Predictive Modeling of the Activation Energy of Petroleum Sludge Pyrolysis Using Artificial Intelligence Approaches

Karim Khiari[1(✉)], Hadjer Aicha Habieb[1], Youcef Sehili[1], Lyes Tarabet[1], and Meriem Braik[2]

[1] Ecole Militaire Polytechnique, Bordj-El-Bahri, 16046 Algiers, Algeria
k_khiari@yahoo.fr
[2] Algerian Petroleum Institute, 35000 Boumerdes, Algeria

Abstract. This study is about using a hybrid artificial intelligence approach to predict the activation energy (Ea) of petroleum sludge pyrolysis. Three modeling techniques were compared: Artificial Neural Networks (ANN), Radial Basis Functions (RBF), and a new hybrid model that combines Radial Basis Functions with a Genetic Algorithm (RBF-GA). Models were trained using physicochemical parameters (elemental composition, proximate analysis) curated from peer-reviewed literature.

In this study, RBF-GA demonstrated superior performance by achieving an determination coefficient (R^2) value of 0.79, having a root mean square error (RMSE) of 21.37 kJ/mol, while the independent ANN model produced an R^2 value of 0.26, and RBF received an R^2 of 0.50. This research highlights how AI-developed meta-models can be utilized as a feasible alternative to traditional isostatic methods for kinetic analysis of pyrolysis, and will serve as a powerful computational tool to support optimizing the valorization of petroleum waste products.

Keywords: Modelisation · Pyrolysis · Waste valorisation · Genetic algorithm

1 Introduction

Petroluem Sludge (PS) is one of the more challenging forms of hazardous waste created throughout the process of producing petroleum, including drilling, transport, storage, and refining. It crisis produce between 60 and 80 million tons of sludge every year worldwide [1]. Due to its complex physicochemical composition, petroleum sludge is known to cause serious environmental and health hazards if not handled appropriately. The conventional methods of disposal include landfills and incineration. These methods are still widely practiced but are associated with significant operating costs, as they may cause soil and groundwater contamination, greenhouse emissions, as well as secondary pollution Environmental Science and Pollution Research [2].

Recently, processes like thermochemical conversion processes (pyrolysis) have proven to be very promising as alternatives to the conversion of petroleum sludge.

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2027
O. Gervasi et al. (Eds.): ICCSA 2026 Workshops, LNCS 16765, pp. 171–180, 2027.
https://doi.org/10.1007/978-3-032-30536-7_11

Pyrolysis is the process of breaking down organic matter using heat in the absence of oxygen; this generates useful products, including pyrolytic oil, syngas, and char. This method produces much more environmentally friendly disposal options than do conventional practices and results in 70–80% recovery of biomass as potential energy sources. However, the efficiency of this process depends on the knowledge of its reaction kinetics [3]. Among the kinetic parameters that govern pyrolysis, activation energy (Ea) is c is the most important one. It reflects the amount of energy required to initiate thermal decomposition, as well as providing insight into the level of reactivity and stability of the sludge components. Determining Ea accurately is important for reactor design, process optimization, and large-scale implementation [4]. Traditionally, Ea is estimated using thermogravimetric analysis (TGA) together with isoconversional methods like Flynn-Wall-Ozawa (FWO) and Kissinger-Akahira-Sunose (KAS). While these model methods are commonly used, they have limitations when applied to materials like petroleum sludge, which are highly heterogeneous. Specifically, they have difficulty capturing nonlinear interactions between various influencing factors, such as elemental composition, proximate properties, catalytic effects, and temperature-controlled heating rates.

To overcome these issues, AI based approaches have begun to be used in the area of kinetics modelling, and especially with use of machine learning techniques that have demonstrated their value in working with nonlinear, multivariate and dimensionally large datasets. Artificial Neural Networks (ANN) have been shown to have tremendous potential in solving complex thermochemical problems. However, the basic ANN are generally associated with the following problems: overfitting, slow convergence, and local minima traps, especially when dealing with relatively small data sets [5].

The present investigation seeks to develop and establish a hybrid artificial intelligence approach for the prediction of the activation energy of petroleum sludge pyrolysis. The intended design of this hybrid approach would be the application of a Radial Basis Function (RBF) networks coupled with a Genetic Algorithm (GA) for optimized parameter selection, unlike the conventional RBF network that has limited flexibility associated with having fixed parameters [6].

The originality of this work lies in three main aspects:

- A comprehensive dataset has been built from the peer-reviewed literature, which combines both elemental and proximate analyses.
- The development of an optimized hybrid RBF-GA architecture specifically tailored for activation energy prediction.
- A systematic comparative evaluation between ANN, RBF, and RBF-GA models using rigorous cross-validation procedures.

By suggesting the application of an artificial intelligence-based metamodel as an auxiliary tool to the conventional isoconversional methods, the present study makes an important contribution to the development of predictive modeling in the thermochemical valorization of wastes, as well as the development of an efficient computational tool for optimizing petroleum sludge pyrolysis processes.

2 Materials and Methods

Through both Elemental and Proximate analyses, all PS samples were subjected to extensive characterization. Elemental analysis measured the content of carbon (C), hydrogen (H), oxygen (O), and nitrogen (N), while proximate analysis included the determination of volatile matter (VM), ash content (A), fixed carbon (FC), and moisture (M). In addition, the samples were analyzed for their biochemical composition in terms of cellulose, hemicellulose, and lignin contents. These physicochemical characteristics collectively form the set of input variables for modeling, denoted as X:

X = [M, VM, FC, A, C, H, N, O, cellulose, hemicellulose, lignin, catalytic ratio].

The Ea, expressed in kJ/mol, was estimated from TGA using isoconversional methods, specifically FWO and KAS. The output of the modeling approach is the activation energy and therefore it is represented by as Y = Ea.

Three different AI techniques were evaluated to predict Ea from the input parameters: ANN, Support Vector Regression (SVR), and Random Forest Regression (RFR). Each method was assessed for its ability to accurately model the complex nonlinear relationships between the physicochemical properties of the PS samples and their corresponding activation energies.

In order to ensure the accuracy of the metamodels during their construction and to eliminate the impact of the nature of the training data on the result, the leave one out cross validation technique (LOOCV), which has been shown to be very effective in the literature [7], is applied. This technique is based on excluding a portion of the data used to create the model for validation purposes.

2.1 Artificial Neural Networks (ANN)

The ANN is a computational model based on the human brain structure and is composed of interconnected nodes referred to as artificial neurons. The artificial neurons exist in layers and interact with each other via weighted connections. The architecture is defined by the number and types of layers and neurons in those layers, the training algorithm used to create the ANN, along with its activation function [8]. The ANN model adopted in this work is based on the multilayer ANN as depicted in Fig. 1 with nonlinear activation functions [9]. The output is determined by Eq. (1).

$$\hat{Y} = f(W_2 \cdot f(W_1 \cdot X + b_1) + b_2) \tag{1}$$

where W_1 and W_2 are the weight matrices, b_1 and b_2 are the biases, and f(.) is an activation function.

Figure 1 illustrates the overall workflow of the proposed approach. It highlights the interaction between the RBF model and the optimization process, providing a global view of the methodology.

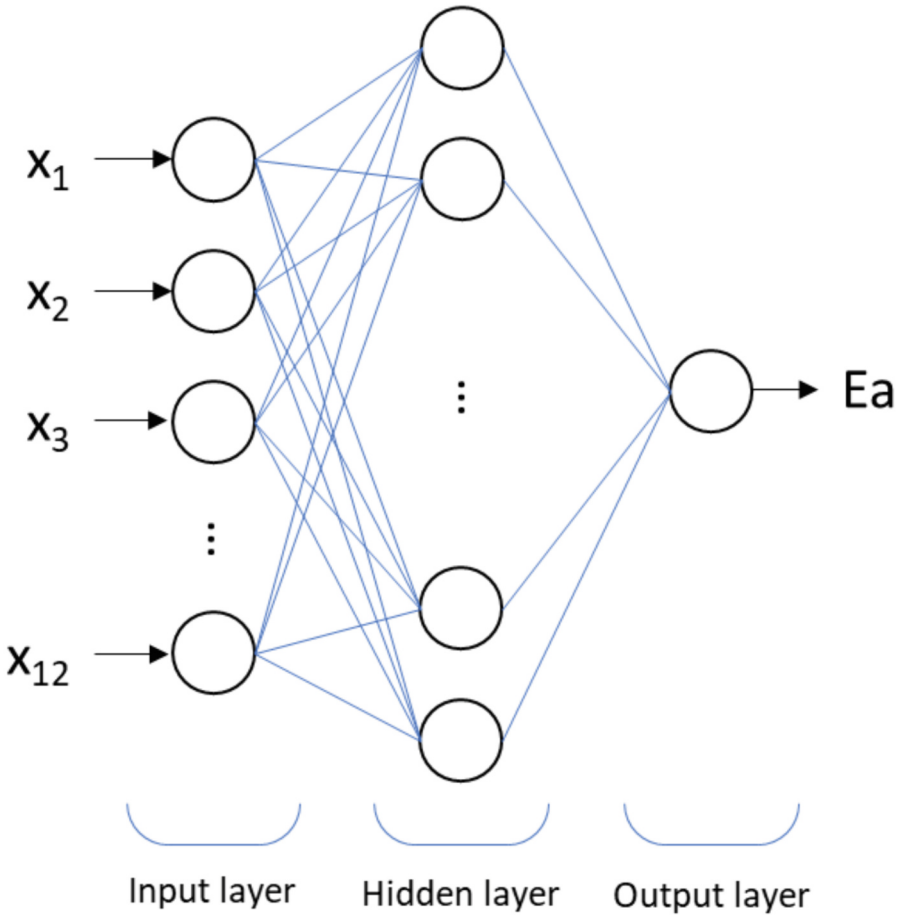

Fig. 1. ANN model architecture.

2.2 Radial Basis Function Networks (RBF)

The RBF uses a set of Gaussian functions, with each function centered around a certain point, ci [10]. These functions are used as activation functions that react to the distance between input vectors and centers.

In this work, the RBF method proposed in [7] has been used. As it has been shown in many comparative studies, RBF is one of the best approximations for problems in a high dimension space. The RBF approximation is given by Eq. (2).

$$f(x) = \sum_{i=1}^{N} \lambda_i \phi(\varepsilon_i x - x_i) \tag{2}$$

where the vector $x_1,\ldots,x_n \in \Re d$ is the set of Design of Experiments (DoE) points; φ is a user-defined radial function; λ is the vector of interpolation coefficients; and ε is the vector of shape factors.

2.3 Genetic Algorithm-Optimized Hybrid RBF Model (RBF-GA)

This model enhances the traditional RBF network by employing a GA to optimize its key parameters: the centers ci, widths σi, and weights wi [9]. In this study, ci represents the center of the radial basis function, σi denotes its spread (width), and wi is the associated weight parameter. Unlike fixed-parameter approaches, this evolutionary method allows the network to dynamically adapt its architecture and parameters based on the underlying data complexity. As a result, it significantly improves the model's flexibility, accuracy, and ability to generalize to nonlinear and high-dimensional problems.

As shown in Fig. 2, the parameters of the RBF are initialized in the first step. The model is constructed according to the experimental data with the RBF tool. According to the technique presented in this work, an error evaluation is performed to measure the convergence (Err_Conv = 1 for convergence achieved, Err_Conv = 0 for convergence not achieved). If convergence is not achieved, while respecting a maximum cost (Cos_Max), the RBF parameters will be optimised based on the genetic algorithm.

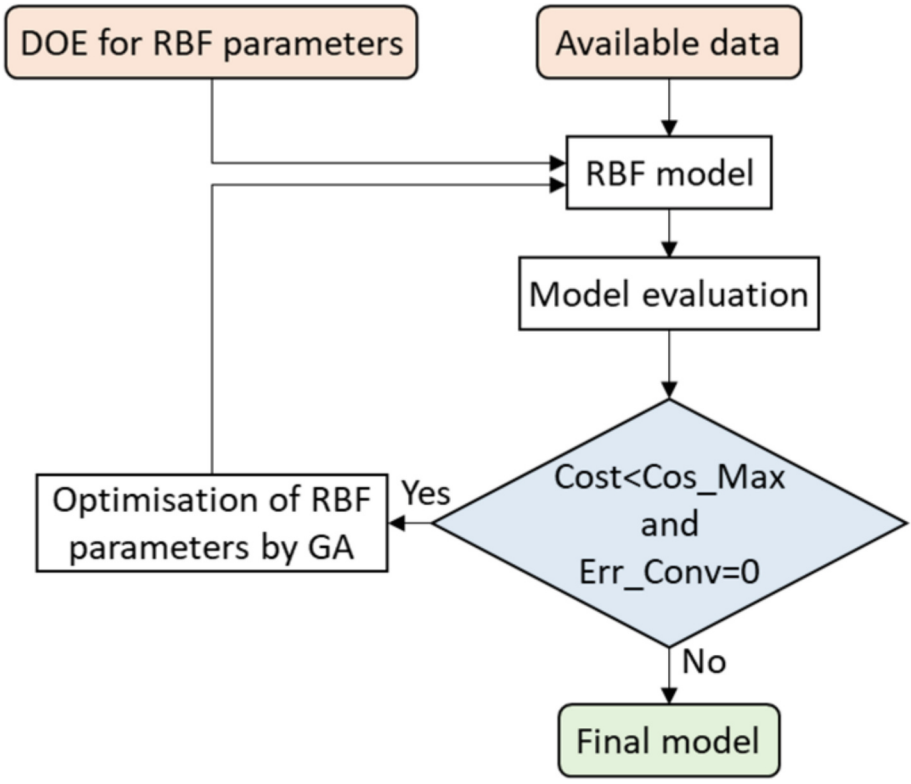

Fig. 2. Organisation chart for the RBF-GA model.

3 Results and Discussion

The comparison between predicted and experimental values showed a noticeable difference, with the model failing to accurately reflect the real data trends; this was quantitatively confirmed by the low coefficient of determination value of $R^2 = 0.26$, which indicates a poor correlation and low prediction ability. Such a result suggests that the ANN, in its current configuration, is not well adapted to the complexity of the input data. To enhance its performance, several improvements could be considered, including hyperparameter tuning (e.g., learning rate, number of hidden layers, activation functions) or the exploration of alternative ANN structures. Implementing some of these changes could potentially help capture non-linear relationships and increase generalization capability (Fig. 3). Cross-validation was used to ensure the reliability of the evaluation. This technique is based primarily on setting aside 5% of the data for validation and training the model using the remainder. This step is repeated several times, selecting different subsets for validation.

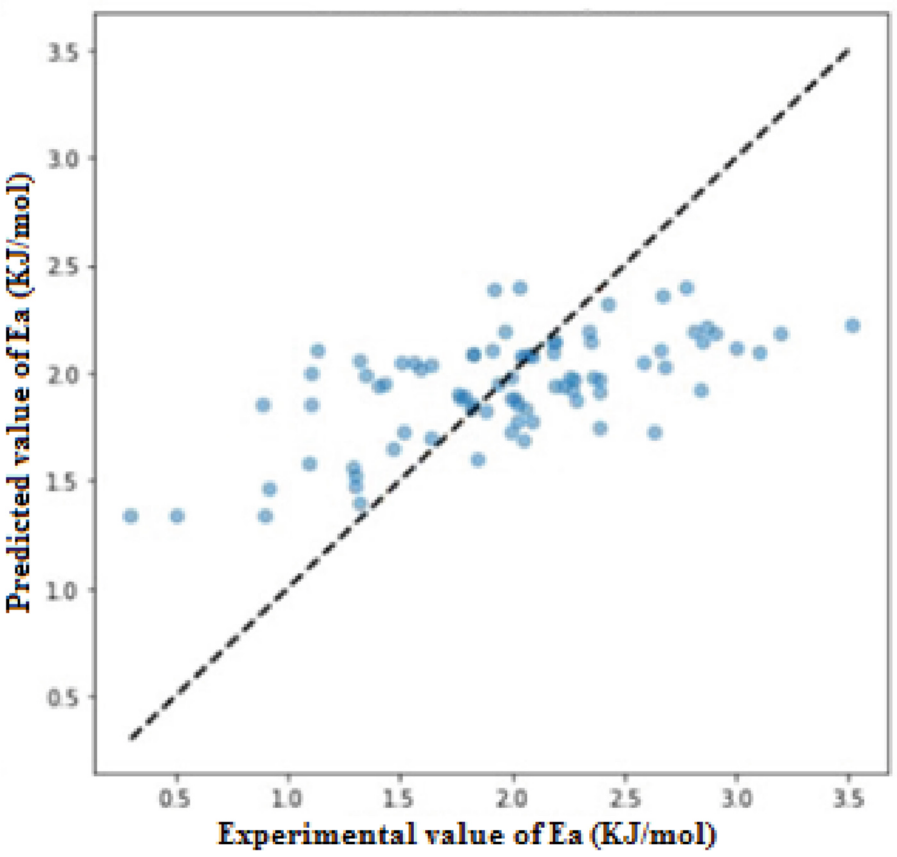

Fig. 3. Activation energy prediction using ANN.

In response to the limitations observed with the ANN model, a RBF model was tested. This approach led to a noticeable improvement, with a coefficient of determination value of $R^2 = 0.5$, reflecting a better correlation with the experimental activation energy values (Fig. 4). Despite this enhanced performance, the RBF model alone was still inadequate for achieving high predictive accuracy, indicating the need for further refinement or the integration of more advanced modeling techniques.

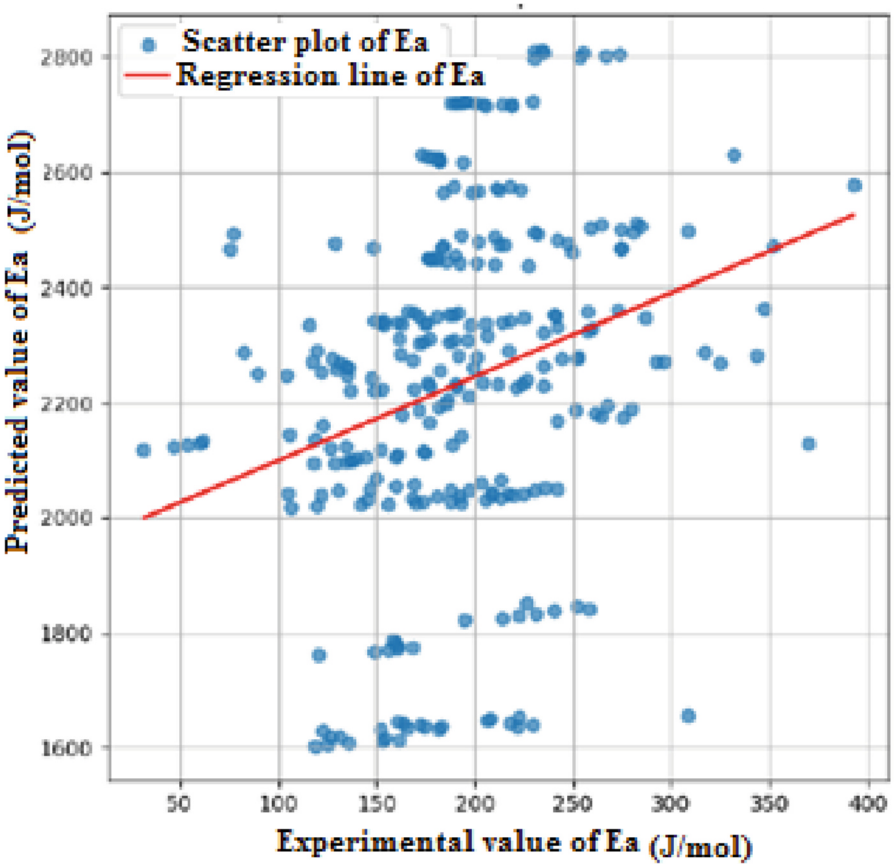

Fig. 4. Prediction of activation energy using RBF.

To overcome the previously observed limitations, a hybrid model combining the RBF network with a GA (RBF-GA) was developed. The results obtained with the RBF-GA model were compared with those of the standard RBF model. This advanced approach allowed for the dynamic optimization of the model's internal parameters, including centers, widths, and weights, leading to a significant enhancement in predictive accuracy. On the test dataset, the RBF-GA model achieved a coefficient of determination value of $R^2 = 0.79$, meaning it could explain 79% of the variability in activation energy values. In addition, root-mean-square error (RMSE) = 21.37 kJ/mol and mean absolute

error (MAE) = 13.45 kJ/mol confirm the predictive accuracy of the model. Cross-validation also shows that the RBF-GA model is robust and has generalization ability, producing an RMSE = 29.91 kJ/mol. Hence, it appears that the hybrid RBF-GA model is appropriate for modeling complex and nonlinear thermogravimetric data with a high degree of accuracy and consistency.

Table 1 presents a detailed comparison between the predicted and experimental activation energy values for different conversion ratios (α).

Table 1. Predicted and experimental activation energy values for different conversion ratios (a).

Conversion Ratio (α)	Predicted Ea (kJ/mol)	Experimental Ea (kJ/mol)
0.20	146.89	160.43
0.25	153.38	170.59
0.30	159.59	175.86
0.35	165.04	184.10
0.40	169.22	190.17
0.45	171.70	194.61
0.50	172.27	193.43
0.55	171.01	199.49
0.60	168.31	197.03
0.65	164.73	200.77
0.70	160.85	203.67
0.75	157.10	207.45
0.80	153.69	205.48

The hybrid RBF-GA model effectively captures the overall trend of Ea variation as a function of the conversion rate (α), demonstrating its ability to simulate complex kinetic behavior. However, a consistent underestimation of Ea values is observed across the entire conversion range. The most significant deviation occurs at $\alpha = 0.8$, where the model predicts an Ea of 153.69 kJ/mol, whereas the experimental value reaches 205.48 kJ/mol. This deviation may be due to the model's limitations in simulating the thermal decomposition of highly stable molecular structures, especially polycyclic aromatic hydrocarbons (PAHs) commonly found in PS. These compounds typically decompose at higher temperatures and conversion rates, which may introduce nonlinearities and kinetic complexities that challenge the model's generalization.

The results show that the RBF-GA model provides better accuracy than the standard RBF model, particularly for higher conversion ratios. This improvement can be explained by the ability of GA to optimize the parameters more effectively.

4 Conclusion

This study compared three modeling approaches for the prediction of Ea during the pyrolysis of petroleum sludge. The model based on ANN showed weak performance, with a coefficient of determination value of only 0.26. The RBF approach provided some improvement ($R^2 = 0.5$), but still insufficient.

The highest accuracy was achieved with the hybrid model RBF-GA, reaching $R^2 = 0.79$ on the test set. This model also reduced errors, with a RMSE of 21.37 kJ/mol and a MAE of 13.45 kJ/mol. Although it slightly underestimates high Ea values, it correctly reproduces the overall trend of the experimental curve, notably the reactivity peak around a conversion rate of 0.5.

These results confirm the relevance of hybrid approaches for modeling complex kinetic phenomena and open up promising perspectives for the development of predictive metamodels applied to the thermal valorization of petroleum residues.

References

1. Di, X., Pan, H., Li, D., Hu, H., Hu, Z., Yan, Y.: Thermochemical recycling of oily sludge by catalytic pyrolysis: a review. Scanning **2021**(1), 1131858 (2021)
2. Teng, Q., Zhang, D., Yang, C.: A review of the application of different treatment processes for oily sludge. Environ. Sci. Pollut. Res. **28**, 121–132 (2021)
3. Murungi, P.I., Sulaimon, A.A.: Petroleum sludge treatment and disposal techniques: a review. Environ. Sci. Pollut. Res. **29**(27), 40358–40372 (2022)
4. Hu, G., Li, J., Zeng, G.: Recent development in the treatment of oily sludge from petroleum industry: a review. J. Hazard. Mater. **261**, 470–490 (2013)
5. Naqvi, S.R., Ammar Taqvi, S.A., Khoja, A.H., et al.: Valorization of wet oily petrochemical sludge via slow pyrolysis: thermo-kinetics assessment and artificial neural network modeling. Front. Energy Res. **9**, 782139 (2022)
6. Garg, A., Osman, M.I., Tai, K., Gao, L.: Hybrid genetic algorithms and artificial neural networks for complex optimization problems. Int. J. Comput. Intell. Syst. **8**(3), 490–503 (2015)
7. Rippa, S.: An algorithm for selecting a good value for the parameter c in radial basis function interpolation. Adv. Comput. Math. **11**, 193–210 (1999)
8. Khalfallah, S., Lehtihet, H.E.: A multi-objective optimization methodology based on multi-mid-range metamodels for multimodal deterministic/robust problems. Struct. Multidisc. Optim. **60**(6), 2373–2389 (2019)

9. Mokashi, I., et al.: Nusselt number analysis from a battery pack cooled by different fluids and multiple back-propagation modelling using feed-forward networks. Int. J. Therm. Sci. **161**, 106738 (2021)
10. Illing, B., Gerstner, W., Brea, J.: Biologically plausible deep learning—but how far can we go with shallow networks? Neural Netw. **118**, 90–101 (2019)