E.B., "A High-Accuracy Thermal Conductivity Model for Water-Based Graphene Na- noplatelet Nanofluids," Energies, vol. 14, no. 16, p. 5178, 2021.
7. Collopy M. and F., "How effective are neural networks at forecasting and prediction? A review and evaluation," J. Forecasting, vol. 17, no. 5-6, pp. 481-495, 1998.
8. Zhang P., Lan C., Xing J., Zeng W., Xue J. and Zheng N., "View Adaptive Neural Networks for High Performance Skeleton-Based Human Action Recognition," in IEEE Transactions on Pattern Anal- ysis and Machine Intelligence, vol. 41, no. 8, pp. 1963-1978, 1 Aug. 2019, doi: 10.1109/TPAMI.2019.2896631.
9. Wu Wenbo, and Yun Pan, "Adaptive Modular Convolutional Neural Network for Image Recogni- tion" Sensors 22, no. 15: 5488, 28 June 2022, https://doi.org/10.3390/s22155488
10. Chandra R. and Zhang M., "Cooperative coevolution of Elman recurrent neural networks for chaotic time series prediction," Neurocomputing, vol. 86, pp. 116-123, Apr. 2012.
11. Elman J.L., "Finding structure in time," Cogn. Sci., vol. 14, no. 2, pp. 179-211, Apr. 1990.

12. Lipton, Z.C., "A Critical Review of Recurrent Neural Networks for Sequence Learning," arXiv preprint, arXiv:1506.00019, 2015.

13. Hochreiter S. and Schmidhuber J., "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735-1780, Nov. 1997.

14. Graves A. and Schmidhuber J., "Framewise phoneme classification with bidirectional LSTM and other neural network architectures," Neural Networks, vol. 18, no. 5-6, pp. 602-610, Jul. 2005.

15. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser Ł., and Polosu- khin I., "Attention is all you need," in Proc. Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 2017, pp. 5998-6008.

16. Lipton Z.C., "A critical review of recurrent neural networks for sequence learning," arXiv preprint, arXiv:1506.00019, Jun. 2015.

17. Qin Y., Song D., Chen H., Cheng W., Jiang G., and Cottrell G., "A dual-stage attention-based recurrent neural network for time series prediction," in Proc. Int. Joint Conf. Artificial Intelligence (IJCAI), Melbourne, Australia, 2017, pp. 2627-2633. doi: 10.24963/ijcai.2017/366

18. Lai Y., Shao L., Xu Y., and Wu X., "Recurrent convolutional neural networks for traffic prediction in transportation networks," IEEE Trans. Intelligent Transportation Systems, vol. 19, no. 5, pp. 1457- 1467, May 2018. doi: 10.1109/TITS.2017.2722160

19. L. Breiman. Random forests. Machine Learning, 45:5–32, 2001.

20. V. Svetnik, A. Liaw, C. Tong, J. Culberson, R. Sheridan, and B. Feuston. Random forest: A classification and regression tool for compound classification and QSAR modeling. Journal of Chemical Information and Computer Sciences, 43:1947–1958, 2003.

21. R. Diaz-Uriarte and S.A. de Andres. Gene selection and classification of microarray data using ´ random forest. BMC Bioinformatics, 7:1471–2105, 2006.

22. R. Genuer, J.-M. Poggi, and C. Tuleau. Random Forests: Some Methodological Insights. arXiv:0811.3619, 2008.

23. R. Genuer, J.-M. Poggi, and C. Tuleau-Malot. Variable selection using random forests. Pattern Recognition Letters, 31:2225–2236, 2010.

24. Y. Freund and R. Shapire. Experiments with a new boosting algorithm. In L. Saitta, editor, Machine Learning: Proceedings of the 13th International Conference, pages 148–156, San Francisco, 1996. Morgan Kaufmann

25. J. Shawe-Taylor and N. Cristianini. Kernel Methods for Pattern Analysis. Cambridge University Press, Cambridge, 2004.

26. Biau, G. "Analysis of a Random Forests Model" Journal of Machine Learning Research, vol. 13, no. 1, 2012, pp. 1063-1095.

27. L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. Classification and Regression Trees. Chapman & Hall, New York, 1984.

28. P. Buhlmann and B. Yu. Analyzing bagging. ¨ The Annals of Statistics, 30:927–961, 2002.

A. Buja and W. Stuetzle. Observations on bagging. Statistica Sinica, 16:323–352, 2006.

29. G. Biau, F. Cerou, and A. Guyader. On the rate of convergence of the bagged nearest neighbor estimate. Journal of Machine Learning Research, 11:687–712, 2010.

30. T. O. Hodson, R. E. Comer, and M. R. Valenstein. Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not. Geoscientific Model Development, 15:5481–5494, 2022.

A. Babu, A. Goyal, N. H. Keskar, et al. XLS-R: Self-supervised Cross-lingual Speech Representation Learning at Scale. arXiv preprint arXiv:2111.09296, 2021.

31. Y. Wang, S. Li, and J. Chen. Using Life Table Techniques to Model Mortality Rates for Small Populations. Living to 100 Monograph, Society of Actuaries, 2017.

32. Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Computer Science, 7, e623. https://doi.org/10.7717/peerj-cs.623

33. Zdravković Milan, Ćirić Ivan, Ignjatović Marko, "Towards explainable AI-assisted operations in district heating systems", IFAC PapersOnLine 54-1 (2021) 390–395, Niš, 2021.

34. M. Simonović, V. Nikolić, and E. Petrović. Input Vector Impact on Short-Term Heat Load Prediction of Small District Heating System. Facta Universitatis, Series: Automatic Control and Robotics, Vol. 15, No. 3, pp. 209–219, 2016.

35. M. Simonović, V. D. Nikolić, E. P. Petrović, and I. T. Ćirić. Heat Load Prediction of Small District Heating System Using Artificial Neural Networks. Thermal Science, Vol. 20, Suppl. 5, pp. S355–S366, 2016. DOI: 10.2298/TSCI151229155S