Journal of Physics: Theories and Applications E-ISSN: 2549-7324 / P-ISSN: 2549-7316
J. Phys.: Theor. Appl. Vol. 9 No. 2 (2025) 93-102 doi: 10.20961/jphystheor-appl.v9i2.105932

Predicting Newtonian cooling with machine learning: a comparative analysis of gradient boosting and random forest models

Eko Sulistya

Physics Department Faculty of Mathematics and Natural Sciences Gadjah Mada University

sulistya@ugm.ac.id

Received 11 July 2025, Revised 8 August 2025, Published 30 September 2025

Abstract: This study investigates the use of artificial intelligence, specifically machine learning models, to predict temperature reduction in Newtonian cooling experiments involving varying volumes of water. Two regression models, Gradient Boosting Regression and Random Forest Regressor, were utilized to learn from empirical data. The findings indicate that both models are capable of accurately predicting cooling behavior, with the Random Forest model demonstrating superior accuracy for the dataset used. The machine learning models effectively represent the theoretical model of Newton's Law of Cooling, which is characterized by an exponential decay curve. Furthermore, the cooling constant for each volume was estimated using curve fitting techniques. This research underscores the potential of AI in modeling complex physical processes, particularly in real-world scenarios where the relationships between physical variables are intricate and challenging to express analytically. With sufficient data, AI can adeptly predict variable changes based on fluctuations in others. As technology continues to advance, AI is poised to assume an increasingly critical role in experimental and industrial applications involving complex physical systems. The novelty of this study lies in its comparative analysis to identify the optimal machine learning model—Gradient Boosting Regression or Random Forest Regressor—for accurately predicting Newtonian cooling behavior. Additionally, this research introduces an automated data acquisition approach using a datalogger, significantly enhancing precision and practicality compared to traditional manual methods involving a stopwatch and thermometer.

Keywords: Newtonian Cooling, Machine Learning, Temperature Prediction, Random Forest, Gradient Boosting.

1. Introduction

Newton's Law of Cooling, a cornerstone of thermodynamics, describes the rate at which an object's temperature changes in relation to the temperature differential between the object and its surroundings (Árpád et al., 2024). This principle, which assumes that the rate of heat transfer is proportional to the temperature difference, is widely employed in a variety of scientific and engineering domains (Li et al., 2012).

However, real-world scenarios frequently deviate from the idealized conditions of Newtonian cooling, necessitating the use of more sophisticated modeling techniques. The growing capabilities of machine learning provide an intriguing opportunity to improve the accuracy and adaptability of cooling analysis, enabling more precise predictions and efficient thermal management strategies (Svensen et al., 2024). By using machine learning algorithms, intricate relationships and nonlinearities that are frequently overlooked by traditional methods can be captured, and computational fluid dynamics can be integrated to get high-fidelity data (Fukami et al., 2020). This opens the door to a more thorough understanding of cooling processes, which has ramifications for anything from the design of electronic devices to the optimization of industrial cooling systems (Kochkov et al., 2021). We will explore the theoretical underpinnings of Newtonian cooling in this study, looking at its limitations and applicability in different situations. We will also look at how machine learning models can be used to improve the precision and effectiveness of cooling analysis.

Traditionally, the Newtonian cooling process has been studied using water cooling experiments, as is done in basic physics laboratories, where measuring cups with volumes of 100 ml, 250 ml, and 600 ml are filled with hot water. Temperature decrease in each measuring cup is read every 2 minutes, which can reduce accuracy and precision because time and temperature must be read simultaneously. The 2-minute time interval is also wide, potentially eliminating important data, especially at the beginning of the cooling process, where the temperature drops quickly. Due to the limited time allocated for practical work (generally 3 hours), the data obtained is less than optimal, which results in a less smooth temperature drop characteristic curve, and the value of the cooling coefficient obtained through the least square method is imprecise, making it difficult to validate Newton's Law of Cooling.

This study distinguishes itself through the application of machine learning methodologies to water cooling experiments, an approach that necessitates a substantial volume of data. To this end, a datalogger is employed to capture time and temperature data, supplanting conventional manual readings. Current technological solutions, comprising Arduino microcontrollers, datalogger modules equipped with integrated RTC and SD cards, and DS18B20 waterproof temperature sensors, facilitate the periodic measurement of water temperature and the logging of data in CSV format to an SD card, as implemented herein. By integrating machine learning models, this research addresses the limitations in accuracy and resolution inherent in traditional cooling experiments, thereby yielding more precise and robust predictive tools for thermal management.

2. Methodology

The methodology implemented in this investigation encompasses data collection, model selection, model training, and performance evaluation. The apparatus employed comprises three measuring glasses (100 ml, 250 ml, and 600 ml), each integrated with a waterproof temperature sensor DS18B20; a temperature sensor to measure the ambient temperature; an Arduino Uno microcontroller; and a datalogger module featuring an integrated Real-Time Clock and Secure Digital card module. In the experimental setup, the temperature sensors are connected to the Arduino, along with the datalogger module. Subsequently, the temperature sensors are introduced into the measuring glasses containing hot water, and the Arduino code is executed to record temperature variations over time. Following data acquisition, the collected data is used to train a machine learning model, with meticulous attention to the selection of the model architecture, hyperparameters, and training algorithms. For training purposes, the elapsed time from the experiment's commencement serves as the feature, while the corresponding water temperature functions as the label. The data, stored in CSV format on the SD card, is then transferred to a computer for subsequent processing and analysis. The elapsed time since the experiment began is used as the input feature, and the corresponding water temperature is the target variable.

The schematic of the data acquisition system, including the Arduino, temperature sensor, and SD card datalogger module are shown in Figure 1. The primary objective of this experimental setup is to collect high-fidelity temperature data over time, which will subsequently be used to train and validate machine learning models for predicting cooling behavior based on Newton's Law of Cooling (Loisel et al., 2021).

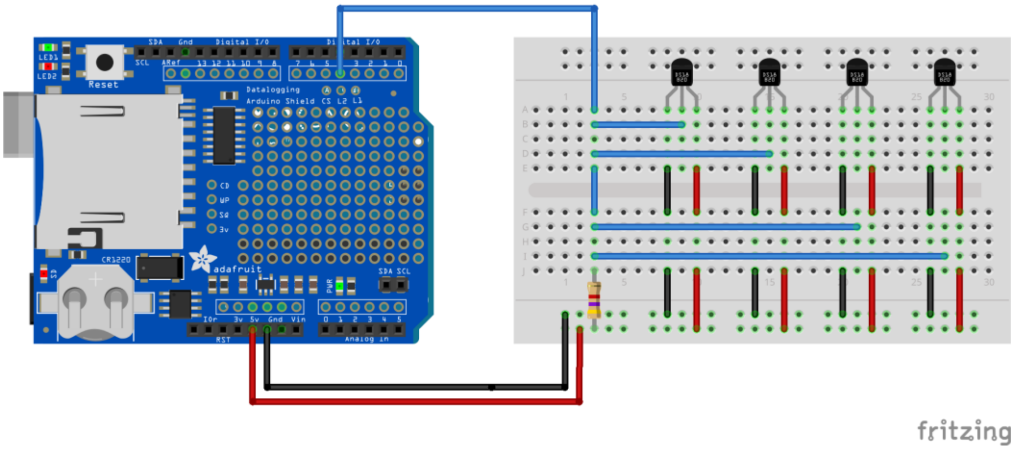

Figure 1. Schematic of the Data Acquisition System

The Arduino Uno is situated below the datalogger module. The DS18B20 sensors, represented as transistors in schematics, are housed in waterproof metal casings with attached cables for practical use, ensuring they are waterproof. A 4.7 kOhms resistor functions as a pull-up resistor, with the sensor's data output connected to digital pin 4. Three sensors are deployed to measure the water temperature in measuring glasseses containing varying water volumes, while the fourth sensor tracks the ambient room temperature, offering essential context for understanding the cooling dynamics. The incorporation of multiple temperature sensors, along with the SD card module's data logging capabilities, enables the development of a comprehensive dataset for training advanced machine learning algorithms. These algorithms can identify complex thermal patterns and forecast cooling trends across different experimental conditions. This meticulous arrangement guarantees the system's ability to precisely capture the thermal behavior of water samples, which is vital for training machine learning algorithms to accurately simulate the cooling process. The DS18B20 sensors, recognized for their accuracy and affordability, are well-suited for this application, providing acceptable temperature measurement precision. The Arduino-based system, which integrates the DS18B20 temperature sensors and an SD card module for data logging, delivers a dependable platform for real-time temperature monitoring and data collection in water cooling experiments.

3. Results and Discussion

The collected data, encompassing temperature readings from multiple sensors and varying water volumes, forms the empirical basis for developing and validating machine learning models to predict cooling behavior based on Newton's Law of Cooling. Figure 2 shows the trend of the temperature readings from all three experimental setups, demonstrating a clear exponential decay consistent with Newton's Law of Cooling, while the ambient temperature remained relatively stable throughout the experimental duration. The initial rapid temperature drop is gradually reduced as the water temperature approaches ambient temperature, demonstrating the cooling process's expected asymptotic behavior. The graph also compares the behavior of temperatures in each volume, showing that different volumes have a small impact on cooling rates.

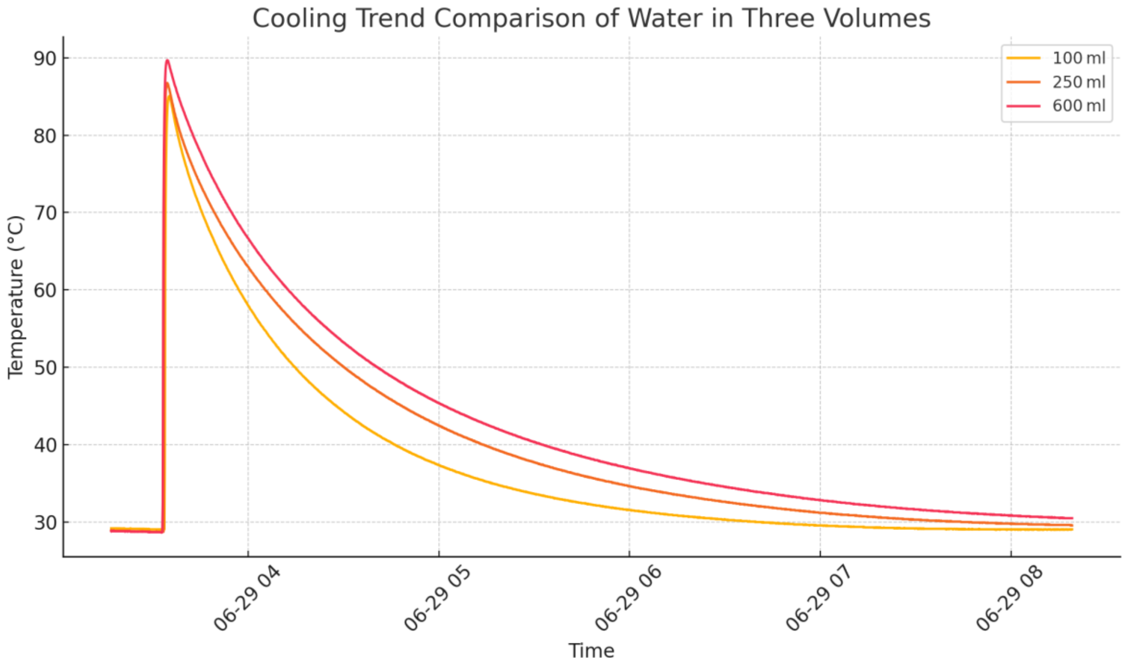

Figure 2. Temperature Data Trend for all Three Volumes

The initial data points were excluded to focus on the stabilized temperature data, ensuring accurate modeling of the cooling process. This exclusion of initial data points refines the dataset, allowing for a more accurate representation of the system's dynamic cooling behavior under steady-state conditions, which is crucial for robust model training. Further, this initial data exclusion helps in eliminating noise and transient effects that are not representative of the actual cooling process governed by Newton's Law. The data collected from these experiments can be used to validate theoretical models of heat transfer, which often rely on principles such as Fourier's law of heat conduction and the Stefan-Boltzmann law in addition to Newton's law of cooling. This ensures that the analysis focuses on the well-defined cooling phase, thereby enhancing the precision and reliability of the machine learning models developed (Guo et al., 2015). The collected data is divided into training and testing datasets to assess the machine learning model. The training set, which is larger, is used to teach the model, while the testing set evaluates its performance with new data to ensure the model can generalize effectively.

A crucial component of this study involves the development and application of machine learning models to predict and analyze Newtonian cooling dynamics. These models offer a data-driven approach to understanding the complex relationships between temperature, time, and environmental conditions, surpassing the limitations of traditional analytical methods (Raad et al., 2019). It is clear from Figure 2 that the temperature data trend shows exponential decay so we can use machine learning models to predict those changes. Two model algorithms will be applied which are Gradient Boost Regression and Random Forest Regressor, while Linear Regression is obviously not applicable in this case.

The procedures in using machine learning are collecting data, preprocessing, selecting a model, training, validation and parameter tuning, testing, and deployment. Machine learning methods have become indispensable tools in various engineering and scientific domains, offering robust capabilities for automated decision-making in complex systems (Bunyan et al., 2025). The selection of appropriate machine learning algorithms and the optimization of their parameters are critical steps in achieving high prediction accuracy and reliable performance (Zeng, 2024).

Gradient Boosting Regression is an ensemble learning technique that constructs a predictive model through the sequential combination of multiple weak learners, typically decision trees (Sha et al., 2021) while Random Forest Regressor operates by constructing a multitude of decision trees during the training phase and outputting the average prediction of the individual trees (Zhang et al., 2025). The two models are programmed with python and using libraries such as pandas, numpy, matplotlib, and scikit-learn. The dataset is split into training and testing sets, with 80% of the data used for training and 20% reserved for testing, while the Root Mean Squared Error is used to evaluate the performance of the models.

Figure 3 shows the visualization of the predicted values and experimental data for the Gradient Boosting Regression model, while Figure 4 shows the visualization of the predicted values and experimental data for the Random Forest Regression model. The visualization results demonstrate the predictive performance of both models, highlighting their ability to capture the underlying patterns in the temperature data. It appears that the Random Forest Regressor is more accurate in predicting temperature.

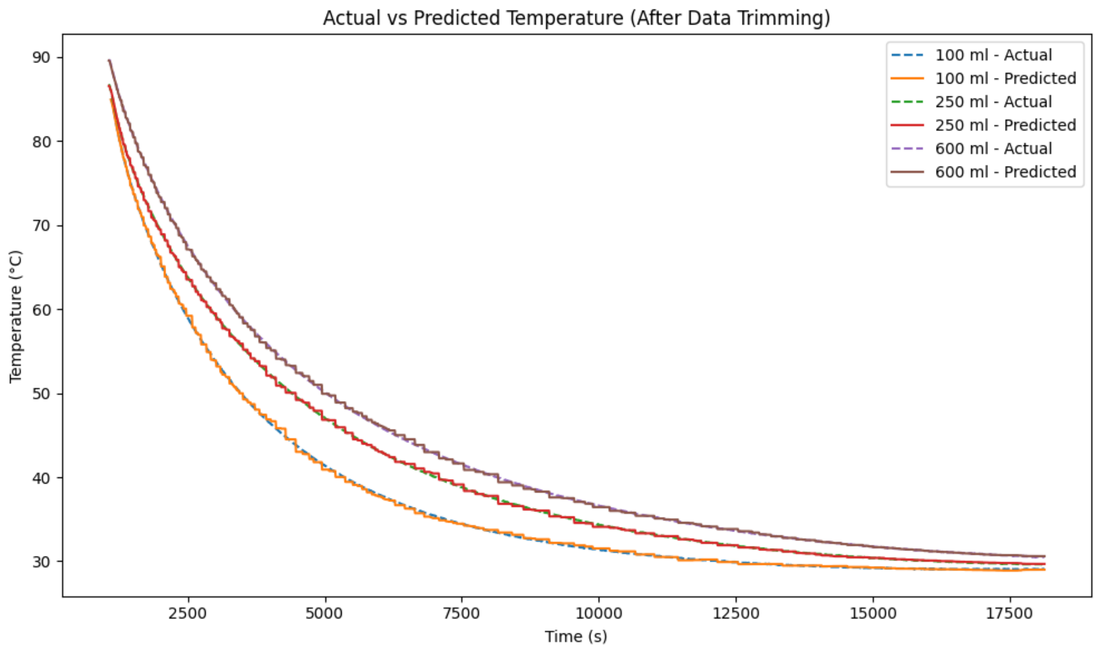

Figure 3. Comparison of Actual Versus Predicted Temperature using Gradient Boosting Regression Model

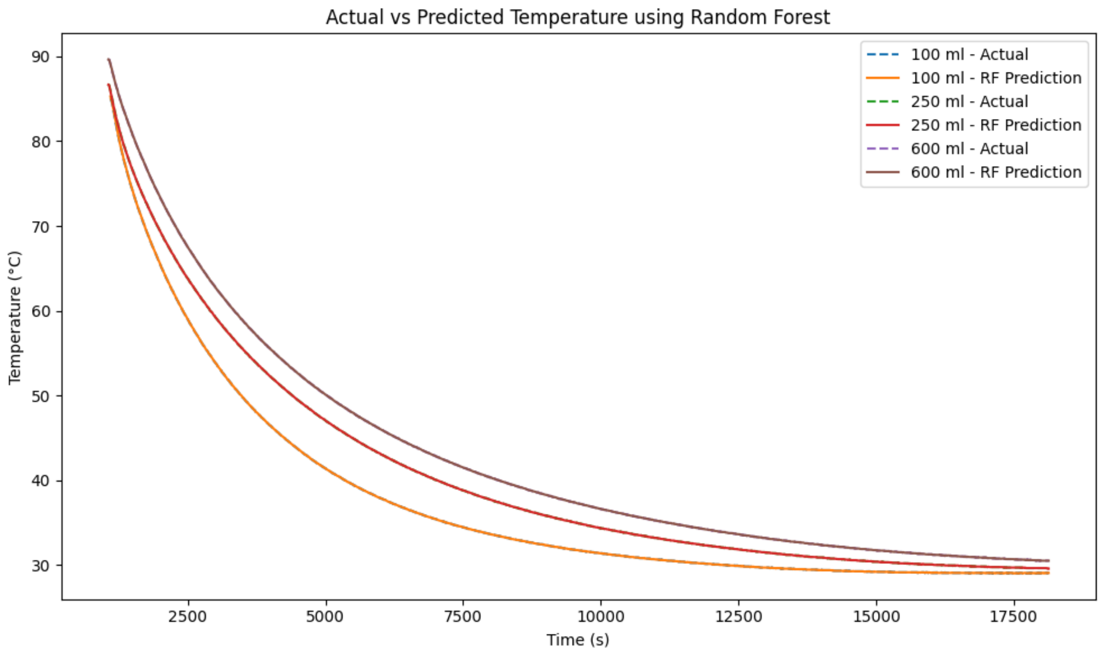

Figure 4. Comparison of Actual Versus Predicted Temperature using the Random Forest Regression Model.

The accuracy of the machine learning model is determined by the values of MAE and Root Mean Squared Error—the smaller the value, the more accurate the model (Già & Papurello, 2022)—and R^2—the larger the value, the more accurate the model (Samadi et al., 2023). The results of the two models have been shown in Table 1.

Table 1. Testing Results of Gradient Boosting Regression and Random Forest Regression

Metric	Gradient Boosting Regression	Random Forest Regression
Mean Absolute Error	0.1346	0.0091
Root Mean Squared Error	0.1868	0.0156
R-squared	0.9998	1.0000

Both models show good performance, but Random Forest Regression is better than Gradient Boosting Regression due to the lower values of error and the higher value of the R^2 value. The superior performance of the Random Forest model might be attributed to its robustness against overfitting and ability to handle small datasets better than Gradient Boosting. Furthermore, potential experimental errors such as sensor delays or slight variations in environmental conditions may contribute to prediction errors. However, other studies suggest that Gradient Boosting typically exhibits superior prediction performance compared to Random Forest (Ahn et al., 2023). This is likely due to the amount of data used and the physics phenomena being studied. The choice of hyperparameters significantly influences the predictive performance of machine learning models, particularly in ensemble methods like Random Forest and Gradient Boosting (Parikh et al., 2019). Thus, the choice of specific parameters can substantially impact the model's performance (Feng et al., 2021).

In the context of proving Newton's Law of Cooling, especially the form of the equation, which involves a cooling constant, cannot be determined by machine learning. Machine learning algorithms, including regression or classification, are designed to find patterns and relationships in data but cannot directly validate or prove the truth of a fundamental law of physics. To validate Newton's Law of Cooling, specifically the cooling constant, for the three water volumes examined, regression analysis was performed using Python with libraries like NumPy, pandas, and SciPy, as shown in Table 2.

Table 2. The cooling constant derived from the three different volumes of water

Volume (ml)	k (1/min)	R^2 (goodness of fit)
100	1.60×10^{-2}	0.56
250	1.50×10^{-2}	0.97
600	1.24×10^{-2}	0.99

Knowing the cooling constant, k, allows for the calculation of an object's cooling rate, a task that can also be accomplished with the machine learning models previously discussed. The application of Newton's Law of Cooling in this study serves as an illustrative example of using artificial intelligence in experiments, particularly when an exponential trend is observed. In scenarios where the relationship between variables is intricate and lacks a straightforward mathematical representation, machine learning can be employed to address the problem effectively. The adoption of machine learning methodologies offers a paradigm shift in experimental design and optimization, enabling researchers to navigate complex parameter spaces and uncover non-intuitive solutions (Barker et al., 2020).

4. Conclusion

The testing and validation results indicate that machine learning models can effectively estimate experimental outcomes. However, physical constants, such as the cooling constant in Newton's Law of Cooling, must be determined through physical experiments and curve fitting. While machine learning models are unsuitable for validating or discovering physical laws, they can estimate physical properties. Integrating physics knowledge into deep learning models can enhance prediction accuracy and ensure physical consistency. This study demonstrates that using machine learning to estimate the temperature change of cooling water offers a quick method for obtaining temperature estimates. This approach saves time and resources compared to traditional numerical simulations or experimental studies.

In addition, the application of machine learning methodologies in this study underscores the importance of data-driven approaches in scientific inquiry, where algorithms can automatically learn and extract relevant features from complex datasets. This integration not only enhances the accuracy and efficiency of predictive modeling but also facilitates the discovery of underlying patterns and relationships that may not be readily apparent through traditional analytical techniques (Brunton et al., 2020). The hybridization of physics-based models with machine learning algorithms offers a synergistic approach to addressing complex scientific problems, combining the interpretability and generalizability of physics-based models with the data-driven adaptability of machine learning (Pawar et al., 2021). Future studies could investigate the use of deep learning models such as neural networks, or hybrid models combining physical laws with machine learning to further enhance the accuracy and interpretability of predictions.

This study has limitations, including the small dataset size and the specific conditions of the experiment that may limit generalization. Additionally, the environmental factors, such as humidity and air flow, were not controlled strictly, which may introduce variability in the results.

References

Ahn, J. M., Kim, J., & Kim, K. (2023). Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting. Toxins, 15(10), 608. https://doi.org/10.3390/toxins15100608

Árpád, I., Kiss, J. T., & Kocsis, D. (2024). Role of the volume-specific surface area in heat transfer objects: A critical thinking-based investigation of Newton's law of cooling. International Journal of Heat and Mass Transfer, 227, 125535. https://doi.org/10.1016/j.ijheatmasstransfer.2024.125535

Barker, A., Style, H., Luksch, K., Sunami, S., Garrick, D., Hill, F., Foot, C. J., & Bentine, E. (2020). Applying machine learning optimization methods to the production of a quantum gas. Machine Learning Science and Technology, 1(1), 15007. https://doi.org/10.1088/2632-2153/ab6432

Brunton, S. L., Hemati, M. S., & Taira, K. (2020). Special issue on machine learning and data-driven methods in fluid dynamics. Theoretical and Computational Fluid Dynamics, 34(4), 333. https://doi.org/10.1007/s00162-020-00542-y

Bunyan, S. T., Khan, Z. H., Al-Haddad, L. A., Dhahad, H. A., Al-Karkhi, M. I., Ogaili, A. A. F., & Al-Sharify, Z. T. (2025). Intelligent Thermal Condition Monitoring for Predictive Maintenance of Gas Turbines Using Machine Learning. Machines, 13(5), 401. https://doi.org/10.3390/machines13050401

Feng, Y., Duan, Q., Chen, X., Yakkali, S. S., & Wang, J. (2021). Space cooling energy usage prediction based on utility data for residential buildings using machine learning methods. Applied Energy, 291, 116814. https://doi.org/10.1016/j.apenergy.2021.116814

Fukami, K., Fukagata, K., & Taira, K. (2020). Assessment of supervised machine learning methods for fluid flows. Theoretical and Computational Fluid Dynamics, 34(4), 497. https://doi.org/10.1007/s00162-020-00518-y

Già, S. D., & Papurello, D. (2022). Hybrid Models for Indoor Temperature Prediction Using Long Short Term Memory Networks—Case Study Energy Center. Buildings, 12(7), 933. https://doi.org/10.3390/buildings12070933

Guo, T., Shang, B., Duan, B., & Luo, X. (2015). Design and testing of a liquid cooled garment for hot environments. Journal of Thermal Biology, 47. https://doi.org/10.1016/j.jtherbio.2015.01.003

Kochkov, D., Smith, J., Alieva, A., Wang, M., Brenner, M. P., & Hoyer, S. (2021). Machine learning–accelerated computational fluid dynamics. Proceedings of the National Academy of Sciences, 118(21). https://doi.org/10.1073/pnas.2101784118

Loisel, J., Duret, S., Cornuéjols, A., Cagnon, D., Tardet, M., Derens-Bertheau, E., & Laguerre, O. (2021). Cold chain break detection and analysis: Can machine learning help? Trends in Food Science & Technology, 112, 391. https://doi.org/10.1016/j.tifs.2021.03.052

Parikh, R. B., Manz, C. R., Chivers, C., Regli, S. H., Braun, J., Draugelis, M., Schuchter, L. M., Shulman, L. N., Navathe, A. S., Patel, M. S., & O'Connor, N. (2019). Machine Learning Approaches to Predict 6-Month Mortality Among Patients With Cancer. JAMA Network Open, 2(10). https://doi.org/10.1001/jamanetworkopen.2019.15997

Pawar, S., San, O., Nair, A., Rasheed, A., & Kvamsdal, T. (2021). Model fusion with physics-guided machine learning. arXiv (Cornell University). http://export.arxiv.org/pdf/2104.04574

Raad, R., Itani, M., Ghaddar, N., & Ghali, K. (2019). A novel M-cycle evaporative cooling vest for enhanced comfort of active human in hot environment. International Journal of Thermal Sciences, 142, 1. https://doi.org/10.1016/j.ijthermalsci.2019.04.010

Samadi, B., Raison, M., Mahaudens, P., Detrembleur, C., & Achiche, S. (2023). Development of Machine learning algorithms to identify the Cobb angle in adolescents with idiopathic scoliosis based on lumbosacral joint efforts during gait (Case study). arXiv (Cornell University). https://doi.org/10.48550/arxiv.2301.12588

Sha, H., Moujahed, M., & Qi, D. (2021). Machine learning-based cooling load prediction and optimal control for mechanical ventilative cooling in high-rise buildings. Energy and Buildings, 242, 110980. https://doi.org/10.1016/j.enbuild.2021.110980

Svensen, J. L., Silva, W. R. L. da, Merino, J. P., Sampath, D., & Jørgensen, J. B. (2024). A Dynamic Cooler Model for Cement Clinker Production. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2409.09076

Zeng, X. (2024). A Review on Design of Sustainable Advanced Materials by Using Artificial Intelligence [Review of A Review on Design of Sustainable Advanced Materials by Using Artificial Intelligence]. Deleted Journal, 1(1), 10006. https://doi.org/10.35534/amsm.2024.10006

Zhang, J., Yang, M., Dong, N., & Wang, Y. (2025). Machine-Learning-Based Ensemble Prediction of the Snow Water Equivalent in the Upper Yalong River Basin. Sustainability, 17(9), 3779. https://doi.org/10.3390/su17093779