

The role of machine learning algorithms in enhancing energy efficiency and minimizing operational costs in smart infrastructure

Sri Harsha Giri^{1*}, Dr. Nilam Nimraj Ghuge²

1 Research Scholar, Department of Electrical, University Of Technology, Jaipur, Rajasthan, India

harshaguys1988@gmail.com

2 Professor, Department of Electrical Engineering, University Of Technology, Jaipur, Rajasthan, India

Abstract: Construction is a significant consumer of energy and a significant contributor to carbon emissions; thus, it is imperative that intelligent solutions be implemented in order to enhance energy efficiency and reduce operating costs in the construction sector. Through the use of machine learning strategies, the objective of this study is to investigate the most effective ways in which smart infrastructure systems may maximize their energy efficiency. A quantitative strategy that was driven by data was used by our team. This method includes predictive modeling, feature engineering, and statistical correlation analysis, among other things. The temperature, lighting, heating, ventilation, and air conditioning (HVAC) systems, total occupancy, and interior equipment are the most significant aspects that influence energy use, as shown by Spearman's rank correlation. A number of machine learning models, including the XGB Regressor, the Keras Regressor, the Neural Network (MLP), the Random Forest Regressor, and the Support Vector Machine (SVR), were evaluated based on their performance using the Root Mean Square Error (RMSE). It was via the use of feature selection and feature importance procedures that the performance of the model was considerably enhanced. Based on the findings, it was determined that the optimization of Neural Network (MLP) algorithms with Variance Threshold and LASSO algorithms yielded the highest level of prediction accuracy, accompanied by the lowest RMSE values. On the other hand, statistical feature extraction brought to an improvement in generalization in ensemble models such as XGBoost and Random Forest. The findings indicate that the use of appropriate feature engineering approaches has a significant impact on the outcomes of operational efficiency and prediction accuracy. This research highlights the real-world potential of energy management systems that are driven by machine learning to reduce operational expenditures and energy use in situations that are associated with smart infrastructure.

Keywords: Machine Learning, Smart Infrastructure, Energy Efficiency, Operational Cost Reduction, Feature Selection, Neural Networks, Smart Buildings, Energy Optimization

INTRODUCTION

The building sector is a major target for efforts to limit the influence that energy consumption and carbon dioxide emissions have on the climate [1]. This is because the construction industry

is responsible for a considerable amount of the world's energy consumption and emissions of carbon dioxide. Infrastructure solutions that are both environmentally friendly and good for the environment are in high demand as a result of the increasing demand brought on by growing urbanization and the problems posed by climate change. At this point in time, it is very necessary to develop intelligent building energy management systems in order to achieve global sustainability and carbon reduction expectations [2]. Smart infrastructure systems provide significant opportunities to increase energy efficiency without compromising occupant comfort or operational productivity. This is particularly true in smart buildings that include artificial intelligence capabilities, automated systems, and high-tech sensors [3, 4]. One of the most significant opportunities is to maximize energy efficiency.

The use of artificial intelligence (AI) has lately gained significance as a method for optimizing the use of energy in smart infrastructure. In order to improve the effectiveness of the system, artificial intelligence systems are able to evaluate vast quantities of data that are supplied by sensors, forecast patterns of energy use, and implement control strategies in real time [5]. Integrated energy systems, lighting control, occupancy-based management, and optimization of heating, ventilation, and air conditioning (HVAC) systems are some of the domains in which machine learning approaches, such as supervised learning, deep learning, and reinforcement learning, have shown significant capacity for improvement [6, 7]. As Internet of Things (IoT) frameworks and edge computing technologies continue to advance, it is becoming more practical to integrate energy management systems that are driven by artificial intelligence (AI) in buildings that are already in operation [8]. By laying the foundations for data-driven decision-making and efficient infrastructure management, these technological innovations bring about the necessary foundation.

However, despite all of the progress that has been achieved, there is still a significant amount of work to be done before artificial intelligence-based energy optimization can be implemented on a large scale. Many study works, which are limited to simulation environments or controlled experimental settings [9, 10], have found very little validation in real-world operational circumstances. This is because the research works have been conducted in these contexts. Data quality restrictions, difficulties in connecting with legacy infrastructure, high implementation costs, cybersecurity risks, and concerns regarding occupant acceptance are some of the challenges that are preventing widespread deployment of AI-enabled systems [11, 12]. But these are only some of the issues that are preventing widespread deployment. As a result,

converting advances in prediction accuracy into tangible reductions in operational expenses continues to be a fight that goes on indefinitely.

Despite the fact that previous systematic reviews have explored the use of artificial intelligence in building energy management, these studies often had limitations such as a limited focus, the absence of quantitative synthesis, and an insufficient evaluation of the challenges involved with using the technology. As an example, Wang et al. [13] examined 45 different pieces of research; however, they did not do anything like a meta-analysis or a comprehensive investigation into how to put all of this information into practice. Due to the fact that prior assessments have not adequately addressed methodology, building typology, dataset, and performance measure heterogeneity, it is also difficult to evaluate artificial intelligence systems across a variety of smart infrastructure contexts.

The purpose of this project is to address such gaps by examining the ways in which applications of machine learning algorithms might potentially improve smart infrastructure systems in terms of decreased running costs and better energy efficiency. The purpose of this study is to experimentally evaluate the predictive performance of a number of different machine learning models, such as Neural Networks (MLP), Keras Regressor, Random Forest Regressor, Support Vector Machine (SVR), and XGB Regressor. The metrics that are used in this research are RMSE metrics. In order to successfully complete the assessment process, feature engineering, comprehensive statistical analysis, and comparative testing are all required. Through a combination of statistical correlation analysis, feature selection procedures, and feature significance methodologies, the main objective of the research is to reduce the performance gap that exists between the theoretical capabilities of artificial intelligence and the actual energy optimization that occurs under smart infrastructure conditions.

OBJECTIVES

1. To assess and contrast how well various machine learning algorithms estimate energy use in smart infrastructure systems.
2. To examine how feature significance and selection strategies affect enhancing forecast accuracy and lowering operating energy expenses.

RESEARCH METHODOLOGY

This research adopts a quantitative, data-driven approach in order to evaluate the ways in which machine learning algorithms contribute to the enhanced energy efficiency and lower operational costs of smart infrastructure systems. The method includes a number of different components, including the collecting of data, the cleaning of data, analysis, feature engineering, model construction, and evaluation of performance.

1. Data Collection

The operational and environmental parameters that are included in the dataset that was used for this research are the ones that have an impact on the amount of energy that is used by buildings. Several crucial aspects include the amount of energy that is used by the heating, ventilation, and air conditioning (HVAC) system, the total occupancy, the lighting load, the consumption of interior equipment, and climatic variables such as temperature. Specifically, the structured time-series operational data that is represented in the dataset was given by smart infrastructure monitoring systems and simulation results [14].

2. Data Preprocessing

To guarantee the accuracy and trustworthiness of the data, preprocessing was carried out. The processes outlined below were carried out:

- Removal of missing and inconsistent records
- Outlier detection and elimination through percentile analysis
- Normalization and scaling of numerical variables
- Seasonal analysis to understand consumption trends
- Verification of data distribution using statistical measures

For the purpose of ensuring that everything was statistically consistent, we used distribution analysis, measures of central tendency (mean and median), and standard deviation. Following the completion of the preprocessing, a dataset that was free of any significant outliers was obtained.

3. Statistical Analysis

We carried out a correlation study with the help of Spearman's rank correlation coefficient in order to discover the differences and similarities that exist between the variables. Temperature, lighting, heating, ventilation, and air conditioning (HVAC), total occupancy, and interior equipment are the most significant elements that influence energy use, as shown by the heatmap in Figure 1. Through the use of this method, independent predictors were discovered and multicollinearity was eliminated [15].

4. Feature Engineering

In order to enhance the performance of the model, two different feature optimization approaches were utilized:

a) Feature Selection Techniques:

- Backward Elimination
- Mutual Information
- Recursive Feature Elimination (RFE)
- Variance Threshold

b) Feature Importance Methods:

- Decision Tree importance
- Extra Trees importance
- Random Forest importance
- LASSO regularization
- Greedy selection

Optimized datasets for modeling were generated by using these strategies to eliminate redundancy.

5. Machine Learning Model Development

There were five different machine learning algorithms tested and evaluated:

- Keras Regressor
- Neural Network (MLP)
- Random Forest Regressor
- Support Vector Machine (SVR)
- XGBRegressor

In order to evaluate the extent to which the dataset may be generalized, a training set and a testing set were designed and constructed from the dataset. The use of hyperparameter modification resulted in an increase in the precision of the forecasts received.

6. Model Evaluation

The root mean square error, often known as RMSE, was the major statistic that was used in order to assess the effectiveness of the models. The Root Mean Square Error (RMSE) was selected because of its capacity to precisely measure the occurrence of prediction mistakes and its more robust penalization of more significant deviations.

Three evaluation phases were conducted:

1. Performance using feature selection datasets (Table 1)
2. Performance using feature importance datasets (Table 2)
3. Performance using statistically derived datasets (Table 3)

7. Comparative Analysis

The RMSE values were used to analyze the models in order to determine which strategy is the most effective for predicting energy use. According to the results, the approaches used in feature engineering have an effect on the outcomes of model performance and operational efficiency.

RESULT

The available variables were statistically analyzed, with a correlation analysis using Spearman's coefficient used initially [16]. Without considering interdependencies, our research found that HVAC, overall occupancy, lighting, and interior equipment are the most

critical factors for energy use. Figure 1's heatmap also shows that temperature is the most important climatic variable.

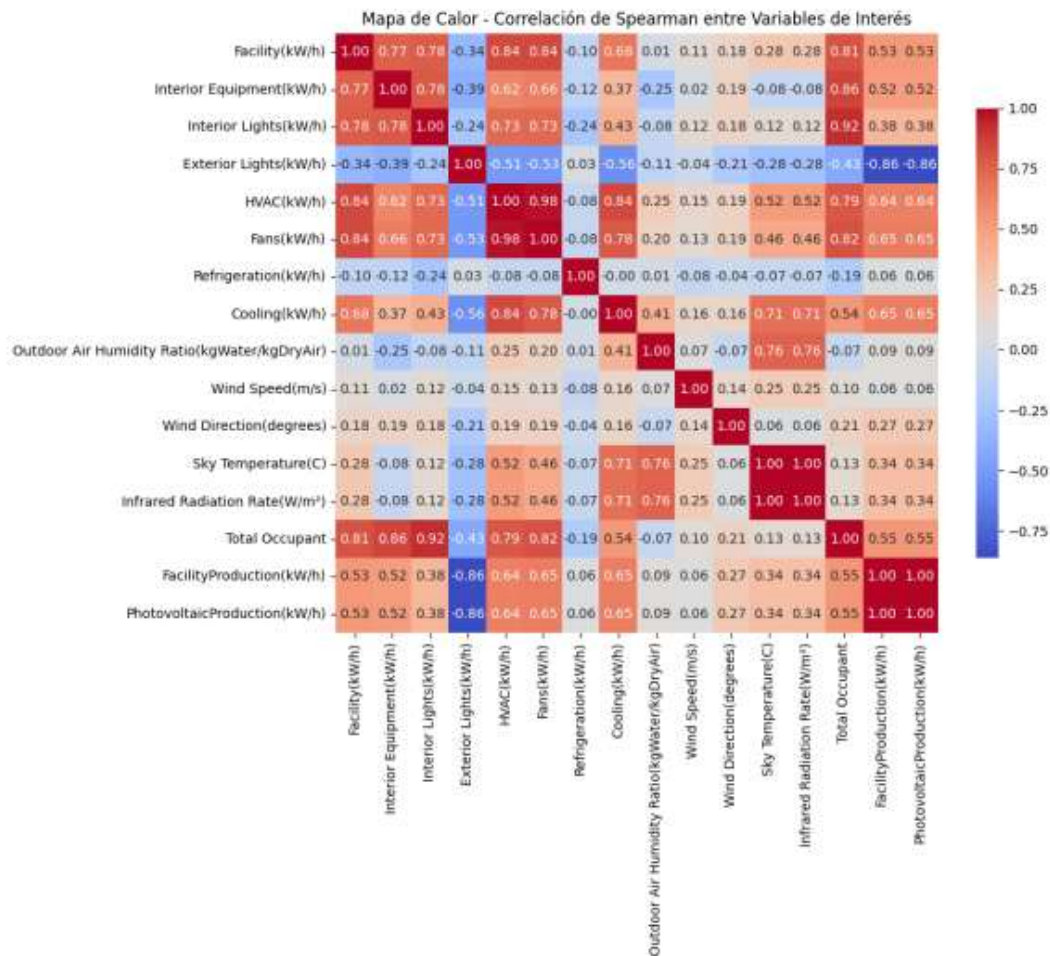


Figure 1. Heat Map of the Spearman coefficient.

This research, which includes measures of central tendency, data distribution analysis, percentile analysis, and seasonal analysis, results in the production of a dataset that is devoid of records that are particularly out of the ordinary. In order to evaluate the performance of the Keras Regressor, MLP, Random Forest Regressor, SVR, and XGBRegressor models during the modeling phase, we make use of the root-mean-squared error (RMSE) metric. The results of this investigation are shown in Table 1 [17], which makes use of the datasets obtained by feature selection methods.

Table 1. RMSE Using Techniques for Feature Selection

Model	Backward Elimination	Mutual Information	Recursive Feature Elimination (RFE)	Variance Threshold
Keras Regressor	1.275428848	1.335675642	1.678062718	1.195154758
Neural Network (MLP)	0.699321023	0.711946748	0.91663323	0.671586502
Random Forest Regressor	3.870392689	3.723547237	3.704117998	3.885881702
Support Vector Machine (SVR)	28.68902389	23.8256986	26.52675301	29.13542355
XGB Regressor	3.987850317	3.97702206	3.522636004	3.955199328

The root-mean-squared errors (RMSEs) for a number of different machine learning models that were trained using a range of feature selection procedures are shown in Table 1. It does not matter whatever method of feature selection is used; the Neural Network (MLP) model always performs better than the others. This is due to the fact that a reduced root mean square error (RMSE) indicates excellent prediction performance [18]. The Neural Network (MLP) with Variance Threshold obtains the highest overall performance, with an RMSE of 0.6716 and Backward Elimination coming in at 0.6993. Also achieved is the highest overall performance.

Despite the fact that MLP routinely achieves higher accuracy, the Keras Regressor comes very near to matching it, particularly when the Variance Threshold method is used (1.1952). Ensemble models, on the other hand, perform modestly; RMSE values that fall between 3.5 and 4.0 show that models such as Random Forest Regressor and XGB Regressor have a lower level of prediction accuracy. With RMSE values that are more than 23 in each and every case, the Support Vector Machine (SVR) has the worst performance of all the models, which indicates that it is not a suitable match for this dataset [19]. When it comes to smart infrastructure systems, the results indicate that neural network-based models, in particular MLP with Variance Threshold feature selection, are the most effective method for reducing prediction error and improving energy efficiency modeling.

Table 2. RMSE Using Methods of Feature Importance

Model	Decision Tree	Extra Trees	Random Forest	LASSO	Greedy Selection
Keras Regressor	1.69834858	0.95076538	1.119732593	1.12139474	2.25581072
Neural Network (MLP)	0.90851014	0.82108634	0.792092153	0.71370946	2.11893499
Random Forest Regressor	3.72209417	3.66911008	3.712471535	3.58525127	3.45318088
Support Vector Machine (SVR)	26.6863709	21.3416018	23.8256986	25.2648513	22.4777267
XGB Regressor	3.93959529	4.00906769	3.982473177	3.87915773	3.11526544

The results of the machine learning models that were selected based on the relevance of their features are shown in Table 2 [20]. Models that have a root-mean-squared error (RMSE) that is lower are considered to be more accurate. The Neural Network (MLP) obtains the lowest RMSE of 0.7137 when using LASSO. This is followed by the significance of Random Forest at 0.7921 and Extra Trees at 0.8211, indicating great performance across the majority of feature importance methodologies. The Keras Regressor performs quite well, particularly when the Extra Trees technique (0.9508) is used, despite the fact that it is not nearly as exact as Machine Learning. Ensemble models such as XGB Regressor and Random Forest Regressor display adequate performance, despite the fact that their prediction accuracy is lower than that of neural networks [21].

Root-mean-squared errors (RMSEs) of these models often lie between 3.1 and 4.0. In every single case, the Root Mean Square Error (RMSE) values are more than 21, which indicates that the Support Vector Machine (SVR) is not appropriate for this dataset. It is possible to reach the highest level of prediction accuracy by combining LASSO-based feature importance with logistic regression. with light of this, it can be deduced that the use of regularization-based feature selection enhances the effectiveness of neural networks with regard to the modeling of data about energy [22].

Table 3. Statistical Analysis and RMSE

Model	RMSE
Neural Network (MLP)	12.5498995
Support Vector Machine (SVR)	25.6832274
XGBRegressor	8.74757473
Random Forest Regressor	8.87646993
Keras Regressor	14.8418609

Following the completion of statistical analysis, a dataset was obtained, and Table 3 presents the RMSE values of selected models that made use of that dataset [23]. When it comes to the extraction of statistical features, tree-based ensemble techniques are superior to other models, as stated in [24]. With an RMSE of 8.7476 and 8.8765, respectively, the XGB Regressor and the Random Forest Regressor are both the ones that achieve the lowest value. A acceptable level of performance is shown by the Root Mean Square Error (RMSE) values of 12.5499 for the Neural Network (MLP) and 14.8419 for the Keras Regressor.

The root-mean-squared error (RMSE) of the Support Vector Machine (SVR) is the highest, coming up at 25.6832. This was precisely what happened. According to the findings of these research [25], ensemble learning models such as XGBoost and Random Forest exhibit enhanced generalizability when features derived from statistical analysis are applied rather than automated feature selection procedures. This demonstrates how it is essential for smart infrastructure systems to make use of the appropriate feature engineering approaches in order to increase the accuracy of their models and the efficiency with which they operate.

CONCLUSION

The purpose of this study is to emphasize the significance of machine learning algorithms in smart infrastructure systems for the purpose of enhancing energy efficiency and reducing operational expenses. After doing comprehensive statistical analysis and comparative modeling, it was shown that neural network-based models, namely the MLP paired with better feature selection methodologies, had the highest level of accuracy in terms of prediction. Furthermore, the findings indicate that the approaches for feature engineering have a

significant influence on the efficiency with which models function. Specifically, the results demonstrate that automated feature selection methods provide better prediction errors than statistical methods alone. Throughout all of the evaluation rounds, Support Vector Machine consistently performed badly in compared to other ensemble techniques such as XGBoost and Random Forest, which demonstrated strong generalization capabilities. A strong foundation for intelligent energy management is provided by cutting-edge machine learning techniques and efficient feature optimization when taken together. This foundation helps in the development of environmentally friendly infrastructure and the maintenance of buildings at a cheap cost.

References

1. International Energy Agency (2023) Energy efficiency 2023. <https://www.iea.org/reports/energy-efficiency-2023>
2. Global Alliance for Buildings and Construction (2024) 2024 global status report for buildings and construction. UN Environment Programme (UNEP).
3. Chen Y, Li Y, Hong T (2020) Review of reinforcement learning applications in building energy management. *Energy and AI* 1:100006. <https://doi.org/10.1016/j.egyai.2020.100006>
4. Wang Y, Zhang Y, Li H, Chen X (2021) A review of smart building energy management systems using artificial intelligence. *Energy Build* 250:111298. <https://doi.org/10.1016/j.enbuild.2021.111298>
5. McKinsey Global Institute (2023) The age of AI: Artificial intelligence and the future of building operations. McKinsey & Company.
6. Ahmad MW, Mourshed M, Rezgui Y (2017) Trees vs neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build* 147:77–89
7. Zhao Y, Liu C, Khosravi A (2022) Reinforcement learning for energy management in smart buildings: a comprehensive review. *Sustain Cities Soc* 76:103424. <https://doi.org/10.1016/j.scs.2021.103424>

8. Deloitte (2024) Smart buildings technology trends 2024: AI and automation in commercial real estate. Deloitte Insights.
9. Energy Information Administration (2020) 2018 Commercial Buildings Energy Consumption Survey (CBECS). US Department of Energy. <https://www.eia.gov/consumption/commercial/>
10. National Renewable Energy Laboratory (2023) Building AI integration: Technology readiness assessment. NREL Technical Report.
11. Department of Energy (2024) Building technologies roadmap: Artificial intelligence integration strategies. DOE Office of Energy Efficiency and Renewable Energy.
12. US Department of Energy (2023) EnergyPlus version 23.1. Building Technologies Office. <https://energyplus.net/>
13. Wang Y, Zhang Y, Li H, Chen X (2021) A review of smart building energy management systems using artificial intelligence. *Energy Build* 250:111298. <https://doi.org/10.1016/j.enbuild.2021.111298>
14. Bastidas Paz, C. H., & Chinchero Villacís, H. F. (2023). El Sistema de Gestión de Energía para Edificios Inteligentes utilizando Generación Distribuida. *Latin-American Journal of Computing*, 10(1), Article 1.
15. Bendaoud, N. M. M., Farah, N., & Ben Ahmed, S. (2022). Applying load profiles propagation to machine learning based electrical energy forecasting. *Electric Power Systems Research*, 203, 107635. <https://doi.org/10.1016/j.epsr.2021.107635>
16. 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>
17. Castillo De La Barrera, J. C. (2023). Metodologías de ahorro energético aplicadas a los sistemas HVAC utilizando inteligencia artificial: Una revisión del estado del arte. <https://repositorio.itm.edu.co/handle/20.500.12622/6023>

18. Centeno, M. A. C., & Ramírez, L. G. B. (2024). Smart energy soportado por Inteligencia Artificial: Un futuro sostenible y eficiente. Encuentro Internacional de Educación en Ingeniería. <https://doi.org/10.26507/paper.3926>
19. Chen, G., Lewis, F. L., Feng, E. N., & Song, Y. (2015). Distributed Optimal Active Power Control of Multiple Generation Systems. *IEEE Transactions on Industrial Electronics*, 62(11), 7079–7090. *IEEE Transactions on Industrial Electronics*. <https://doi.org/10.1109/TIE.2015.2431631>
20. Chinchero, H. F., & Alonso, J. M. (2020). A Review on Energy Management Methodologies for LED Lighting Systems in Smart Buildings. 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), 1–6. <https://doi.org/10.1109/EEEIC/ICPSEurope49358.2020.9160796>
21. Deb, S., & Gao, X.-Z. (2022). A hybrid ant lion optimization chicken swarm optimization algorithm for charger placement problem. *Complex & Intelligent Systems*, 8(4), Article 4. <https://doi.org/10.1007/s40747-021-00510-x>
22. Deepanraj, B., Senthilkumar, N., Jarin, T., Gurel, A. E., Sundar, L. S., & Anand, A. V. (2022). Intelligent wild geese algorithm with deep learning driven short term load forecasting for sustainable energy management in microgrids. *Sustainable Computing: Informatics and Systems*, 36, 100813. <https://doi.org/10.1016/j.suscom.2022.100813>
23. EnergyPlus. (n.d.). EnergyPlus. Retrieved December 11, 2024, from <https://energyplus.net/>
24. Freire, F. S. B., Fiallos, J. N. C., Caiza, C. I. Q., & Peña Herrera, H. R. R. (2023). Unidades recurrentes cerradas (GRU) vs redes neuronales artificiales en la predicción de la generación eléctrica de la Central Hidroeléctrica Illuchi. *AlfaPublicaciones*, 5(3), Article 3. <https://doi.org/10.33262/ap.v5i3.395>
25. Hotz, N. (2018, September 11). What is TDSP? Data Science PM. <https://www.datascience-pm.com/tdsp/>