Journal of Molecular Liquids 358 (2022) 119208

Contents lists available at ScienceDirect

Journal of Molecular Liquids

journal homepage: www.elsevier.com/locate/molliq

Development and validation of machine learning models for prediction of nanomedicine solubility in supercritical solvent for advanced pharmaceutical manufacturing

Wenlin Liu [a,*], Ruijuan Zhao [a], Xiankun Su [a], Abdullah Mohamed [b], Tazeddinova Diana [c,d]

[a] Guizhou Academy of Tobacco Science, Guiyang 550000, China
[b] Research Centre, Future University in Egypt, New Cairo 11845, Egypt
[c] Department of Technology and Catering Organization, South Ural State University, Chelyabinsk, Russian Federation
[d] Zhangir Khan Agrarian Technical University, Uralsk, Kazakhstan

ARTICLE INFO

Article history:
Received 16 March 2022
Revised 11 April 2022
Accepted 19 April 2022
Available online 25 April 2022

Keywords:
Modeling
Drug solubility
Supercritical process
Drug manufacture
Nanomedicine

ABSTRACT

Salsalate solubility in supercritical carbon dioxide was studied in this research by computational simulation to correlate the solubility to input parameters including temperature and pressure. A dataset was collected from resources and the models were correlated to the data. Both training and validation steps have been performed to implement the computational tasks. Indeed, we are dealing with a dataset with two P and T inputs and one output (drug solubility), and 32 data points. We chose three models based on Gaussian Process Regression (GPR) including simple (raw) GPR, Ada-boosted GPR, and Bagged GPR as two ensemble methods for correlation of the solubility data. All hyper-parameters were tuned for more general models and models evaluated with standard metrics. The GPR, Adaboost + GPR, and Bagging + GPR have scores of 0.9779, 0.9992, and 0.9795 using R-squared, respectively. Also, in terms of RMSE models have error rates of 1.25×10^{-4}, 1.20×10^{-4}, and 1.29×10^{-4}, respectively. Finally, considering the standard criteria and visual analysis, the boosted GPR model is selected as the main model. The optimal values found as (P = 400, T = 338.0, Y = 0.003879).

1. Introduction

Development of supercritical processing and its application in pharmaceutical manufacturing has gained momentum recently, due to the superior characteristics of supercritical processing which can be considered as green technology [1–4]. The technology is attractive as it utilizes mainly CO_2 as the solvent for the processing drug formation such as solid dosage oral formulations. The drug particles can be prepared in nanosized which can enhance its solubility in aqueous solutions specifically for poorly water-soluble drugs [5–11]. Basically, the drug particles at nano scale provide higher free energy due to their high surface area which is favorable for enhanced solubility in aqueous media [12–18]. For preparation of nanomedicine, supercritical technology using CO_2 as solvent can be employed by which the drug solubility in the solvent is a crucial parameter prior to design of process and operation [19,20]. The solubility can be measured by different techniques among which gravimetric method is the major method due to its simplicity and reliability [21,22].

The solubility of drugs in supercritical solvents can be also interrogated using computational techniques such as thermodynamic models, semi-empirical correlations, and machine learning techniques. Recently, semi-empirical methods have been extensively used for correlation of drug solubility in supercritical CO_2 to the operational parameters, i.e., pressure and temperature [19,21–28]. Also, it has been reported that machine learning models can be successfully used in correlation of drug solubility in supercritical carbon dioxide. In terms of fitting accuracy, machine learning models have shown to be more accurate compared to other models such as thermodynamic models [29,30]. Therefore, development of machine learning models for prediction of drug solubility is a great subject of research with practical application in pharmaceutical manufacturing of solid dosage oral formulations [31–36].

Machine learning (ML) methods can be used like other computational techniques for prediction and understanding the relationship between the process parameters and response in physical, chemical, and biochemical processes [31,37–45]. These techniques

* Corresponding author.
E-mail address: lwl225@163.com (W. Liu).

https://doi.org/10.1016/j.molliq.2022.119208
0167-7322/

use database technology, statistical analysis, and some other disciplines to extract information, correlations, and hidden patterns of different datasets [46,47]. Modeling and forecasting are often accomplished using two strategies. The classic approach is built on a single independent model, whereas the modern ensemble methods such as bagging, and boosting are built on several models [48]. The established results show that these learning methods outperform the typical individual ML models [49–51]. The ensemble models initially train numerous weak (base) predictors with training data and then combine them to form a strong predictor [52–54]. In this study, we used kriging or Gaussian process regression both as a standalone model and base model for bagging and boosting models.

The Gaussian process regression model is a helpful nonparametric Bayesian model that can be employed in exploration and exploitation. One of GPR's key advantages is its ability to process an accurate response to a model's input variables. This method may display a vast variety of correlations between input features and output targets by employing a theoretically boundless quantity of input features and letting the data to control the complexity of predictors with Bayesian inference [55–58].

Boosting is a popular and necessary approach in ensemble learning, also known as improved learning. Boosting improves prediction outcomes by incorporating key factors. AdaBoost is a common Boosting approach that combines the estimations of multiple weak learners to create more efficient and applicable estimates. AdaBoost is a repeated strategy that trains a large number of base (weak) estimators for the whole training data set and then aggregates these base estimators to create a more robust final robust learner [51,59,60].

Furthermore, Bagging is a widely used ensembles for modeling that uses the bootstrap sample selection (for developing base models) strategy and aggregating advantages. The initial training set replaces part models during the bagging process. In this method, a data point may be used in training multiple weak predictors and some other data points may never be used. The final result is then computed by averaging the weak model outputs [61,62].

2. Data set for the computational study

The experimental data set contains 32 data points, each with two P and T inputs and one output (Table 1). This dataset is the same as reported in [63] for the solubility of Salsalate drug in supercritical CO_2 at different operational conditions as listed in Table 1. Indeed, we have taken the data and used them for training and validation in this work. Also, the pairwise Pearson correlation between input and outputs is shown in Fig. 1.

Table 1
The whole data set used in this work [63].

No	x1 = P	x2 = T	y
1	120	3.08×10^2	1.07×10^{-4}
2	120	3.18×10^2	7.07×10^{-5}
3	120	3.28×10^2	6.12×10^{-5}
4	120	3.38×10^2	3.77×10^{-5}
5	160	3.08×10^2	1.98×10^{-4}
6	160	3.18×10^2	2.17×10^{-4}
7	160	3.28×10^2	1.87×10^{-4}
8	160	3.38×10^2	1.66×10^{-4}
9	200	3.08×10^2	2.47×10^{-4}
10	200	3.18×10^2	3.59×10^{-4}
11	200	3.28×10^2	5.14×10^{-4}
12	200	3.38×10^2	5.93×10^{-4}
13	240	3.08×10^2	2.95×10^{-4}
14	240	3.18×10^2	4.83×10^{-4}
15	240	3.28×10^2	6.85×10^{-4}
16	240	3.38×10^2	1.02×10^{-3}
17	280	3.08×10^2	3.88×10^{-4}
18	280	3.18×10^2	6.74×10^{-4}
19	280	3.28×10^2	1.05×10^{-3}
20	280	3.38×10^2	1.60×10^{-3}
21	320	3.08×10^2	4.71×10^{-4}
22	320	3.18×10^2	8.59×10^{-4}
23	320	3.28×10^2	1.39×10^{-3}
24	320	3.38×10^2	2.11×10^{-3}
25	360	3.08×10^2	5.35×10^{-4}
26	360	3.18×10^2	9.58×10^{-4}
27	360	3.28×10^2	1.69×10^{-3}
28	360	3.38×10^2	2.52×10^{-3}
29	400	3.08×10^2	5.77×10^{-4}
30	400	3.18×10^2	1.28×10^{-3}
31	400	3.28×10^2	2.25×10^{-3}
32	400	3.38×10^2	3.88×10^{-3}

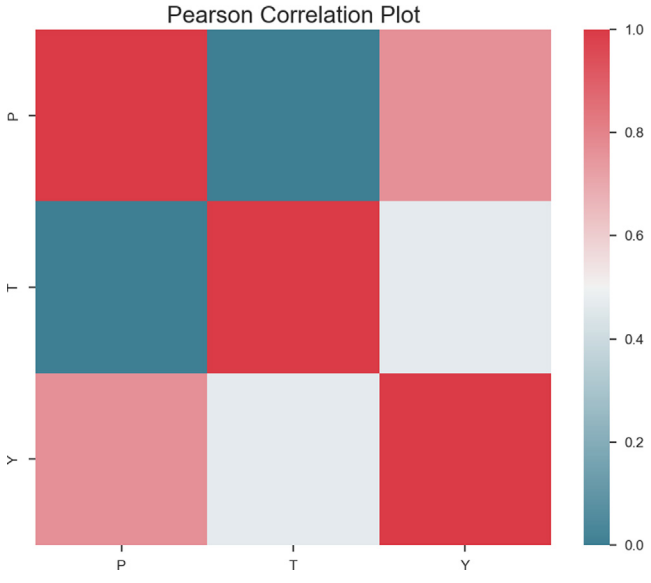

Fig. 1. Pairwise correlation plot.

3. Computational method

3.1. Gaussian process regression (GPR)

Non-parametric learning algorithms have grown in prominence in machine learning over the last decade, particularly the Bayesian approach for Gaussian Process (GP) models [64]. GPR, a complete Bayesian learning method, is used in a variety of applications including multivariate regression, experiment design, and model approximation [65,66]. A process is called Gaussian Process if the model outputs' joint probability distribution is Gaussian. The fundamental profit of GPR is that it incorporates several various ML tasks such as uncertainty estimation, model development, and tuning of hyper-parameters [67–69].

GPR operates under a probabilistic framework, which takes a training dataset $D = [(y_i,x_i)\ n = 1,\ldots,I]$ including I input data points $x_i \epsilon R^d$ as input and output y_i (a noisy scalar), and develops a predictor that generalizes to the distribution of the target value for at unseen data points. Using the GPR, the computation of y_i is acquired as follows:

$$y_i = f(x_i) + \varepsilon_i$$

It is important to mention that Gaussian Process Regressor assumes the latent function values $f(x_i)$ as random variables, and the relevant input variables x_i are used to index them .ε_i indicates the Gaussian noise (variance σ^2_{noise} and mean = 0) i.e., $\varepsilon_i\ N(0, \sigma^2_{noise})$.

The fundamental benefit of employing the Gaussian prior assumption is that functions may be defined easily using a mean function *m(x)* and a covariance function *cov(x, x′)*. Let x^* reflect a random sample of input variables; consequently, the output variable is estimated by the predictive probability distribution $p(y^* | X, y, x)$ with mean and variance [70]:

$$\widehat{y}* = m(x*) + k_*^T (K + \sigma_n^2 I)^{-1} (y - m(x*))$$

$$\sigma_{y*}^2 = k_* + \sigma_n^2 - k_*^T (K + \sigma_n^2 I)_{k*}^{-1}$$

Here, I denotes the identity matrix, k_* stands for some vectors which is defined by $[k_*]_i = \mathrm{cov}(x_i, x*)$, $k_* = \mathrm{cov}(x_i, x*)$ and K shows a covariance matrix of $[K]_{i,j} = cov(x_i, x_j)$. When the aforementioned descriptions are taken into account, it is evident that the model output is dependent on the training group *X, y*, as opposed to canonical regression approaches in which the parameters are only used to derive the model estimation [70].

To acquire reliable estimations, the covariance and mean function parameters are retrieved from the provided dataset. As a result, these parameters, known colloquially as hyper-parameters, completely dictate the predictive probability distribution. The hyper-parameter values are calculated by maximizing the log-likelihood function of the training group [70]:

$$\log p(y|X) = -\frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} y - \frac{1}{2} \log(|K + \sigma_n^2 I|) - \frac{n}{2} \log(2\pi)$$

Here, n is the quantity of training subsets.

3.2. Bagging and boosting ensembles

Machine learning (ML) techniques use ensemble techniques to improve recognition and prediction reliability. These methods frequently help to alleviate the over-fitting problem by integrating and aggregating numerous weak learners. Creating multiple sub-models' components (base predictors) (1, … , M) by intelligently updating training data can be used for developing some better learners. Particularly, the ideal parametric or predictive model can be obtained by combining qualifying sub-models using averaging or voting combination procedures [71].

The boosting approach, like the bagging technique, creates an ensemble model that results in the construction of some of learners with more accuracy than a single model.

In this study, we used a simple bagging ensemble on the GPR model alongside Adaboost as a boosting ensemble on the same model. Adaboost [72] updates the distribution of the training data set. In condition when εt is the miss-estimation rate at iteration t, next the weights for miss-estimated examples in the training subset will be modified by the factor $\beta t = \frac{1-\varepsilon t}{\varepsilon t}$. The whole sum of the modified weights is set to one. The estimators *C1, C2, … , Ct* are combined based on the weighted averaging. *Ct* is weighted by $\log(\beta t)$. The iterations will be finished if εt N > 0.5 and iteration T is altered to $t - 1$. If $\varepsilon t = 0$, then trial T becomes t. As a result, the new ensemble estimators developed by Adaboost are forced to focus on more difficult instances. The performance of each estimator is weighted in the voting for the final outcome [50,51,73].

4. Results and discussion

The explained models in the previous sections have been utilized in order to correlate the solubility of Salsalate to the pressure and temperature. Ensemble model hyper-parameters can be correlated with the appropriate quantity of weak predictors or learning rates, as well as other such parameters that especially influence ensemble techniques. The three introduced models were tuned to their hyper-parameters using the test and evaluated by standard criteria. These criteria are presented below. There are also hyper-parameters related directly to GPR such as alpha value. The three selected GPR-based models were tuned to their hyper-parameters and evaluated using standard criteria using the test. These standards are outlined below.

Certainty, R^2 is the most often utilized criterion for judging the performance success of estimation outcomes. It displays how closely the patterns of the estimated results correspond to the trends of the observed data [74]:

$$R^2 = 1 - \frac{\sum (y_i - x_i)^2}{\sum \left(x_i - \bar{x}_i\right)^2}$$

The standard deviation of the difference between the predicted and expected outputs is used to calculate the RMSE. It provides a measure of predictability for the expected values [75].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2}$$

MAPE is a widely used metric due to the scale independence and interpretability. It is a statistical metric used to determine the performance of an estimation model. MAPE defines error size as a float between 0 and 1 or as a percentage [76]:

$$MAPE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{x_i - y_i}{x_i} \right|$$

y_i and x_i indicate estimated and expected outputs, respectively. $\underline{x_i}$ is the average of expected data. n reflects the size of dataset.

The above criteria have been implemented for the developed models and the statistical parameters are listed in Table 2 for three models. It is clearly seen that the model of Adaboost + GPR has the highest value of R^2 which is equal to 0.999 confirming the fitting accuracy of this model compared to the other two models, as listed in Table 2.

Figs. 2–4 display the comparison of actual (expected) and predicted values of three implemented models. Comparing them, we can see BAGGING + GPR and AdaBoost + GPR have the relatively same result, and both are better than the GPR model (without ensembles). According to these facts and results in Table 2 finally, AdaBoost + GPR was selected as the most accurate model in this study among other models. It is also seen that all models have great capability in prediction of Salsalate solubility in supercritical CO_2 at various temperatures and pressures.

Furthermore, the optimum points obtained by the model (AdaBoost + GPR) are presented in Table 3. Also, the 3D and 2D plots of the drug solubility as function of pressure and temperature are illustrated in Figs. 5–7. Moreover, the residual of fitting is represented in Fig. 8. It is illustrated that the solubility of Salsalate in the solvent is enhanced with both temperature and pressure. However, it is seen that the influence of temperature (x2) is more significant compared to the effect of pressure (x1) on the drug

Table 2
Final model results.

Models	MAPE	RMSE	R^2
GPR	7.2389E-02	1.2500E-04	0.9779
Adaboost + GPR	7.2831E-02	1.2038E-04	0.9992
Bagging + GPR	7.329E-02	1.2945E-04	0.9795

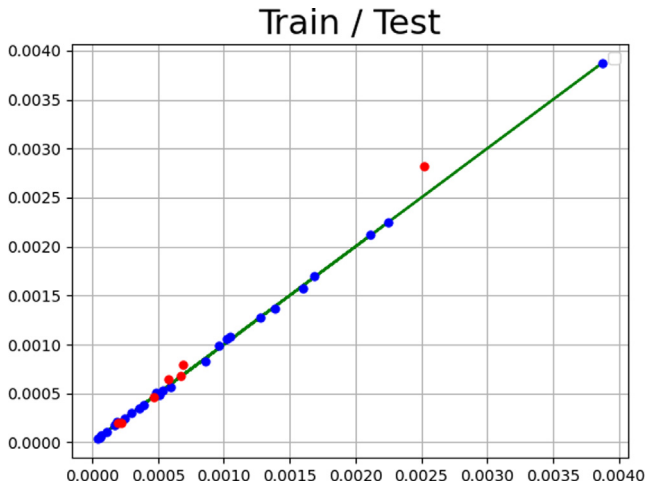

Fig. 2. Actual versus GPR predicted.

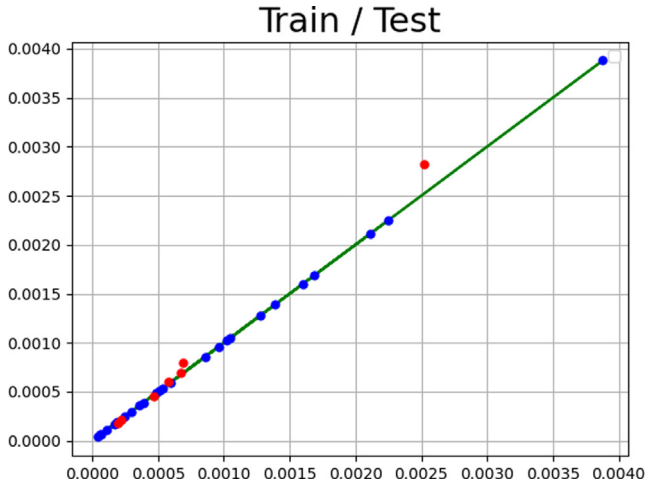

Fig. 3. Actual versus ADABOOST + GPR predicted.

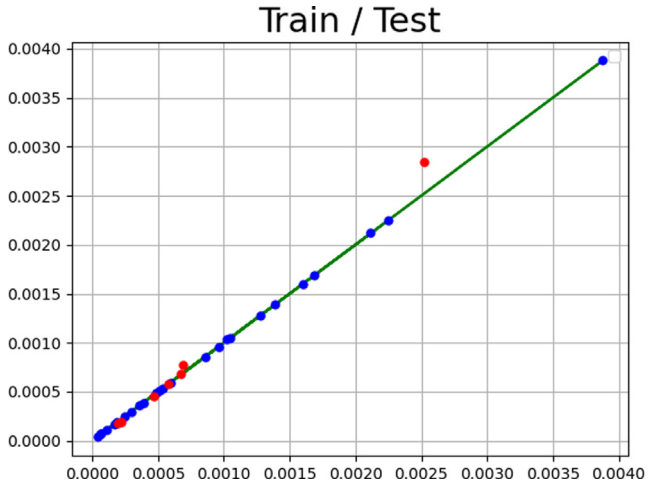

Fig. 4. Actual versus BAGGING + GPR predicted.

Table 3
Final optimal values of the paramours for maximum response.

X1 = P	X2 = T	Y
400	338.0	0.003879

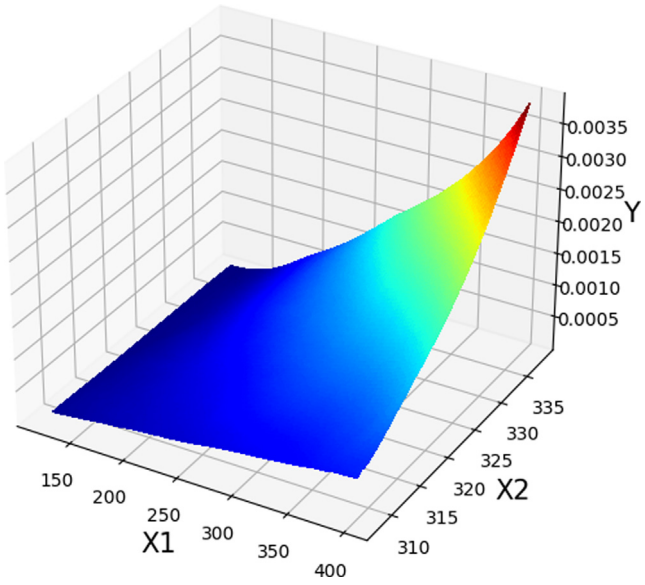

Fig. 5. 3D projection of inputs/outputs.

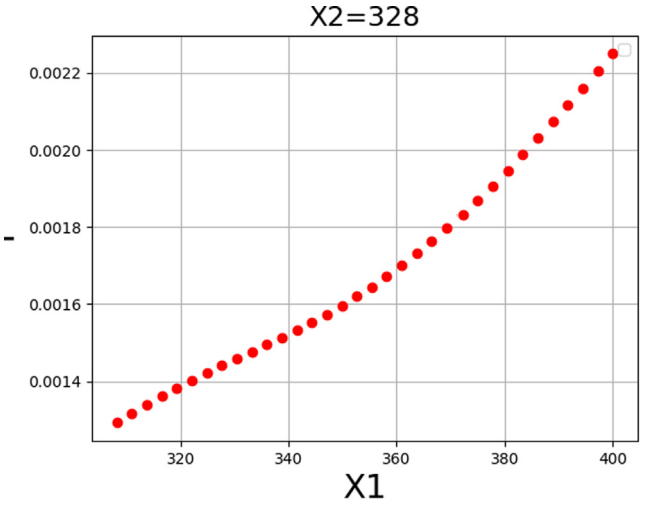

Fig. 6. Trends for X1.

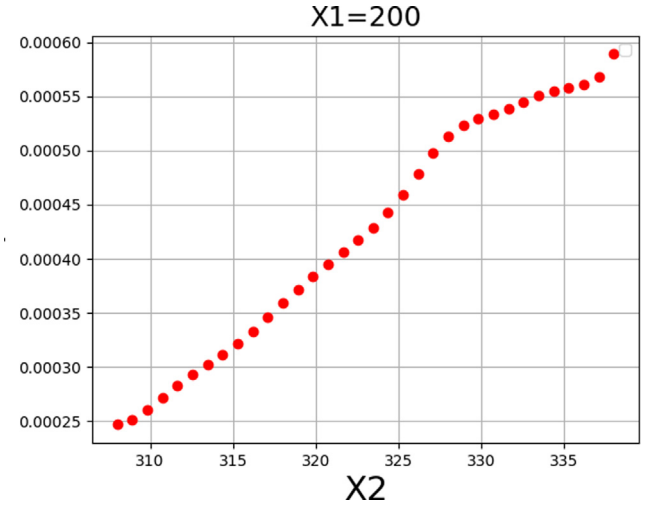

Fig. 7. Trends for X2.

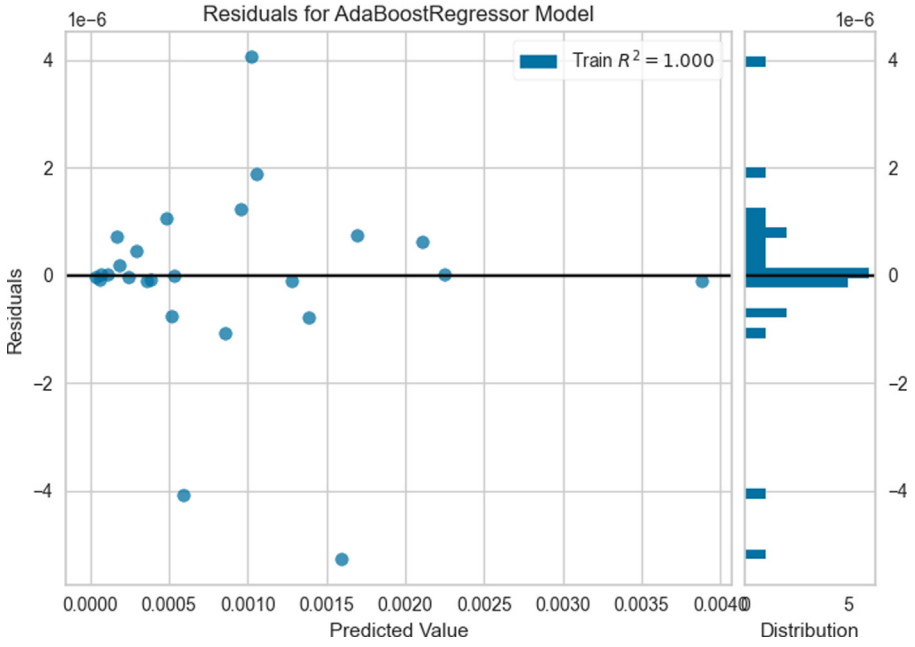

Fig. 8. Residuals for ADABOOST + GPR.

solubility. It is also indicated that the highest Salsalate solubility is obtained at the maximum values of pressure and temperature as listed in Table 3.

5. conclusion

We developed computational models for estimation of drug solubility in supercritical carbon dioxide as the green solvent with view of nanomedicine production. Salsalate was considered as the model drug in this work. We are dealing with a dataset containing two P and T inputs, one output, and 32 data points in this study. As two ensemble approaches, we used three Gaussian Process Regression (GPR) models: simple (raw) GPR, Ada-boosted GPR, and Bagged GPR. All hyperparameters have been tweaked for a more general model, and models have been evaluated using standard metrics. In terms of R-squared, the GPR, Adaboost + GPR, and Bagging + GPR have values of 0.9779, 0.9992, and 0.9795, respectively. Furthermore, the RMSE models have error ratings of 1.25E-04, 1.20E-04, and 1.29E-04, respectively. As the third metric, in MAPE models have 7.2389E-02, 7.2831E-02, and 7.329E-02, respectively. Finally, the boosted GPR model is chosen as the main model based on the usual criteria and visual analysis. The ideal values were discovered to be (P = 400, T = 338.0, Y = 0.003879).

CRediT authorship contribution statement

Wenlin Liu: Writing – original draft, Methodology, Resources, Validation. **Ruijuan Zhao:** Conceptualization, Formal analysis, Writing – review & editing. **Xiankun Su:** Writing – original draft, Conceptualization, Software. **Abdullah Mohamed:** Formal analysis, Validation, Supervision. **Tazeddinova Diana:** Writing – review & editing, Resources, Formal analysis.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

Tazeddinova Diana is thankful to the South Ural State University (Russian Federation) and Zhangir Khan Agrarian Technical University (Republic of Kazakhstan) for their help and support in completing this work.

References

[1] A. Zeinolabedini Hezave, H. Rajaei, M. Lashkarbolooki, F. Esmaeilzadeh, Analyzing the solubility of fluoxetine hydrochloride in supercritical carbon dioxide, J. Supercrit. Fluids 73 (2013) 57–62.
[2] B.-Q. Chen, R.K. Kankala, S.-B. Wang, A.-Z. Chen, Continuous nanonization of lonidamine by modified-rapid expansion of supercritical solution process, J. Supercrit. Fluids 133 (2018) 486–493.
[3] A.Z. Hezave, F. Esmaeilzadeh, Crystallization of micro particles of sulindac using rapid expansion of supercritical solution, J. Cryst. Growth 312 (22) (2010) 3373–3383.
[4] S.M. Abuzar, S.-M. Hyun, J.-H. Kim, H.J. Park, M.-S. Kim, J.-S. Park, S.-J. Hwang, Enhancing the solubility and bioavailability of poorly water-soluble drugs using supercritical antisolvent (SAS) process, Int. J. Pharm. 538 (1–2) (2018) 1–13.
[5] Q. Zou, P. Xing, L. Wei, B. Liu, Gene2vec: gene subsequence embedding for prediction of mammalian N6-methyladenosine sites from mRNA, RNA 25 (2) (2019) 205–218.
[6] C. Huang et al., Gold nanoparticles-loaded polyvinylpyrrolidone/ethylcellulose coaxial electrospun nanofibers with enhanced osteogenic capability for bone tissue regeneration, Mater. Des. 212 (2021) 110240.
[7] C. Huang, Y. Su, J. Shi, C. Yuan, S. Zhai, Q. Yong, Revealing the effects of centuries of ageing on the chemical structural features of lignin in archaeological fir woods, New J. Chem. 43 (8) (2019) 3520–3528.
[8] R. Shaikh et al., Chapter 2 – The development of a pharmaceutical oral solid dosage forms, in: R. Singh, Z. Yuan (Eds.), Computer Aided Chemical Engineering, Elsevier, 2018, pp. 27–65..
[9] R. Shaikh, G.M. Walker, D.M. Croker, Continuous, simultaneous cocrystallization and formulation of Theophylline and 4-Aminobenzoic acid pharmaceutical cocrystals using twin screw melt granulation, Eur. J. Pharm. Sci. 137 (2019) 104981.
[10] S. Shirazian, H.Y. Ismail, M. Singh, R. Shaikh, D.M. Croker, G.M. Walker, Multi-dimensional population balance modelling of pharmaceutical formulations for continuous twin-screw wet granulation: Determination of liquid distribution, Int. J. Pharm. 566 (2019) 352–360.
[11] R. Shaikh et al., Understanding solid-state processing of pharmaceutical cocrystals via milling: Role of tablet excipients, Int. J. Pharm. 601 (2021) 120514.
[12] H. Dong, L. Zheng, P. Yu, Q. Jiang, Y. Wu, C. Huang, B. Yin, Characterization and application of lignin-carbohydrate complexes from lignocellulosic materials as antioxidants for scavenging in vitro and in vivo reactive oxygen species, ACS Sust. Chem. Eng. 8 (1) (2020) 256–266.

[13] T. Li, W. Yin, S. Gao, Y. Sun, P. Xu, S. Wu, H. Kong, G. Yang, G. Wei, The combination of two-dimensional nanomaterials with metal oxide nanoparticles for gas sensors: a review, Nanomaterials 12 (6) (2022) 982.
[14] B. Bai et al., Coupled thermo-hydro-mechanical mechanism in view of the soil particle rearrangement of granular thermodynamics, Comput. Geotech. 137 (2021) 104272.
[15] M. Latif, D. Zhang, G. Gamez, Flowing atmospheric-pressure afterglow drift tube ion mobility spectrometry, Anal. Chim. Acta 1163 (2021) 338507.
[16] C. Huang et al., Lignin-enzyme interaction: a roadblock for efficient enzymatic hydrolysis of lignocellulosics, Renew. Sustain. Energy Rev. 154 (2022) 111822.
[17] J.G. Ethier, P. Nourian, R. Islam, R. Khare, J.D. Schieber, Microrheology analysis in molecular dynamics simulations: finite box size correction, J. Rheol. 65 (6) (2021) 1255–1267.
[18] D.S. Devarajan et al., Molecular simulation of nanocolloid rheology: viscosity, viscoelasticity, and time-concentration superposition, J. Rheol. 64 (3) (2020) 529–543.
[19] D. Antoine et al., Rapid, Point-of-Care scFv-SERS Assay for Femtogram Level Detection of SARS-CoV-2, ACS sensors 7 (3) (2022) 866–873, https://doi.org/10.1021/acssensors.1c02664.
[20] Maryam Salehi Pourbavarsad et al., Nitrogen oxidation and carbon removal from high strength nitrogen habitation wastewater with nitrification in membrane aerated biological reactors, Journal of Environmental Chemical Engineering 9 (5) (2021), https://doi.org/10.1016/j.jece.2021.106271 106271.
[21] S. Zabihi, S. Hosseini, M. Pishnamazi, M. Lashkarbolooki, F. Borousan, A. Marjani, R. Pelalak, S. Shirazian, Tenoxicam (mobiflex) solubility in carbon dioxide under supercritical conditions, J. Chem. Eng. Data 66 (2) (2021) 990–998.
[22] S.A. Sajadian, M. Amani, N. Saadati Ardestani, S. Shirazian, Experimental analysis and thermodynamic modelling of lenalidomide solubility in supercritical carbon dioxide, Arab. J. Chem. 15 (6) (2022) 103821.
[23] M. Pishnamazi et al., Chloroquine (antimalaria medication with anti SARS-CoV activity) solubility in supercritical carbon dioxide, J. Mol. Liq. 322 (2021) 114539.
[24] M. Pishnamazi, S. Zabihi, S. Jamshidian, M. Lashkarbolooki, F. Borousan, A. Marjani, Evaluation of supercritical technology for the preparation of nanomedicine: etoricoxib analysis, Chem. Eng. Technol. 44 (3) (2021) 559–564.
[25] S. Zabihi, Y. Rahnama, A. Sharafi, F. Borousan, A. Zeinolabedini Hezave, S. Shirazian, Experimental solubility measurements of Fenoprofen in supercritical carbon dioxide, J. Chem. Eng. Data 65 (4) (2020) 1425–1434.
[26] M. Pishnamazi, S. Zabihi, S. Jamshidian, H.Z. Hezaveh, A.Z. Hezave, S. Shirazian, Measuring solubility of a chemotherapy-anti cancer drug (busulfan) in supercritical carbon dioxide, J. Mol. Liq. 317 (2020) 113954.
[27] A. Khoshmaram et al., Supercritical process for preparation of nanomedicine: Oxaprozin case study. Chemical Engineering & Technology, n/a(n/a)..
[28] M. Pishnamazi, S. Zabihi, P. Sarafzadeh, F. Borousan, A. Marjani, R. Pelalak, S. Shirazian, Using static method to measure tolmetin solubility at different pressures and temperatures in supercritical carbon dioxide, Sci. Rep. 10 (1) (2020).
[29] H. Zhu et al., Machine learning based simulation of an anti-cancer drug (busulfan) solubility in supercritical carbon dioxide: ANFIS model and experimental validation, J. Mol. Liq. 338 (2021) 116731.
[30] A. Sadeghi, C.-H. Su, A. Khan, M.d. Lutfor Rahman, M. Sani Sarjadi, S.M. Sarkar, Machine learning simulation of pharmaceutical solubility in supercritical carbon dioxide: prediction and experimental validation for busulfan drug, Arabian J. Chem. 15 (1) (2022) 103502.
[31] M. Yosofvand, S. Liyanage, N.S. Kalupahana, S. Scoggin, N. Moustaid-Moussa, H. Moussa, AdipoGauge software for analysis of biological microscopic images, Adipocyte 9 (1) (2020) 360–373.
[32] M. Latif, D. Zhang, G. Gamez, Field-switching repeller flowing atmospheric-pressure afterglow drift tube ion mobility spectrometry, J. Am. Soc. Mass Spectrom. 1163 (2021) 338507.
[33] S. He, F. Guo, Q. Zou, HuiDing, MRMD2.0: a python tool for machine learning with feature ranking and reduction, Curr. Bioinform. 15 (10) (2021) 1213–1221.
[34] Y.u. Lin, H. Song, F. Ke, W. Yan, Z. Liu, F. Cai, Optimal caching scheme in D2D networks with multiple robot helpers, Comput. Commun. 181 (2022) 132–142.
[35] H. Wu, X. Li, F. Sun, H. Zheng, Y. Zhao, Optimization design method of machine tool static geometric accuracy using tolerance modeling, Int. J. Adv. Manuf. Technol. 118 (5-6) (2022) 1793–1809.
[36] V. Andalib, J. Sarkar, A system with two spare units, two repair facilities, and two types of repairers, Mathematics 10 (6) (2022) 852.
[37] P. Nourian, R. Islam, R. Khare, Implementation of active probe rheology simulation technique for determining the viscoelastic moduli of soft matter, J. Rheol. 65 (4) (2021) 617–632.
[38] D. Zhang, M. Latif, G. Gamez, Instantaneous differentiation of functional isomers via reactive flowing atmospheric pressure afterglow mass spectrometry, Anal. Chem. 93 (29) (2021) 9986–9994.
[39] X. Wang et al., Preparing printable bacterial cellulose based gelatin gel to promote in vivo bone regeneration, Carbohydr. Polym. 270 (2021) 118342.
[40] Y. Hu, B. Yan, Z. Stephen Chen, L. Wang, W. Tang and Caoxing Huang, Recent technologies for the extraction and separation of polyphenols in different plants: a review, J. Renew. Mater. 10 (6) (2022) 1471–1490.
[41] V. Andalib, J. Sarkar, A repairable system supported by two spare units and serviced by two types of repairers, J. Statist. Theory Appl. 20 (2) (2021) 180–192.
[42] W. Lin, J. Yang, Y. Zheng, C. Huang, Q. Yong, Understanding the effects of different residual lignin fractions in acid-pretreated bamboo residues on its enzymatic digestibility, Biotechnol. Biofuels 14 (1) (2021).
[43] C. Huang, S. Tang, W. Zhang, Y. Tao, C. Lai, X.i. Li, Q. Yong, Unveiling the structural properties of lignin-carbohydrate complexes in bamboo residues and its functionality as antioxidants and immunostimulants, ACS Sust. Chem. Eng. 6 (9) (2018) 12522–12531.
[44] S. Shirazian, M. Kuhs, S. Darwish, D. Croker, G.M. Walker, Artificial neural network modelling of continuous wet granulation using a twin-screw extruder, Int. J. Pharm. 521 (1-2) (2017) 102–109.
[45] M. Babanezhad, M. Rezakazemi, N. Hajilary, S. Shirazian, Liquid-phase chemical reactors: development of 3D hybrid model based on CFD-adaptive network-based fuzzy inference system, Can. J. Chem. Eng. 97 (S1) (2019) 1676–1684.
[46] Ali Hosin Alibak et al., Simulation the adsorption capacity of polyvinyl alcohol/carboxymethyl cellulose based hydrogels towards methylene blue in aqueous solutions using cascade correlation neural network (CCNN) technique, Journal of Cleaner Production 337 (2022), https://doi.org/10.1016/j.jclepro.2022.130509 130509.
[47] Srimantha E. Mudiyanselage et al., Automated Workers' Ergonomic Risk Assessment in Manual Material Handling Using sEMG Wearable Sensors and Machine Learning, Electronics 10 (20) (2021), https://doi.org/10.3390/electronics10202558 2558.
[48] Behnam Niknam et al., Effect of variations internal pressure on cracking radiant coils distortion, Structures 34 (2021) 4986–4998, https://doi.org/10.1016/j.istruc.2021.10.083.
[49] I.K. Ahani, M. Salari, A. Shadman, An ensemble multi-step-ahead forecasting system for fine particulate matter in urban areas, J. Cleaner Prod. 263 (2020) 120983.
[50] Q. Wu, C.J.C. Burges, K.M. Svore, J. Gao, Adapting boosting for information retrieval measures, Inf. Retrieval 13 (3) (2010) 254–270.
[51] Y. Cao, Q.-G. Miao, J.-C. Liu, L. Gao, Advance and prospects of AdaBoost algorithm, Acta Autom. Sin. 39 (6) (2013) 745–758.
[52] T.G. Dietterich, Ensemble methods in machine learning, International Workshop on Multiple Classifier Systems, Springer, 2000.
[53] L. Wehenkel, D. Ernst, P. Geurts, Ensembles of extremely randomized trees and some generic applications, Proceedings of Robust Methods for Power System State Estimation and Load Forecasting, 2006.
[54] Z.-H. Zhou, Ensemble Methods: Foundations and Algorithms, 2019, Chapman and Hall/CRC..
[55] S.J. Gershman, D.M. Blei, A tutorial on Bayesian nonparametric models, J. Math. Psychol. 56 (1) (2012) 1–12.
[56] C.K.I. Williams, Prediction with Gaussian processes: From linear regression to linear prediction and beyond, in: M.I. Jordan (Ed.), Learning in Graphical Models, Springer Netherlands, Dordrecht, 1998, pp. 599–621.
[57] E. Schulz, M. Speekenbrink, A. Krause, A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions, J. Math. Psychol. 85 (2018) 1–16.
[58] N.-D. Hoang, A.-D. Pham, Q.-L. Nguyen, Q.-N. Pham, Estimating compressive strength of high performance concrete with Gaussian process regression model, Adv. Civil Eng. 2016 (2016) 1–8.
[59] P. Bartlett, Y. Freund, W.S. Lee, R.E. Schapire, Boosting the margin:a new explanation for the effectiveness of voting methods, Ann. Stat. 26 (5) (1998).
[60] D.-C. Feng et al., Failure mode classification and bearing capacity prediction for reinforced concrete columns based on ensemble machine learning algorithm, Adv. Eng. Inf. 45 (2020) 101126.
[61] L. Breiman, Bagging predictors, Mach. Learn. 24 (2) (1996) 123–140.
[62] H. Erdal, İ. Karahanoğlu, Bagging ensemble models for bank profitability: an emprical research on Turkish development and investment banks, Appl. Soft Comput. 49 (2016) 861–867.
[63] S. Zabihi et al., Measuring salsalate solubility in supercritical carbon dioxide: experimental and thermodynamic modelling, J. Chem. Thermodyn. 152 (2021) 106271.
[64] J. Yuan, K. Wang, T. Yu, M. Fang, Reliable multi-objective optimization of high-speed WEDM process based on Gaussian process regression, Int. J. Mach. Tools Manuf 48 (1) (2008) 47–60.
[65] C.E. Rasmussen, C.K. Williams, Gaussian Processes for Machine Learning, MIT Press [Google Scholar], Cambridge, MA, 2006.
[66] J. Quinonero-Candela, C.E. Rasmussen, A unifying view of sparse approximate Gaussian process regression, J. Mach. Learn. Res. 6 (2005) 1939–1959.
[67] M. Ebden, Gaussian Processes: A Quick Introduction, arXiv preprint arXiv:1505.02965, 2015..
[68] C.K. Williams, C.E. Rasmussen, Gaussian processes for regression, 1996..
[69] A.G. Wilson, D.A. Knowles, Z. Ghahramani, Gaussian Process Regression Networks, arXiv preprint arXiv:1110.4411, 2011..
[70] S. Zhu, E.K. Nyarko, M. Hadzima-Nyarko, Modelling daily water temperature from air temperature for the Missouri River, PeerJ 6 (2018) e4894.
[71] F. Farooq et al., Predictive modeling for sustainable high-performance concrete from industrial wastes: a comparison and optimization of models using ensemble learners, J. Cleaner Prod. 292 (2021) 126032.
[72] Y. Freund, R.E. Schapire, Experiments with a new boosting algorithm. in icml, Citeseer, 1996..
[73] A. Dargahi-Zarandi et al., Modeling minimum miscibility pressure of pure/impure CO2-crude oil systems using adaptive boosting support vector

regression: application to gas injection processes, J. Petrol. Sci. Eng. 184 (2020) 106499.
[74] S.G. Gouda, Z. Hussein, S. Luo, Q. Yuan, Model selection for accurate daily global solar radiation prediction in China, J. Cleaner Prod. 221 (2019) 132–144.
[75] H. Zang et al., Application of functional deep belief network for estimating daily global solar radiation: a case study in China, Energy 191 (2020) 116502.
[76] S. Kim, H. Kim, A new metric of absolute percentage error for intermittent demand forecasts, Int. J. Forecast. 32 (3) (2016) 669–679.