Hindawi
International Journal of Chemical Engineering
Volume 2022, Article ID 8264297, 10 pages
https://doi.org/10.1155/2022/8264297

Research Article

On the Investigation of Effective Factors on Electronic Structure Properties of Transition Metal Complexes: Robust Modeling Using GPR Approach

Jianjun Wang[1] **and Mohammad Mahdi Molla Jafari**[2]

[1]*School of Petroleum and Petrochemical, Lanzhou University of Technology, Lanzhou 730050, China*
[2]*Department of Petroleum Engineering, Ahwaz Faculty of Petroleum Engineering, Petroleum University of Technology (PUT), Ahwaz, Iran*

Correspondence should be addressed to Jianjun Wang; wangjianjun_vip@outlook.com and Mohammad Mahdi Molla Jafari; mohammad.molajafari@afp.put.ac.ir

Received 9 December 2021; Revised 8 January 2022; Accepted 12 January 2022; Published 29 January 2022

Academic Editor: Alireza Baghban

Materials discovery is usually done using high-throughput computational screening. The use of costly and complex direct density functional theory (DFT) simulation methods has been commonly used to determine subtle trends in spin-state ordering and inorganic bonding of inorganic materials and, in general, to predict the electronic structure properties of transition metal complexes. A Gaussian process regression (GPR) framework consisting of four kernel functions is introduced for spin-state splitting estimation through inorganic chemistry-appropriate empirical inputs. To this end, the present study reviewed an extensive range of data values from earlier works. According to statistical analysis, the GPR model showed very good performance. The coefficients of determination were calculated to be 0.986 for the exponential and Matern kernel functions, suggesting the highest predictive power of these methods. Moreover, the sensitivity of output to inputs was measured. Artificial intelligence (AI) helped accurately predict the target values through various input ranges.

1. Introduction

Novel compounds, catalysts [1], and materials [2] are routinely discovered via high-throughput computer screening [3, 4]. Numerous screening and recognition experiments still rely on first-principles modeling, but the increased computational expense simulation means that only a narrow subset of the chemical domain can be explored [5, 6]. Lower thresholds of hypothesis, such as machine-learning designs, have emerged as alternatives to traditional methods for efficiently evaluating the latest candidate substances to speed up the exploration [7]. Computational chemists have recently discovered a broad range of uses for artificial neural networks (ANNs) [8–10]. The versatility of machine-learning methods to potential energy surfaces and, therefore, force field simulations were first recognized [9, 11–13]. Molecular or heterogeneous catalyst and substance exploration have lately been studied in exchange-correlation functional advancement [8, 14], common Schrödinger equation strategies [15], functional hypothesis for orbital-free density [16, 17], numerous body expansions [18], dynamics velocity [19, 20], and band-gap estimation [21, 22] among others.

The proper identification of widely relevant qualifiers that allow the ANN to be used dynamically beyond particles in the learning collection, e.g., for bigger molecules or those with varied chemical reactions, are essential difficulties for ANNs to substitute direct computation first-principles techniques. ANNs have had the greatest effectiveness thus far beyond proof-of-concept demonstrations developing force fields for well-defined substances, such as water [23, 24]. To make energetic predictions in organic chemistry,

compositional qualifiers like the Coulomb matrix [25] or regional chemical surroundings and adhesive descriptions [26, 27] have been helpful when considering only a small number of mixtures (e.g., C, H, N, and O). Molecular resemblance, force field advancement, numerical structure-activity [28] correlation, and commutative group hypotheses have all been successfully evaluated using cheminformatics in the past. There are just a handful of force fields [29] for transition metal combinations covering the whole spectrum of inorganic chemical bonding interactions [30]. More rigorous construction of qualifiers is needed to accurately anticipate the characteristics of open-shell transition metal combinations since spin state and coordination setting influence binding [31].

In the same way, qualifiers that were effective for organic molecules are ineffective for inorganic crystalline particles [32]. In transition metal combinations, it is well-recognized [33, 34] that the responsiveness of electronic characteristics (such as spin-state separation) correlates strongly with the ligand-atom linkage and ligand-field power [35, 36]. When substituting distantly (e.g., tetraphenyl porphyrin for base porphin), the impact will be restricted because ligands with the identical metal-bonding atom can have vastly distinctive ligand-field powers (for example, C for both weakened field CH_3CN and robust field CO). Therefore, the transition metal complex qualifier collection must cautiously balance metal-proximal and metal-distant qualifiers. A second issue pertains to establishing ANN estimations of first-principles characteristics in transition metal chemistry and associated inorganic substances. Transition metal complexes cannot benefit from efficient correlated wave function theory techniques (e.g., MP2) because optimal procedures for transition metal complexes remain mysterious [37]. In transition metal chemistry, while potential paths for ANNs involve projecting lower-level theory findings to a higher-level hypothesis (e.g., from semiempirical assumption) [38], as has been shown for atomization energies [39] and more recently reaction obstacles [40], appropriate degrees of theory for inference are less apparent. The level of precise (Hartree–Fock, HF) transfer to incorporate in the analysis of transition metal combinations is also unclear. Suggestions range from no interchange to alternatively low or large quantities of accurate interchange in a system-dependent way, notwithstanding inordinate delocalization faults in approximation DFT on transition metal combinations [35, 41, 42], with these amounts being determined by the system. It is true that measuring uncertainty about functional choice in energetic forecasts, particularly the responsiveness of projections to include precise interchange, has garnered a lot of attention lately. To get a direct number and understand how the exchange fraction [33, 34] affects spin-state splitting, one must first determine how responsive it is to interchange. To translate empirical forecasts or provide measurements of accuracy on calculated information, a machine-learning system that anticipates spin-state ordering among interchange rates would be helpful.

As a general rule, any presentation of artificial intelligence in inorganic chemistry, such as for the fast identification of novel spin-crossover combination [43, 44], the use of dye-sensitizers throughout solar panels [45], or the quick assessment of spin-state sequencing to determine the responsiveness of open-shell catalysts, should meet two requirements: (i) qualifiers must integrate metal-proximal and metal-distant properties and (ii) they must also anticipate spin-state sequencing when exchange-correlation blending is taking place. Cheminformatics-inspired transition metal complex structure creation instruments help us make progress toward both of these goals in this study. To educate GPR, as a new method, to anticipate the transition metal complex characteristic, we also developed structure-functional responsiveness correlations in transition metal combinations. In this study, various analyzes have been used to evaluate the proposed models. Our goal is to provide a model with high accuracy in predicting this goal parameter.

2. GPR Model

The present work adopted machine-learning and GPR to handle probabilistic (Bayesian) uncertainties [46, 47]. This approach can simply solve complicated problems. Nonlinear GPR techniques may be employed using small training datasets and integrate new evidence as the data points rise in number [48]. Overfitting is avoided to a great extent as optimization includes fewer hyperparameters in the training phase. The model parameters are determined by the GPR training dataset [49, 50]. Previous data are incorporated into the process along with empirical data to construct the GPR model. GPR operates based on posterior distribution calculations rather than identifying the highest consistency with empirical data, unlike traditional machine-learning algorithms [51].

Let x be the input and y be the output. Also, $T = \{x_{T\cdot i} \cdot y_{T\cdot i}\}_{i=1}^{n}$ denotes a random testing dataset, and $L = \{x_{L\cdot i} \cdot y_{L\cdot i}\}_{i=1}^{n}$ is a random training dataset. GPR begins with [52]

$$y_{L\cdot i} = f\left(x_{L\cdot i}\right) + \varepsilon_{L\cdot i}, \quad i = 1, 2, 3, \ldots, n, \tag{1}$$

where X_L and Y_L are the independent variable and target, respectively. Furthermore, $\sim N\left(0 \cdot \sigma_{\text{noise}}^2 I_n\right)$ denotes the observation noise, σ_{noise}^2 is the noise variance, and I_n is the unit array. As a result, the Gaussian noise model connects y values to $f(x)$. f is assumed to be a random function completely definable by the mean functions and covariance [53]. Similarly,

$$y_{T\cdot i} = f\left(x_{T\cdot i}\right) + \varepsilon_{T\cdot i}, \quad i = 1, 2, 3, \ldots, n, \tag{2}$$

where X_T and Y_T are the testing dataset independent variable and target, respectively, $f(x)$ is the Gaussian process distribution whose kernel function is $k(x \cdot x')$, and mean function is $m(x)$ [54]. Thus,

$$f\left(x_{L\cdot i}\right) \sim GP\left(m(x) \cdot k\left(x \cdot x'\right)\right). \tag{3}$$

Explicit basis functions (BFs) could be employed to determine $m(x)$. It should be noted that $m(x)$ is typically assumed to be zero for simplification purposes, since a constant $m(x)$ is difficult to find [55]. Therefore,

$$f(x_{L,i}) \sim GP(0 \cdot k(x \cdot x')). \tag{4}$$

The integration of (1) and (4) gives the y distribution as [56]

$$y \sim N\left(0 \cdot k(x \cdot x') + \sigma^2_{\text{noise}} I_n\right). \tag{5}$$

Based on the aforementioned parameters [57],

$$\begin{bmatrix} \overrightarrow{f_L} \\ \overrightarrow{f_T} \end{bmatrix} \sim N\left(0 \cdot \begin{bmatrix} k(x_L \cdot x_L) & k(x_L \cdot x_{LT}) \\ k(x_T \cdot x_L) & k(x_T \cdot x_T) \end{bmatrix}\right), \tag{6}$$

$$\begin{bmatrix} \overrightarrow{\varepsilon_L} \\ \overrightarrow{\varepsilon_T} \end{bmatrix} \sim N\left(0 \cdot \begin{bmatrix} \sigma^2_{\text{noise}} I_n & 0 \\ 0 & \sigma^2_{\text{noise}} I_n \end{bmatrix}\right). \tag{7}$$

A Gaussian expression is derived by summing up (6) and (7) [58]:

$$\begin{bmatrix} \overrightarrow{f_L} \\ \overrightarrow{f_T} \end{bmatrix} \sim N \cdot \left(0 \cdot \begin{bmatrix} k(x_L \cdot x_L) + \sigma^2_{\text{noise}} I_n & k(x_L \cdot x_{LT}) \\ k(x_T \cdot x_L) & k(x_T \cdot x_T) + \sigma^2_{\text{noise}} I_n \end{bmatrix}\right). \tag{8}$$

The Gaussian conditioning rule is used to obtain the y_T distribution (where Σ_T is the covariance, and μ_T is the mean) [59]:

$$\begin{aligned} (y_T|y_L) &\sim N(\mu_T \cdot \Sigma_T), \\ \mu_T &= m\left(\overrightarrow{y_T}\right) = k(x_T \cdot x_L)\left(k(x_L \cdot x_L) + \sigma^2_{noise} I_n\right)^{-1} \overrightarrow{y_T}, \\ \Sigma_T &= k(x_T \cdot x_T) = k(x_T \cdot x_T) + \sigma^2_{\text{noise}} I_n - k(x_T \cdot x_L) \\ &\quad \cdot \left(k(x_L \cdot x_L) + \sigma^2_{\text{noise}} I_n\right)^{-1} k(x_L \cdot x_{LT}). \end{aligned} \tag{9}$$

The output estimate of the testing dataset can be obtained by the independent variable and training dataset. The kernel function in the training phase (with asymmetric, invertible matrix) strongly influences GPR predictive performance. The present study implemented the learning technique to identify the most efficient kernel function, manipulating the Matern, exponential, squared exponential, and rational quadratic functions [60, 61].

The Matern kernel is given by

$$\begin{aligned} k_M(x \cdot x') &= \sigma^2 \frac{2^{1-\nu}}{\Gamma_{(\nu)}} \left(\sqrt{2\nu}\ \frac{x - x'}{\ell}\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}\ \frac{x - x'}{\ell}\right), \\ k_E(x \cdot x') &= \sigma^2 \exp\left(-\frac{x - x'}{\ell}\right), \\ k_{RQ}(x \cdot x') &= \sigma^2 \left(1 + \frac{x - x'^2}{2\alpha\ell}\right)^{-\alpha}, \\ k_{SE}(x \cdot x') &= \sigma^2 \exp\left(-\frac{x - x'^2}{\ell^2}\right), \end{aligned} \tag{10}$$

where $\alpha > 0$ is the length scale, $\ell > 0$ is the scale mixture, σ denotes amplitude, and σ^2 is the variance. Moreover, K_ν is the modified Bessel function, ν is a positive variable, and Γ stands for the gamma function. For $\nu = 0.5$, the Matern kernel converts into the exponential kernel function, whereas $\nu = 1.0$ transforms the Matern kernel into the squared kernel function (two particular cases of the Matern kernel) [62, 63].

To maximize mode accuracy, 1/5 of the data was employed as the testing dataset to measure model validity, while the remaining data that were exploited was the training dataset for spin-state splitting evaluation. Details of the data are given elsewhere [64]. Performance evaluation was carried out using MSE, R^2, STD, MRE, and RMSE. These statistical indices are calculated as [65–68]

$$\begin{aligned} R^2 &= 1 - \frac{\sum_{i=1}^{n} \left[x_i^{\text{predicted}} - x_i^{\text{experimental}}\right]^2}{\sum_{i=1}^{n} \left[x_i^{\text{predicted}} - x_m\right]^2}, \\ \text{STD} &= \sqrt{\sum_{i=1}^{n} \frac{\left(x_i^{\text{predicted}} - x_m\right)^2}{n}}, \\ \text{MSE} &= \frac{1}{n} \sum_{i=1}^{n} \left(x_i^{\text{predicted}} - x_i^{\text{experimental}}\right)^2, \\ \text{RMSE} &= \sqrt{\frac{\sum_{i=1}^{n} \left[x_i^{\text{predicted}} - x_i^{\text{experimental}}\right]^2}{n}}, \\ \text{MRE} &= \frac{1}{n} \sum_{i=1}^{n} \frac{\left|x_i^{\text{predicted}} - x_i^{\text{experimental}}\right|}{x_i^{\text{experimental}}}. \end{aligned} \tag{11}$$

3. Accuracy Estimation

A portion of data may show inconsistency with the dataset, with some data being suspected. Such data points majorly imply empirical errors [69, 70]. It is necessary to identify suspected data points since they would diminish predictive performance [71]. To detect suspected (outlier) data, the present study adopted the leverage approach, in which outliers are identified using the hat matrix H and critical leverage limit H^* [72]:

$$\begin{aligned} H &= U\left(U^T U\right)^{-1} U^T, \\ H^* &= \frac{3j}{i+1}, \end{aligned} \tag{12}$$

where U is an $i \times j$ matrix, i denotes the number of parameters, and j stands for the number of training data points [73, 74]. Figure 1 shows William's plot of the standardized residuals versus the hat value in order to evaluate spin-state splitting data accuracy. The reliable region is represented by a critical leverage limit along with standardized results

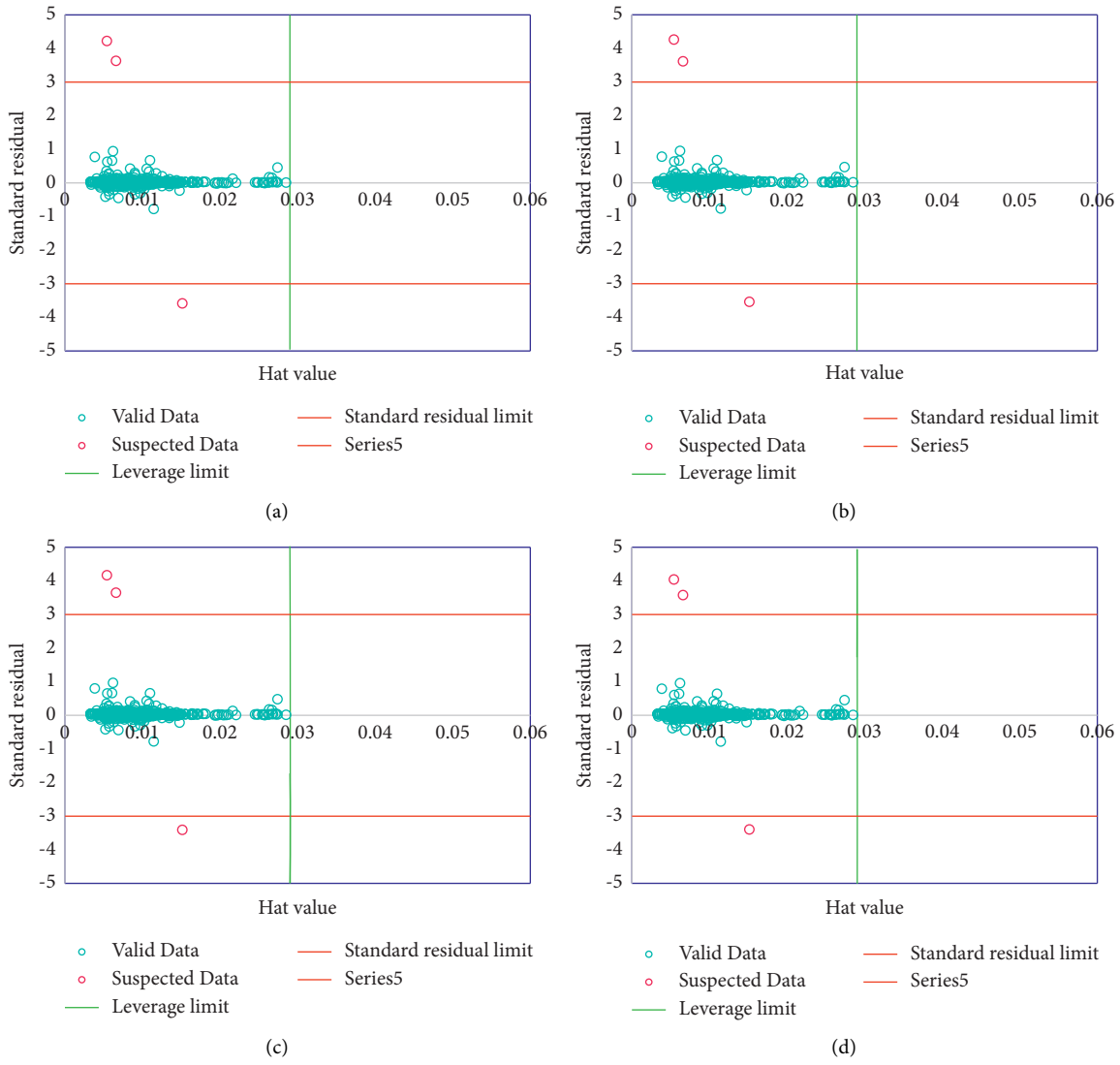

Figure 1: Analysis to determine suspicious points using different kernel functions of the GPR model. (a) Exponential. (b) Matern. (c) Squared exponential. (d) Rational quadratic.

ranging between −3 and +3. As shown, the dataset is concluded to be satisfactory for the model training and testing phases.

4. Results and Discussion

To measure the performance of the model, the present work utilized statistical parameters to evaluate the consistency between the empirical data and the model estimates. Table 1 provides the comparison between the estimates and empirical data. The coefficient of determination was obtained to be 0.985, 0.984, 0.986, and 0.986 for the rational quadratic, squared exponential, Matern, and exponential kernels, respectively. According to the STD, RMSE, MSE, and MRE values, the GPR models showed satisfactory training performance. Moreover, the models should predict spin-state

Table 1: Different statistical analyzes on modeled data with different kernels.

Model	Phase	R^2	MRE (%)	MSE	RMSE	STD
Exponential	Train	0.987	2.494	10.8342	1.5791	21.5100
	Test	0.983	2.679	11.8994	1.6367	19.8056
	Total	0.986	2.540	11.100	1.593	21.084
Matern	Train	0.987	2.797	13.7525	1.6725	27.5004
	Test	0.983	2.974	14.6242	1.7246	24.4109
	Total	0.986	2.841	13.970	1.685	26.729
Squared exponential	Train	0.985	3.287	18.7397	1.8130	34.2635
	Test	0.982	3.407	19.3987	1.8458	34.6710
	Total	0.984	3.317	18.904	1.821	34.365
Rational quadratic	Train	0.986	2.896	14.6700	1.7019	29.4079
	Test	0.983	3.074	15.2871	1.7533	26.2746
	Total	0.985	2.941	14.824	1.715	28.625

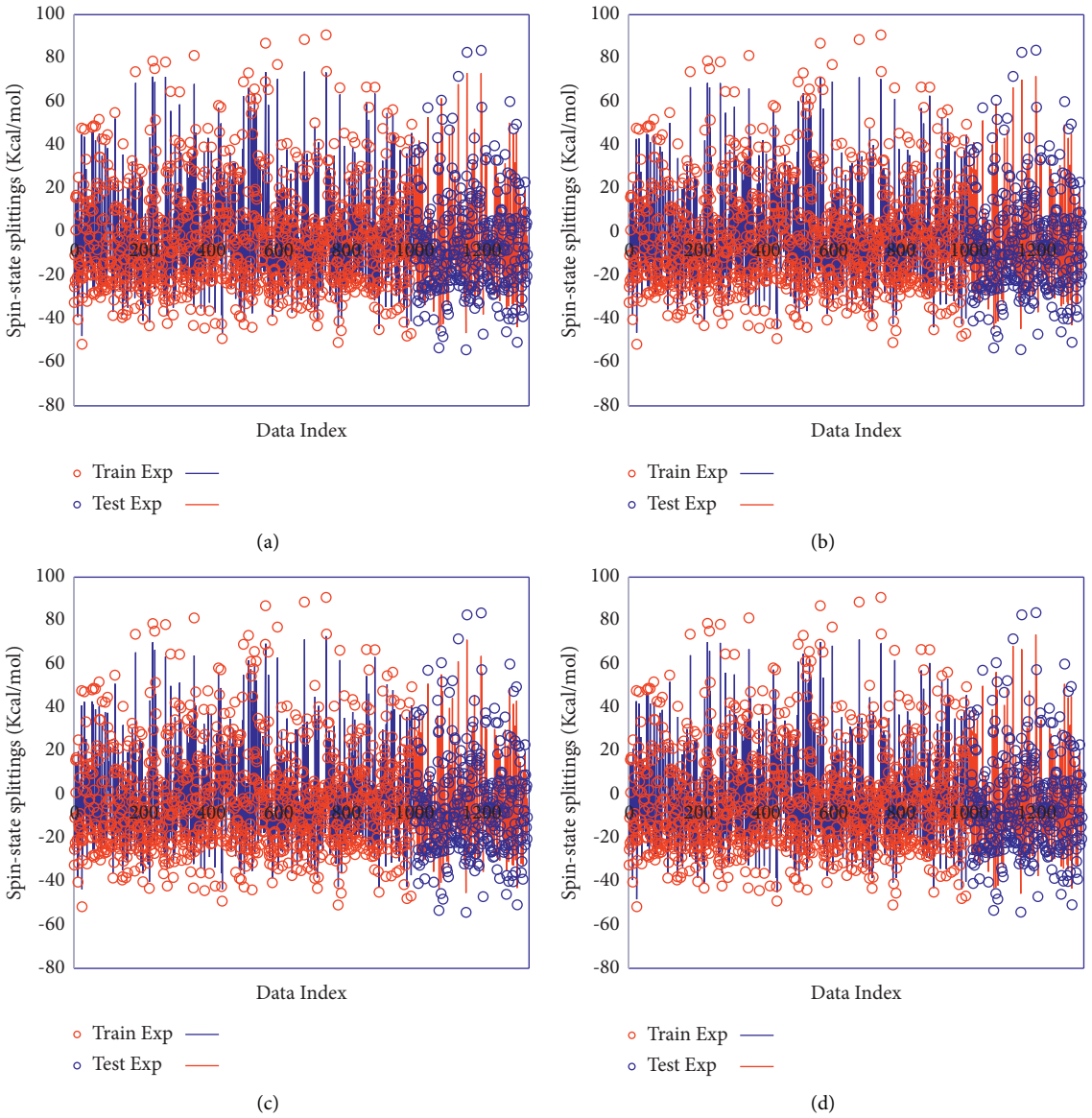

Figure 2: Simultaneous comparison of real data and its corresponding modeled data using different kernels of the GPR model. (a) Exponential. (b) Matern. (c) Squared exponential. (d) Rational quadratic.

splitting accurately. Hence, testing data were used to evaluate the models. The GPR models with the exponential and Matern kernels should have the highest spin-state splitting prediction performance.

Figure 2 shows the comparison between the empirical data and model estimates. As can be seen, the model estimates well agreed with the empirical spin-state splitting data, suggesting high accuracy for the proposed models. As a result, the GPR models can be claimed to have excellent performance in spin-state splitting estimation.

Figure 3 shows the comparison between the empirical data and the predictions of the models. The fitting of the predictions to the corresponding empirical data points was calculated to have correlation coefficients above 0.9816. The fit lines significantly cross the bisector line (45°) as the model accuracy measure. However, the model with the exponential

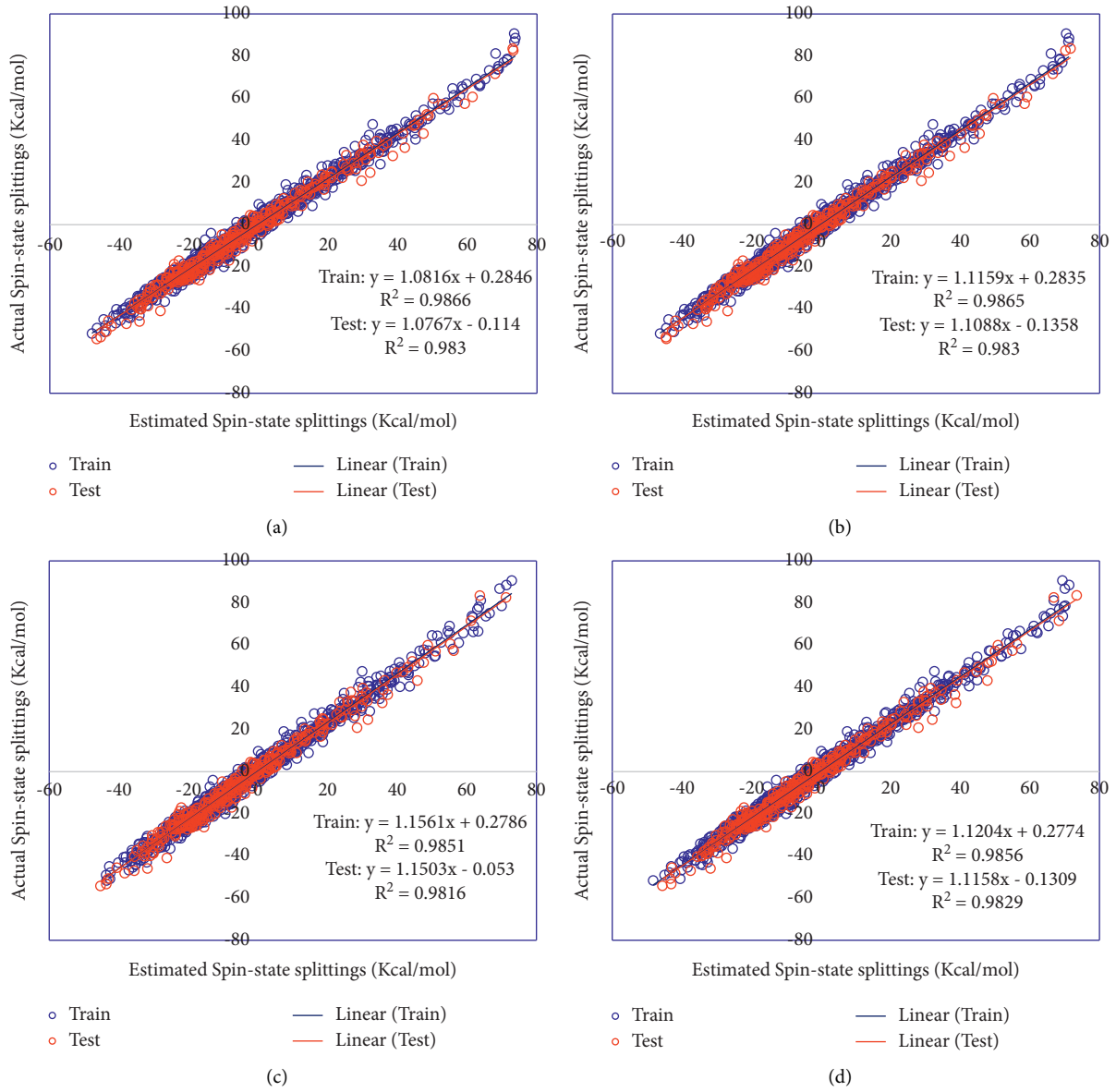

Figure 3: Regression analysis on modeled data designed with different kernels of the GPR model. (a) Exponential. (b) Matern. (c) Squared exponential. (d) Rational quadratic.

and Matern kernel functions showed the largest correlation and thus the highest performance.

The relative deviations of the empirical data and the estimates are shown in Figure 4. According to it, the absolute deviations of the Matern, rational quadratic, and squared exponential kernels were calculated to be below 2000%, whereas the exponential kernel showed an absolute deviation below 1500%.

The GPR models were found to be efficient and effective in the estimation of spin-state splitting. To ensure the spin-state splitting estimation performance of the proposed models with different MOFs, the models were compared to

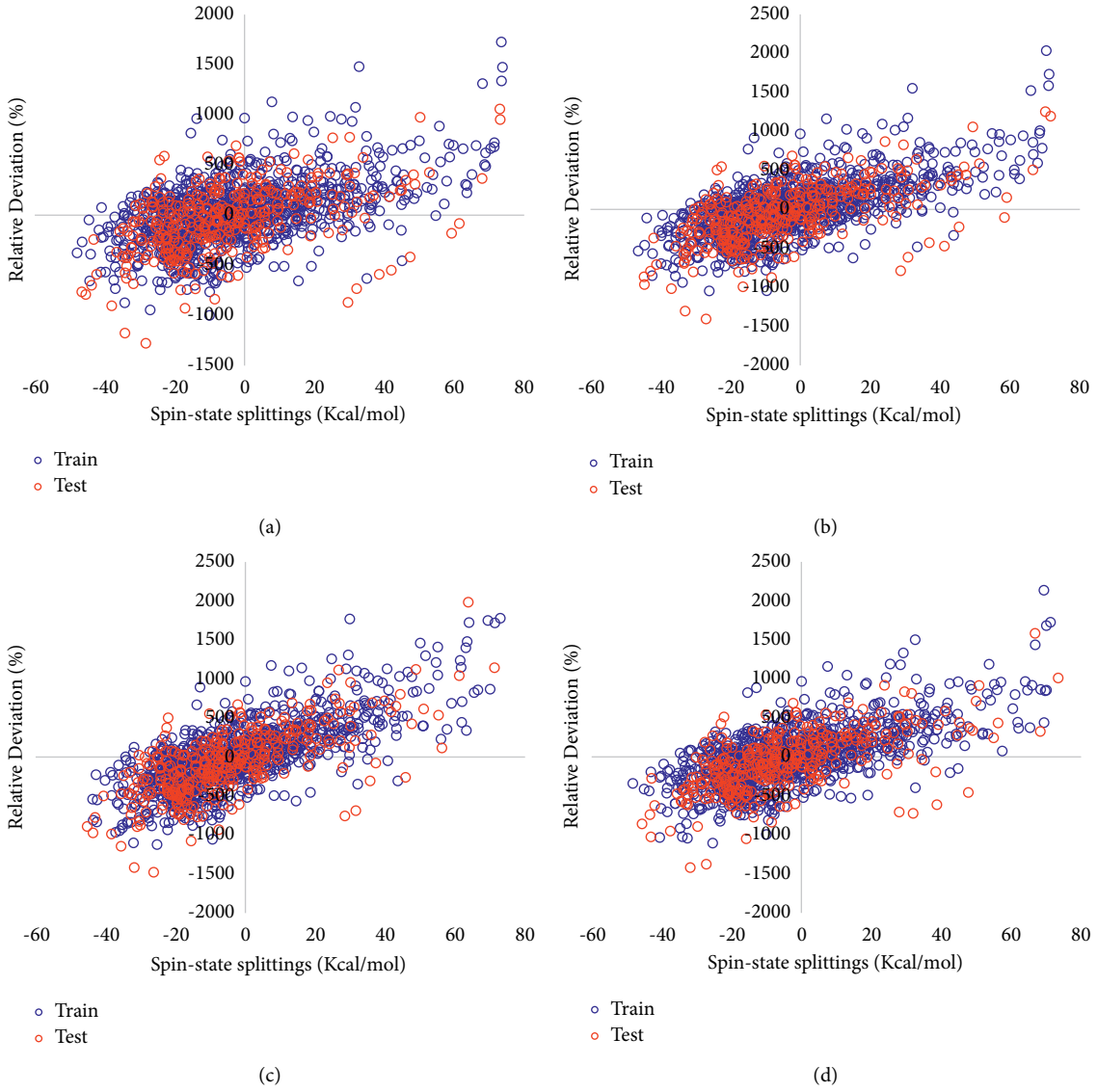

Figure 4: Relative deviation analysis on modeled data designed with different kernels of the GPR model. (a) Exponential. (b) Matern. (c) Squared exponential. (d) Rational quadratic.

earlier studies. Janet and his colleagues used the RMSE statistical parameter to compare LASSO, KRR, SVR, ANN, and KRR models in predicting this parameter [64]. By comparing their results with the results given in Table 1 of our study, it is proved that our proposed models have a higher ability to predict the target data.

5. Conclusion

The present study developed GPR models using four kernel functions, i.e., rational quadratic, Matern, exponential, and squared exponential kernels to evaluate spin-state splitting. As they showed good agreement with the empirical spin-state splitting data, the proposed models were concluded to have high performance. However, the GPR model with the exponential and Matern kernels showed the highest performance. Moreover, a comparison of the models to earlier works in the literature revealed that the proposed GPR models outperformed earlier models.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] J. K. Nørskov and T. Bligaard, *The Catalyst Genome*, Wiley Online Library, Hoboken, NJ, USA, 2013.

[2] A. Jain, S. P. Ong, G. Hautier et al., "Commentary: the materials Project: a materials genome approach to accelerating

materials innovation," *APL Materials*, vol. 1, no. 1, Article ID 011002, 2013.

[3] R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, T. D. Hirzel et al., "Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach," *Nature Materials*, vol. 15, no. 10, pp. 1120–1127, 2016.

[4] E. O. Pyzer-Knapp, K. Li, and A. Aspuru-Guzik, "Learning from the harvard clean energy project: the use of neural networks to accelerate materials discovery," *Advanced Functional Materials*, vol. 25, no. 41, pp. 6495–6502, 2015.

[5] A. M. Virshup, J. Contreras-García, P. Wipf, W. Yang, and D. N. Beratan, "Stochastic voyages into uncharted chemical space produce a representative library of all possible drug-like compounds," *Journal of the American Chemical Society*, vol. 135, no. 19, pp. 7296–7303, 2013.

[6] P. Kirkpatrick and C. Ellis, "Chemical space," *Nature*, vol. 432, no. 7019, pp. 823-824, 2004.

[7] B. Meredig, A. Agrawal, S. Kirklin et al., "Combinatorial screening for new materials in unconstrained composition space with machine learning," *Physical Review B*, vol. 89, no. 9, Article ID 094104, 2014.

[8] L. Li, J. C. Snyder, I. M. Pelaschier et al., "Understanding machine-learned density functionals," *International Journal of Quantum Chemistry*, vol. 116, no. 11, pp. 819–833, 2016.

[9] J. Behler, "Perspective: machine learning potentials for atomistic simulations," *The Journal of Chemical Physics*, vol. 145, no. 17, Article ID 170901, 2016.

[10] A. Lekomtsev, A. Keykhosravi, M. B. Moghaddam, R. Daneshfar, and O. Rezvanjou, "On the prediction of filtration volume of drilling fluids containing different types of nanoparticles by ELM and PSO-LSSVM based models," *Petroleum*, 2021, https://www.sciencedirect.com/science/article/pii/S2405656121000274.

[11] J. S. Smith, O. Isayev, and A. E. Roitberg, "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost," *Chemical Science*, vol. 8, no. 4, pp. 3192–3203, 2017.

[12] F. Mousazadeh, M. Naeem, R. Daneshfar, B. Soulgani, and M. Naseri, "Predicting the condensate viscosity near the wellbore by ELM and ANFIS-PSO strategies," *Journal of Petroleum Science and Engineering*, vol. 204, Article ID 108708, 2021.

[13] R. Setiawan, R. Daneshfar, O. Rezvanjou, S. Ashoori, and M. Naseri, "Surface tension of binary mixtures containing environmentally friendly ionic liquids: insights from artificial intelligence," *Environment, Development and Sustainability*, vol. 23, pp. 1–22, 2021.

[14] J. C. Snyder, M. Rupp, K. Hansen, K. R. Müller, and K. Burke, "Finding density functionals with machine learning," *Physical Review Letters*, vol. 108, no. 25, Article ID 253002, 2012.

[15] K. Mills, M. Spanner, and I. Tamblyn, "Deep learning and the Schrödinger equation," *Physical Review A*, vol. 96, no. 4, Article ID 042113, 2017.

[16] K. Yao and J. Parkhill, "Kinetic energy of hydrocarbons as a function of electron density and convolutional neural networks," *Journal of Chemical Theory and Computation*, vol. 12, no. 3, pp. 1139–1147, 2016.

[17] J. C. Snyder, M. Rupp, K. Hansen, L. Blooston, K. R. Müller, and K. Burke, "Orbital-free bond breaking via machine learning," *The Journal of Chemical Physics*, vol. 139, no. 22, Article ID 224104, 2013.

[18] K. Yao, J. E. Herr, and J. Parkhill, "The many-body expansion combined with neural networks," *The Journal of Chemical Physics*, vol. 146, no. 1, Article ID 014106, 2017.

[19] F. Häse, S. Valleau, E. Pyzer-Knapp, and A. Aspuru-Guzik, "Machine learning exciton dynamics," *Chemical Science*, vol. 7, no. 8, pp. 5139–5147, 2016.

[20] V. Botu and R. Ramprasad, "Adaptive machine learning framework to accelerate ab initio molecular dynamics," *International Journal of Quantum Chemistry*, vol. 115, no. 16, pp. 1074–1083, 2015.

[21] G. Pilania, J. E. Gubernatis, and T. Lookman, "Multi-fidelity machine learning models for accurate bandgap predictions of solids," *Computational Materials Science*, vol. 129, pp. 156–163, 2017.

[22] G. Pilania, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis, and T. Lookman, "Machine learning bandgaps of double perovskites," *Scientific Reports*, vol. 6, no. 1, Article ID 19375, 2016.

[23] T. Morawietz and J. Behler, "A density-functional theory-based neural network potential for water clusters including van der Waals corrections," *The Journal of Physical Chemistry A*, vol. 117, no. 32, pp. 7356–7366, 2013.

[24] T. Morawietz, A. Singraber, C. Dellago, and J. Behler, "How van der Waals interactions determine the unique properties of water," *Proceedings of the National Academy of Sciences*, vol. 113, no. 30, pp. 8368–8373, 2016.

[25] M. Rupp, A. Tkatchenko, K. R. Müller, and O. A. von Lilienfeld, "Fast and accurate modeling of molecular atomization energies with machine learning," *Physical Review Letters*, vol. 108, no. 5, Article ID 058301, 2012.

[26] B. Huang and O. A. Von Lilienfeld, "Communication: understanding molecular representations in machine learning: the role of uniqueness and target similarity," *The Journal of Chemical Physics*, vol. 145, no. 16, Article ID 161102, 2016.

[27] S. De, A. P. Bartók, G. Csányi, and M. Ceriotti, "Comparing molecules and solids across structural and alchemical space," *Physical Chemistry Chemical Physics*, vol. 18, no. 20, pp. 13754–13769, 2016.

[28] G. Maggiora, M. Vogt, D. Stumpfe, and J. Bajorath, "Molecular similarity in medicinal chemistry: miniperspective," *Journal of Medicinal Chemistry*, vol. 57, no. 8, pp. 3186–3204, 2014.

[29] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case, "Development and testing of a general amber force field," *Journal of Computational Chemistry*, vol. 25, no. 9, pp. 1157–1174, 2004.

[30] R. Deeth, "The ligand field molecular mechanics model and the stereoelectronic effects of d and s electrons," *Coordination Chemistry Reviews*, vol. 212, no. 1, pp. 11–34, 2001.

[31] P. Atkins and T. Overton, *Shriver and Atkins' Inorganic Chemistry*, Oxford University Press, Oxford, UK, 2010.

[32] K. T. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. R. Müller, and E. K. U. Gross, "How to represent crystal structures for machine learning: towards fast prediction of electronic properties," *Physical Review B*, vol. 89, no. 20, Article ID 205118, 2014.

[33] E. I. Ioannidis and H. J. Kulik, "Towards quantifying the role of exact exchange in predictions of transition metal complex properties," *The Journal of Chemical Physics*, vol. 143, no. 3, Article ID 034104, 2015.

[34] D. N. Bowman and E. Jakubikova, "Low-spin versus high-spin ground state in pseudo-octahedral iron complexes," *Inorganic Chemistry*, vol. 51, no. 11, pp. 6011–6019, 2012.

[35] T. Z. H. Gani and H. J. Kulik, "Where does the density localize? Convergent behavior for global hybrids, range separation, and DFT+ U," *Journal of Chemical Theory and Computation*, vol. 12, no. 12, pp. 5931–5945, 2016.

[36] E. I. Ioannidis and H. J. Kulik, "Ligand-field-dependent behavior of meta-GGA exchange in transition-metal complex spin-state ordering," *The Journal of Physical Chemistry A*, vol. 121, no. 4, pp. 874–884, 2017.

[37] W. Huang, D.-H. Xing, J.-B. Lu, B. Long, W. H. E. Schwarz, and J. Li, "How much can density functional approximations (DFA) fail? The extreme case of the FeO4 species," *Journal of Chemical Theory and Computation*, vol. 12, no. 4, pp. 1525–1533, 2016.

[38] J. J. P. Stewart, "Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters," *Journal of Molecular Modeling*, vol. 19, no. 1, pp. 1–32, 2013.

[39] R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Big data meets quantum chemistry approximations: the Δ-machine learning approach," *Journal of Chemical Theory and Computation*, vol. 11, no. 5, pp. 2087–2096, 2015.

[40] L. Shen, J. Wu, and W. Yang, "Multiscale quantum mechanics/molecular mechanics simulations with neural networks," *Journal of Chemical Theory and Computation*, vol. 12, no. 10, pp. 4934–4946, 2016.

[41] H. J. Kulik, "Perspective: treating electron over-delocalization with the DFT+ U method," *The Journal of Chemical Physics*, vol. 142, no. 24, Article ID 240901, 2015.

[42] A. J. Cohen, P. Mori-Sánchez, and W. Yang, "Insights into current limitations of density functional theory," *Science*, vol. 321, no. 5890, pp. 792–794, 2008.

[43] M. A. Halcrow, "Structure: function relationships in molecular spin-crossover complexes," *Chemical Society Reviews*, vol. 40, no. 7, pp. 4119–4142, 2011.

[44] J.-F. Létard, P. Guionneau, and L. Goux-Capes, "Towards spin crossover applications," *Spin Crossover in Transition Metal Compounds III*, pp. 221–249, Springer, Berlin, Germany, 2004.

[45] C. A. Bignozzi, R. Argazzi, R. Boaretto et al., "The role of transition metal complexes in dye sensitized solar devices," *Coordination Chemistry Reviews*, vol. 257, no. 9-10, pp. 1472–1492, 2013.

[46] K. Liu, Y. Li, X. Hu, M. Lucu, and W. Dhammika Widanage, "Gaussian process regression with automatic relevance determination kernel for calendar aging prediction of lithium-ion batteries," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 3767–3777, 2019.

[47] M. Taki, A. Rohani, F. Soheili-Fard, and A. Abdeshahi, "Assessment of energy consumption and modeling of output energy for wheat production by neural network (MLP and RBF) and Gaussian process regression (GPR) models," *Journal of Cleaner Production*, vol. 172, pp. 3028–3041, 2018.

[48] W. Gao, M. Karbasi, M. Hasanipanah, X. Zhang, and J. Guo, "Developing GPR model for forecasting the rock fragmentation in surface mines," *Engineering with Computers*, vol. 34, no. 2, pp. 339–345, 2018.

[49] J. Park, D. Lechevalier, R. Ak et al., "Gaussian process regression (GPR) representation in predictive model markup language (PMML)," *Smart and Sustainable Manufacturing Systems*, vol. 1, no. 1, pp. 121–141, 2017.

[50] J. Verrelst, J. Muñoz, L. Alonso et al., "Machine learning regression algorithms for biophysical parameter retrieval: opportunities for sentinel −2 and −3," *Remote Sensing of Environment*, vol. 118, pp. 127–139, 2012.

[51] A. Rohani, M. Taki, and M. Abdollahpour, "A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part: I)," *Renewable Energy*, vol. 115, pp. 411–422, 2018.

[52] N.-D. Hoang, A.-D. Pham, Q.-L. Nguyen, and Q.-N. Pham, "Estimating compressive strength of high performance concrete with Gaussian process regression model," *Advances in Civil Engineering*, vol. 2016, Article ID 2861380, 8 pages, 2016.

[53] D. Liu, J. Pang, J. Zhou, Y. Peng, and M. Pecht, "Prognostics for state of health estimation of lithium-ion batteries based on combination Gaussian process functional regression," *Microelectronics Reliability*, vol. 53, no. 6, pp. 832–839, 2013.

[54] Y. Chen, Y. B. Tian, Z. Qiang, and L. Xu, "Optimisation of reflection coefficient of microstrip antennas based on KBNN exploiting GPR model," *IET Microwaves, Antennas & Propagation*, vol. 12, no. 4, pp. 602–606, 2018.

[55] S. D. Suresh, A. Qasim, B. Lal, S. M. Imran, and K. S. Foo, "Application of Gaussian process regression (GPR) in gas hydrate mitigation," *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, vol. 88, no. 2, pp. 27–37, 2021.

[56] B. B. Sahoo, R. Jha, A. Singh, and D. Kumar, "Application of support vector regression for modeling low flow time series," *KSCE Journal of Civil Engineering*, vol. 23, no. 2, pp. 923–934, 2019.

[57] M. Karbasi, "Forecasting of multi-step ahead reference evapotranspiration using wavelet-Gaussian process regression model," *Water Resources Management*, vol. 32, no. 3, pp. 1035–1052, 2018.

[58] D. V. Dao, H. Adeli, H.-B. Ly et al., "A sensitivity and robustness analysis of GPR and ANN for high-performance concrete compressive strength prediction using a Monte Carlo simulation," *Sustainability*, vol. 12, no. 3, p. 830, 2020.

[59] N. S. Raghavendra and P. C. Deka, "Multistep ahead groundwater level time-series forecasting using Gaussian process regression and ANFIS," in *Advanced Computing and Systems for Security*, pp. 289–302, Springer, Berlin, Germany, 2016.

[60] M. Jamei, I. Ahmadianfar, I. Adewale Olumegbon, M. Karbasi, and A. Asadi, "On the assessment of specific heat capacity of nanofluids for solar energy applications: application of Gaussian process regression (GPR) approach," *Journal of Energy Storage*, vol. 33, Article ID 102067, 2021.

[61] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, "Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, June 2015.

[62] C. K. Arthur, V. A. Temeng, and Y. Y. Ziggah, "Novel approach to predicting blast-induced ground vibration using Gaussian process regression," *Engineering with Computers*, vol. 36, no. 1, pp. 29–42, 2020.

[63] A. R. Ghanizadeh, N. Heidarabadizadeh, and F. Heravi, "Gaussian process regression (GPR) for auto-estimation of resilient modulus of stabilized base materials," *Journal of Soft Computing in Civil Engineering*, vol. 5, no. 1, pp. 80–94, 2021.

[64] J. P. Janet and H. J. Kulik, "Predicting electronic structure properties of transition metal complexes with neural networks," *Chemical Science*, vol. 8, no. 7, pp. 5137–5152, 2017.

[65] M. Gheytanzadeh, A. Baghban, S. Habibzadeh, A. Mohaddespour, and O. Abida, "Insights into the estimation of capacitance for carbon-based supercapacitors," *RSC Advances*, vol. 11, no. 10, pp. 5479–5486, 2021.

[66] A. Bemani, A. Baghban, S. Shamshirband, A. Mosavi, P. Csiba, and A. R. Várkonyi-Kóczy, "Applying ANN, ANFIS, and LSSVM models for estimation of acid solvent solubility in supercritical CO_2," 2019, https://arxiv.org/abs/1912.05612.

[67] A. Rostami, A. Baghban, A. H. Mohammadi, A. Hemmati-Sarapardeh, and S. Habibzadeh, "Rigorous prognostication of permeability of heterogeneous carbonate oil reservoirs: smart modeling and correlation development," *Fuel*, vol. 236, pp. 110–123, 2019.

[68] S. Alizadeh, I. Alruyemi, R. Daneshfar, M. Mohammadi-Khanaposhtani, and M. Naser, "An insight into the estimation of drilling fluid density at HPHT condition using PSO-, ICA-, and GA-LSSVM strategies," *Scientific Reports*, vol. 11, no. 1, pp. 1–14, 2021.

[69] M. Gheytanzadeh, A. Baghban, S. Habibzadeh et al., "Towards estimation of CO2 adsorption on highly porous MOF-based adsorbents using Gaussian process regression approach," *Scientific Reports*, vol. 11, no. 1, pp. 1–13, 2021.

[70] F. Keivanimehr, A. Baghban, S. Habibzadeh et al., "Oxidation kinetics of water contaminants: new insights from artificial intelligence," *Environmental Progress & Sustainable Energy*, vol. 40, no. 1, Article ID e13491, 2021.

[71] Y. Yang, S. Zheng, Z. Ai, and M. M. Molla Jafari, "On the prediction of biogas production from vegetables, fruits, and food wastes by ANFIS-and LSSVM-based models," *BioMed Research International*, vol. 2021, Article ID 9202127, 8 pages, 2021.

[72] B. Xu, T.-C. Chen, D. Ahangari, S. M. Alizadeh, M. Elveny, and J. Makhdoumi, "Application of a supervised learning machine for accurate prognostication of hydrogen contents of bio-oil," *International Journal of Chemical Engineering*, vol. 2021, Article ID 7548251, 8 pages, 2021.

[73] N. Kardani, A. Bardhan, B. Roy et al., "A novel improved Harris Hawks optimization algorithm coupled with ELM for predicting permeability of tight carbonates," *Engineering with Computers*, pp. 1–24, 2021, https://www.springerprofessional.de/en/a-novel-improved-harris-hawks-optimization-algorithm-coupled-wit/19393262.

[74] N. Kardani, A. Bardhan, P. Samui, M. Nazem, A. Zhou, and D. Jahed Armaghani, "A novel technique based on the improved firefly algorithm coupled with extreme learning machine (ELM-IFF) for predicting the thermal conductivity of soil," *Engineering with Computers*, pp. 1–20, 2021, https://www.springerprofessional.de/en/a-novel-improved-harris-hawks-optimization-algorithm-coupled-wit/19393262.