Journal of Molecular Liquids 370 (2023) 120984

Contents lists available at ScienceDirect

Journal of Molecular Liquids

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

Theoretical modeling study on preparation of nanosized drugs using supercritical-based processing: Determination of solubility of Chlorothiazide in supercritical carbon dioxide

Yuanci Li [a,*], Ameer A. Alameri [b], Zainab A. Farhan [c], Hafidh l. AI_Sadi [d], Manal E. Alosaimi [e], Ahmed Ghaleb AbdalSalam [f], Dheyaa Jumaah Jasim [g], Salema K. Hadrawi [h,i], Muataz Mohammed Al-Taee [j], Ali H. Lafta [k], Hakeem A. Othman [l], Saleh Mousa Alzahrani [l], Ashraf A. Moniem [l], Taha Alqadi [m]

[a] School of Pharmacy, Shaanxi Institute of International Trade and Commerce, Xi'an, Shaanxi 710000, China
[b] Department of Chemistry, College of Science, University of Babylon, Babylon, Iraq
[c] Air conditioning and Refrigeration Techniques Engineering department, Al-Mustaqbal University College, Babylon 51001, Iraq
[d] College of MLT, University of Ahl Al Bayt, Kerbala, Iraq
[e] Department of Basic Science, College of Medicine, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia
[f] Department of Pharmacy, AlNoor University College, Iraq
[g] Al-Amarah University College, Al-Amarah, Iraq
[h] Refrigeration and Air-conditioning Technical Engineering Department, College of Technical Engineering, The Islamic University, Najaf, Iraq
[i] Computer Engineering Department, Imam Reza University, Mashhad, Iran
[j] Department of Medical Laboratories Technology, AL-Nisour University College, Baghdad, Iraq
[k] Technical Engineering College, Al-Ayen University, Thi-Qar, Iraq
[l] Department of Mathematics, AL-Qunfudhah University College, Umm Al-Qura University, Saudi Arabia
[m] Department of Biology, Adham University College, Umm Al-Qura University, Saudi Arabia

ARTICLE INFO

Article history:
Received 1 October 2022
Revised 16 November 2022
Accepted 1 December 2022
Available online 7 December 2022

Keywords:
Supercritical carbon dioxide
Chlorothiazide
Solubility
Correlation
Nanomedicine

ABSTRACT

Preparation of drug nanoparticles has been studied and evaluated in this study based on supercritical-based processing as green technology. Computational works have been conducted to evaluate the possibility of manufacturing nanomedicine using this novel technology, and the results are compared with experimental measurements. Chlorothiazide, used as a diuretic and as an antihypertensive was considered as model drug in this work. For the modeling, we used a small data set consisting of two input features, namely temperature and pressure, and one output, namely solubility, in order to analyze the data. Tree ensemble models, including bagging and boosting based on decision trees, have been selected to analyze and model the data. Extremely randomized Trees (Extra Tree), Adaptive Boosting (AdaBoost), and Gradient Boosting models are specifically chosen for this modeling. The hyperparameters of the models were optimized with the help of genetic algorithm (GA) and finally the optimal models were obtained for each of the three methods. Finally, the models were evaluated with different methods. Based on the evaluations, the gradient boosting model showed the best results, and its score was 0.9820 with the coefficient of determination (R^2-score) criterion. Also, the error of the final model with the MEA criterion is 1.51×10^{-2}, with the RMSE criterion equal to 2.51×10^{-2}, and the MAPE error value is 1.59×10^{-2}.

1. Introduction

Measurement of the drug solubility is a key step towards drug manufacturing development. This task is usually done by a variety of methods among which the method of gravimetric is facile and straightforward to be conducted. The method of gravimetric helps finds the amount of solubility for different drugs in particular solvents of interest. Given that crystallization is the key step in production of small-molecular pharmaceuticals, the solubility measurement is required to conduct the crystallization step. Indeed, crystallization is a kind of solid–liquid separation in which the solubility is the key parameter which should be reduced in order to remove the solid from the solution [1–3].

The measurement of drug solubility can be also useful for preparation of nanomedicine which is of great importance for pharmaceutical industry as it can enhance the drug solubility and consequently the drug bioavailability [4]. Using the bottom-up approach, the drug

* Corresponding author.
E-mail address: liyuanci@126.com (Y. Li).

https://doi.org/10.1016/j.molliq.2022.120984
0167-7322/

particles need to be dissolved in a proper solvent, and then nanoparticles are formed in the process [5–7]. Therefore, measuring the drug solubility is the prerequisite step to assess whether the drug is a suitable candidate for the nanonization [8,9]. In recent years, the solubility measurements in supercritical solvents have been proposed for a new and green method for preparation of drug nanoparticles. The process mainly uses CO_2 as the supercritical solvent, and the medicine is dissolved in it for preparation of drug nanoparticles. However, research have been done on measuring and correlating the solubility data via different approaches [10–14].

Given that the experimental measurements of medicine solubility in supercritical solvents is costly and require huge amount of materials and time, computational models can be employed for better performance, and saving time and cost of measurements [15,16]. The computational models can be used based on the mechanistic models, or statistical models to fit the solubility data. The method of thermodynamics and machine learning are among the most commonly used methods for correlation of solubility data for different drugs in supercritical solvents, such as carbon dioxide. For these models, several data points are required for the training, and the model can predict the solubility for the entire range of operational conditions. In this area, usually temperature and pressure are regarded as the most important operational parameters to change the drug solubility values.

Machine learning (ML) has started to gain a great deal of traction in many scientific disciplines as these methods are increasingly replacing classical computing techniques in a broader range of scientific sectors as they take on the roles previously played by them [17]. There is no doubt that ensemble methods are one of the most trending groups of ML methods on the market today. By aggregating predictions from multiple base models, an ensemble can improve the generality and accuracy of a model by increasing the generalization. Among the most helpful methods of ML algorithms of this type, bagging and boosting are the ones that are most commonly used [18].

Multiple base learners can be trained concurrently using bagging (bootstrap aggregation). Ensemble models created by this method are more robust than core models. Bagging involves dividing the original dataset into many subsets (bags) with replacements. Each subset is then generated with a basic model. All models are then run independently, and their projections are combined to make the final predictions [19,20].

Boosting, on the other hand, works with the entire dataset. To begin with, all data points are equally weighted. Every time an iteration is performed, data points with errors are given a higher weight. The predictions are based on a novel model [21,22].

Here, for modeling the process, we utilized three bagging and boosting techniques:

Adaptive Boosting (AdaBoost).

Extremely Randomized Trees (Extra Tree).

Gradient Boosting.

The models have been used to correlate the solubility of a medicine, which is Chlorothiazide in CO_2 as the solvent at supercritical state. The inputs to the models are temperature and pressure of the process, while the only considered output is the Chlorothiazide solubility in the solvent. Therefore, multiples models with two inputs and one output are created to choose the best one in terms of accuracy.

2. Material and methods

2.1. Dataset of drug solubility

We used a dataset in this work for the model development. There are two inputs that are used in this study, which are temperature (T) and pressure (P), which are all displayed in Table 1, which is the data set used in this study. The drug used in this study is Chlorothiazide, and its solubility values have been taken from literature [40]. Aside from that, the only output is a measure of solubility. Indeed, the variations of drug solubility versus temperature and pressure are used in the model to understand the relationship between these parameters. Also, the explanation of the variables of this data set is shown in Fig. 1. Fig. 2 also shows the outlier analysis of this data, where only 2 data points are identified as outliers and these two are not used in the training steps. Once the model has been trained, then we will use it to describe the effect of T and P on the variations of Chlorothiazide solubility.

2.2. Methods of computations

As we mentioned before, in this research, we have used multiple ensemble methods based on decision trees, which we will discuss briefly in the rest of this section. The methods have been chosen based on the number of dataset and also their fitting capability for this solubility dataset. We will try to choose the best model in terms of accuracy by comparing them.

Multiple base (or weak) prediction models can be coupled to build an ensemble learning model, which outperforms a single model in terms of performance. According to Schapire and Freund [23] an ensemble model based on updating instance weights according to prior predictors can improve the performance of base predictors [24,25].

Models can be adaptively enhanced to address a variety of problems, as implied by the title. A simple model's simplicity of structure makes it a reliable generalizer. In spite of their ease of use, they cannot handle complex problems due to their inherent bias. In contrast, complicated models are more likely to be overfitted and are more difficult to implement in practice because of their complexity [26]. A technique called AdaBoost (Adaptive Boosting) is proposed to solve such problems [27,28].

Boosting is another ensemble learning technique that can be used to improve the learning process through ensemble learning. The method of prediction entails a sequence of base predictors instead of a single predictor, which will be averaged together, in order to increase the accuracy of the prediction, rather than relying on a single predictor [29,30]. There is a stage-wise process that is adopted by this method where base estimators (in this case decision trees) are successively fitted in order to remove bias from the model. During the optimization process, a new learner is introduced at each phase in order to optimize the loss function in order

Table 1
The Data points of drug solubility used in this work [40].

T (K)	P (bar)	y ($\times 10^5$)
308	130	0.488
308	170	0.597
308	210	0.661
308	250	0.741
308	290	0.761
318	130	0.466
318	170	0.628
318	210	0.721
318	250	0.79
318	290	0.818
328	130	0.48
328	170	0.62
328	210	0.795
328	250	0.847
328	290	0.919
338	130	0.417
338	170	0.641
338	210	0.851
338	250	0.911
338	290	1.012

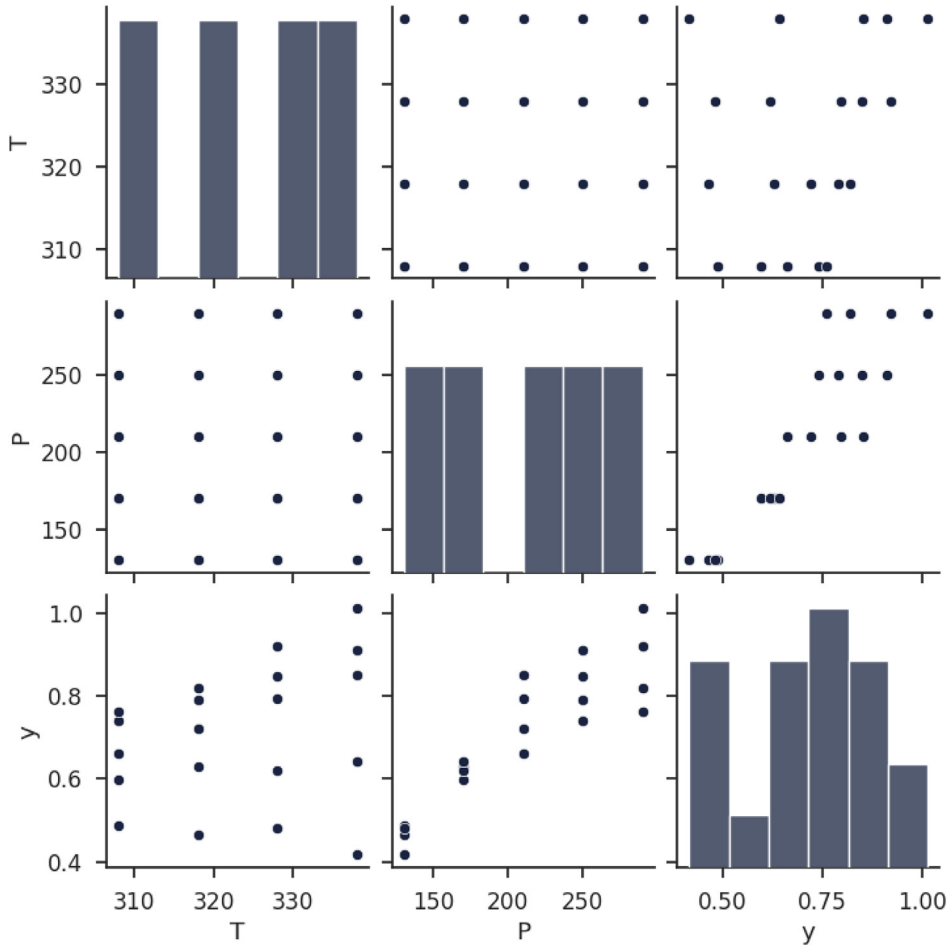

Fig. 1. Distribution of variables.

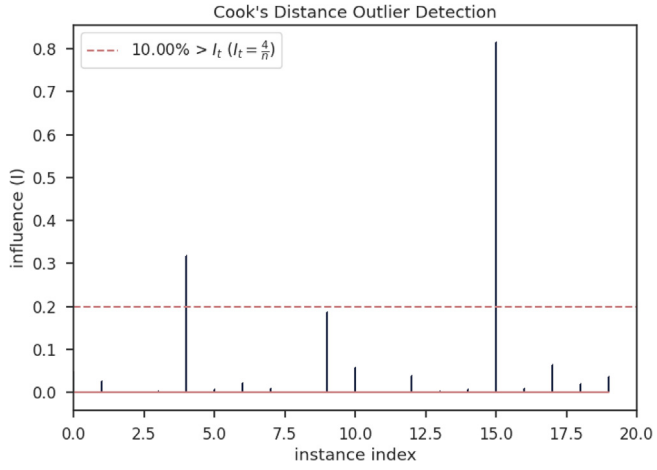

Fig. 2. Cook Distance outlier analysis of the solubility dataset.

to reach the optimal solution. As a result of using the training data, the first learner tries to reduce the loss function to the smallest value possible as a result of using the training data [31–34]. Based on the errors, the following estimators are applied. A gradient boosting algorithm is presented below [32–36]:

Initialize $F_0(x) = \text{argmin}_p \sum_{i-1}^{N} L(y_i, P)$

For $m \in [1, 2, \cdots, M]$:

1. Negative gradient calculation

$\bar{y}_i = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F_{x_i}}\right]$ 2. Create a model

$a_m = \text{argmin}_{a,\beta} \sum_{i=1}^{N} \left[\bar{y} - \beta h(x_i, a_m)\right]^2$ 3. Assign a gradient descent step size to

$p_k = \text{argmin}_p \sum_{i=1}^{N} L(y_i, F_{m-1}(x_i) + ph(x_i, a))$ 4. Modify the estimation of $F(x)$

$F_m(x) = F_{m-1}(x) + p_k h(x, a_m)$Output: Regression function aggregated as $F_m(x)$

Table 2
Optimal Hyper-parameters of the employed models.

Models	Number of Trees	Max depth	Splitter	Learning Rate
AdaBoost	240	16	Best	0.81
Extra Tree	600	5	Random	–
Gradient Boost	210	22	Best	1.29

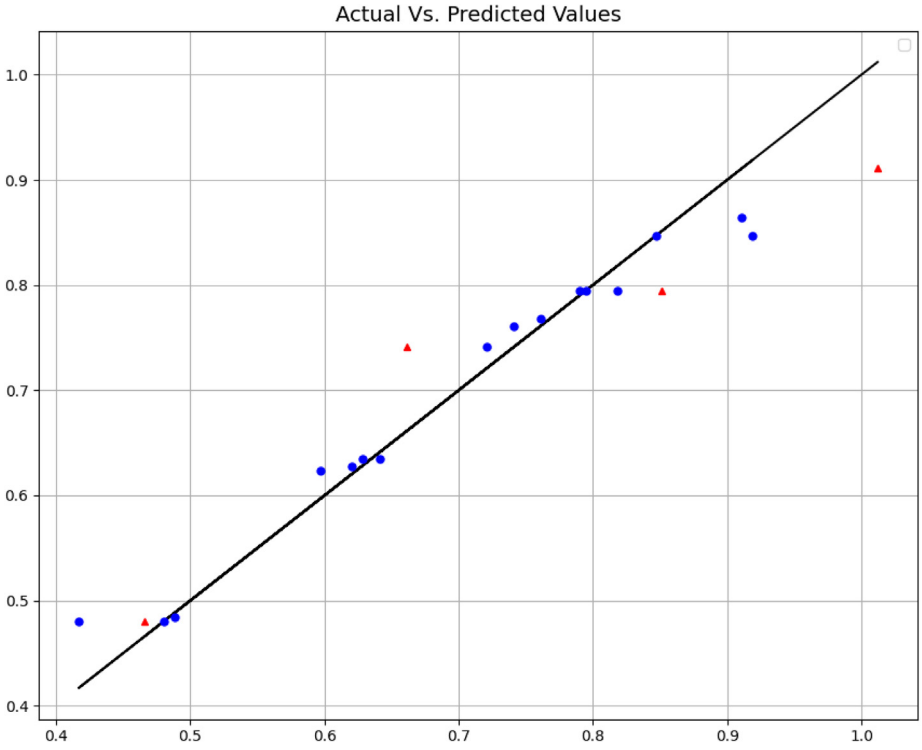

Fig. 3. (AdaBoost) Experimental and modeled values.

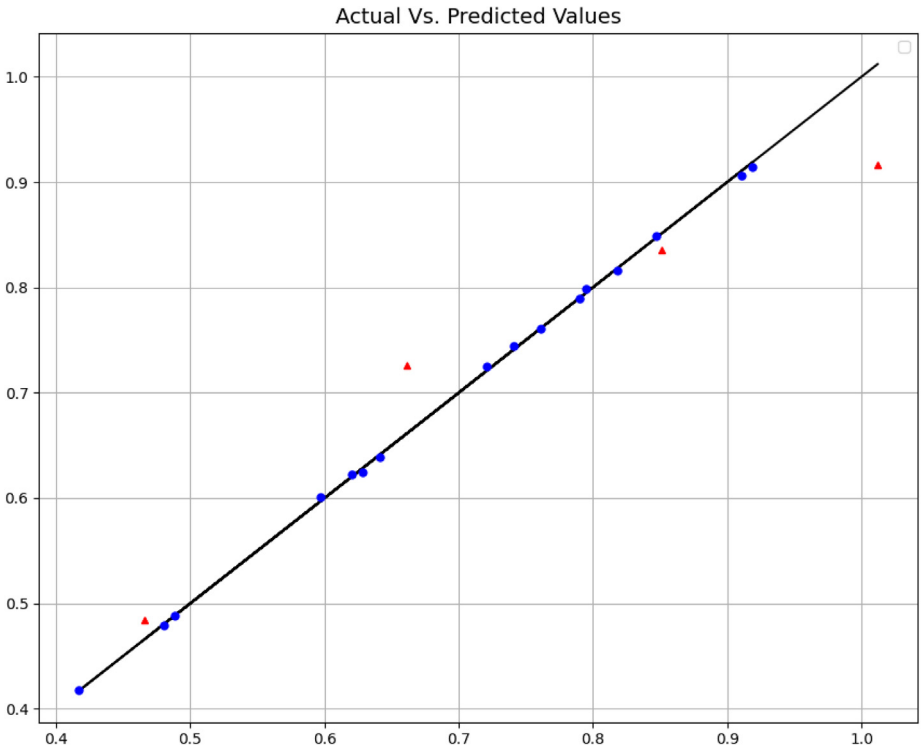

Fig. 4. (Extra Tree) Experimental and modeled values.

A tree-based approach, Extra Trees (ET) are similar to random forests. In order for ET to be able to categorize and analyze data in a way that is relevant to the user, it has to highly randomize both the particularities of each tree node and the cut point decision during its division [37,38].

As far as the way they grow multiple trees and divide nodes using random subsets of functions is concerned, both models are identical. However, the significant differences is that ET relies on randomized splits rather than bootstrap observations, instead of optimum splits [34,39].

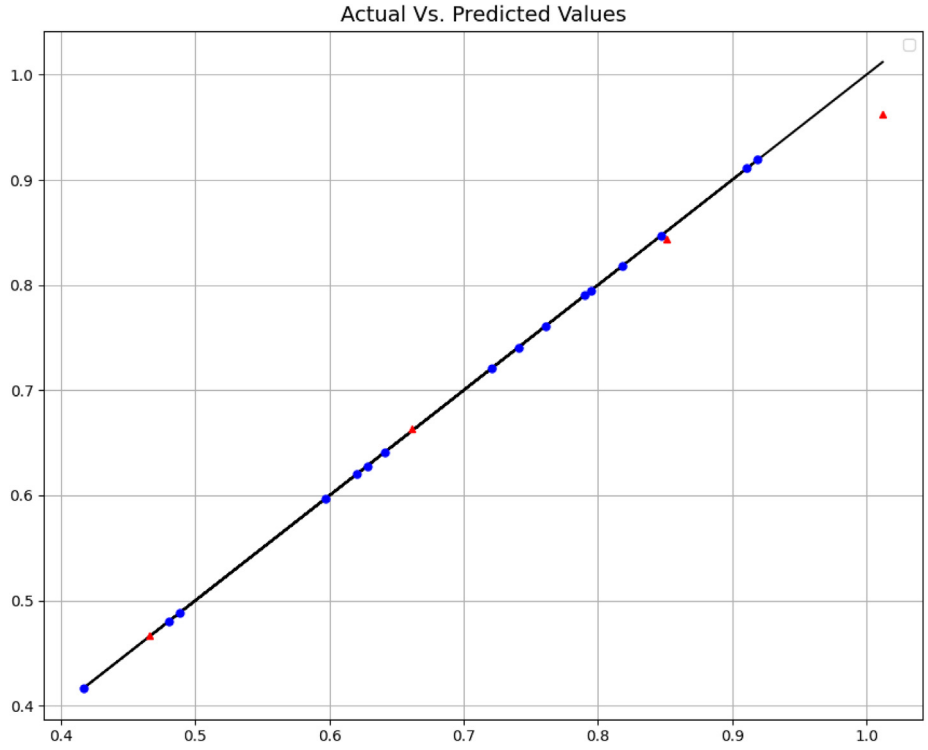

Fig. 5. (Gradient Boosting) Experimental and modeled values.

Table 3
Coefficient of determinations of three tuned models.

Models	Train R^2	Test R^2
AdaBoost	0.9496	0.7998
Extra Tree	0.9996	0.8681
Gradient Boost	0.9998	0.9820

Table 4
Final Model Results.

Models	MAE	RMSE	MAPE
AdaBoost	6.27×10^{-2}	7.05×10^{-2}	7.91×10^{-2}
Extra Tree	4.86×10^{-2}	5.90×10^{-2}	6.25×10^{-2}
Gradient Boost	1.51×10^{-2}	2.51×10^{-2}	1.59×10^{-2}

3. Results and discussions

The introduced models need to be optimized in terms of their hyper-parameters to increase their accuracy and generality. For this purpose, we have used the genetic algorithm in this research. The final results of the optimal parameters are indicated in Table 2.

Figs. 3 to 5 show the distance of the actual values from the values predicted by the models (red for training and blue for testing). These figures show the fact that gradient boosting is more accurate and more general than the other two models. It is also observed that all the tuned models perform well in correlation of the solubility dataset, and great R^2 more than 0.9 has been reported for the

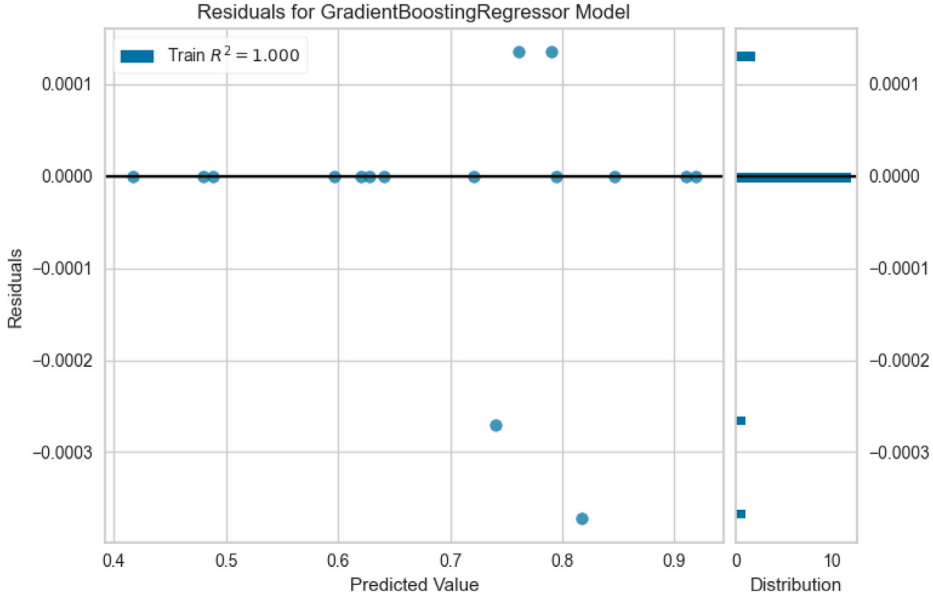

Fig. 6. Gradient boosting (Best model) residuals plot.

training step. Tables 3 and 4 also show the numerical comparison of the models and statistical analysis, which confirm this fact that the methods are properly selected and tuned for describing drug solubility in the supercritical CO_2.

The residual of gradient boosting model is illustrated in Fig. 6, and the 3D representations of the models' outputs are indicated in Figs. 7–9 for the three employed models in this work. It is clearly indicated that all models can show the variations of drug solubility with the temperature and pressure, with almost direct relationship. In fact, the drug solubility should be increased with enhancing the pressure (X2) due to increasing the density and solvation of the solvent which can take more drug to be dissolved at higher pressure. This trend can be clearly observed in the 3D plots which confirm the validity of the developed models in this study.

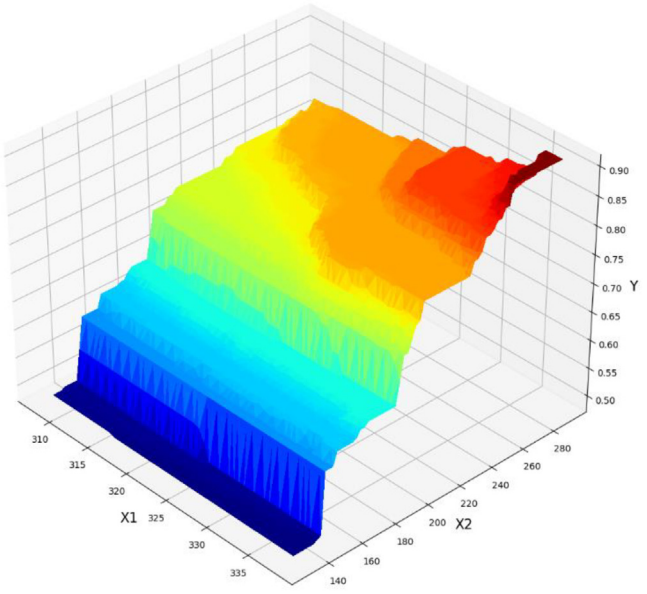

Fig. 7. Final 3d Surface (AdaBoost).

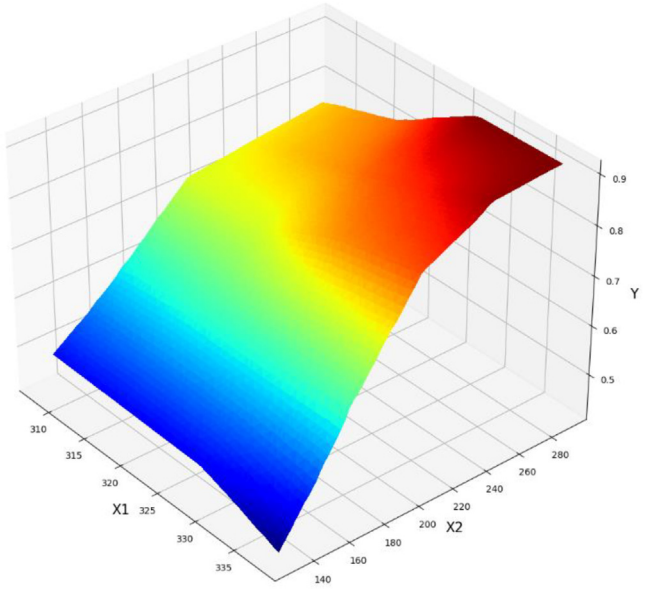

Fig. 8. Final 3d Surface (Extra Tree).

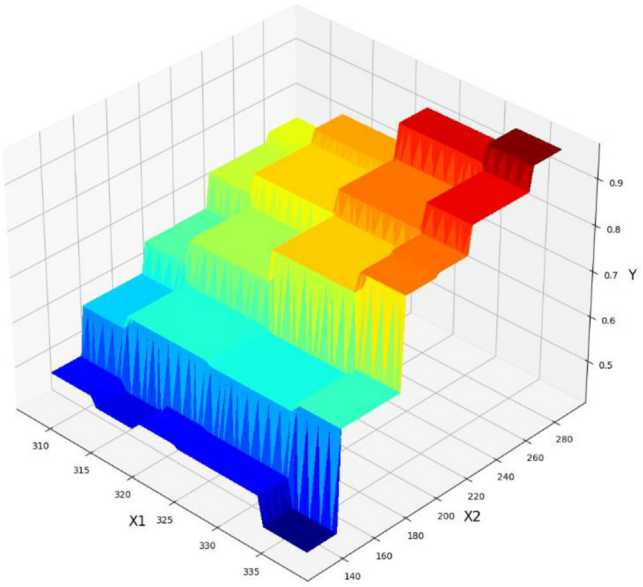

Fig. 9. Final 3d Surface (Gradient Boosting).

4. Conclusions

We optimized a number of machine learning methods for the correlation of a drug solubility to the temperature and pressure. The data set for this study consisted of two inputs, namely temperature and pressure, and one output, namely solubility of drug which is Chlorothiazide. To analyze and model the data, a number of ensemble methods have been chosen, including bagging and boosting based on decision trees. We have specifically selected Extremely Random Trees (Extra Tree), Adaptive Boosting (AdaBoost), and Gradient Boosting models for this modeling task. Using a genetic algorithm (GA), the hyperparameters of the models were optimized and finally, the optimal models were determined for each of the three methods. Lastly, different evaluation methods were applied to the models. According to the evaluations, the gradient boosting model had the best results, with a coefficient of determination (R^2) value of 0.9820. Also, the error of the final model with the MEA criterion is 1.51×10^{-2}, with the RMSE criterion equal to 2.51×10^{-2}, and the MAPE error value is 1.59×10^{-2}.

CRediT authorship contribution statement

Yuanci Li: Conceptualization, Writing – original draft, Investigation. **Ameer A. Alameri:** Writing – original draft, Formal analysis, Resources. **Zainab A. Farhan:** Conceptualization, Project administration, Formal analysis. **Hafidh I. Al_Sadi:** Software, Validation, Writing – original draft. **Manal E. Alosaimi:** Writing – review & editing, Software, Validation, Supervision. **Ahmed Ghaleb AbdalSalam:** Resources, Validation, Software, Conceptualization. **Dheyaa Jumaah Jasim:** Writing – review & editing, Conceptualization, Formal analysis. **Salema K. Hadrawi:** Investigation, Writing – review & editing, Validation. **Muataz Mohammed Al-Taee:** Writing – review & editing, Formal analysis, Software. **Ali H. Lafta:** Conceptualization, Software, Investigation, Validation. **Hakeem A. Othman:** Writing – review & editing, Resources, Validation. **Saleh Mousa Alzahrani:** Writing – review & editing, Formal analysis, Data curation. **Ashraf A. Moniem:** Project administration, Formal analysis, Funding acquisition. **Taha Alqadi:** Supervision, Funding acquisition, Formal analysis.

Data availability

All data are within the published paper.

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.

Acknowledgment

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant code: 22UQU4330052DSR08.

References

[1] M. Garg, A.S. Rathore, Process development in the QbD paradigm: Implementing design of experiments (DoE) in anti-solvent crystallization for production of pharmaceuticals, Journal of Crystal Growth 571 (2021).
[2] A.K. Thakur et al., A critical review on thermodynamic and hydrodynamic modeling and simulation of liquid antisolvent crystallization of pharmaceutical compounds, Journal of Molecular Liquids 362 (2022).
[3] K. Yamaguchi et al., Influence of the crystallization tendencies of pharmaceutical glasses on the applicability of the Adam-Gibbs-Vogel and Vogel-Tammann-Fulcher equations in the prediction of their long-term physical stability, International Journal of Pharmaceutics 626 (2022).
[4] Ramadan, A.G., et al., Biochemical and histopathological alterations induced by subchronic exposure to zinc oxide nanoparticle in male rats and assessment of its genotoxicicty. Journal of Umm Al-Qura University for Applied Sciences, 2022.
[5] W.-F. Lai, W.-T. Wong, Property-Tuneable Microgels Fabricated by Using Flow-Focusing Microfluidic Geometry for Bioactive Agent Delivery, Pharmaceutics 13 (6) (2021) 787.
[6] W.-F. Lai, Development of Hydrogels with Self-Healing Properties for Delivery of Bioactive Agents, Molecular Pharmaceutics 18 (5) (2021) 1833–1841.
[7] S.K. Singh, J.W. Lillard, R. Singh, Reversal of drug resistance by planetary ball milled (PBM) nanoparticle loaded with resveratrol and docetaxel in prostate cancer, Cancer Letters 427 (2018) 49–62.
[8] D. Kim et al., Utilization of a fattigation platform gelatin-oleic acid sodium salt conjugate as a novel solubilizing adjuvant for poorly water-soluble drugs via self-assembly and nanonization, International Journal of Pharmaceutics 575 (2020).
[9] J.-B. Park et al., pH-independent controlled release tablets containing nanonizing valsartan solid dispersions for less variable bioavailability in humans, Journal of Drug Delivery Science and Technology 46 (2018) 365–377.
[10] T. Wang, C.-H. Su, Medium Gaussian SVM, Wide Neural Network and stepwise linear method in estimation of Lornoxicam pharmaceutical solubility in supercritical solvent, Journal of Molecular Liquids 349 (2022).
[11] W. Wang et al., Co-precipitation of 10-hydroxycamptothecin and poly (l-lactic acid) by supercritical CO2 anti-solvent process using dichloromethane/ethanol co-solvent, The Journal of Supercritical Fluids 74 (2013) 137–144.
[12] Z. Zhao et al., Multi support vector models to estimate solubility of Busulfan drug in supercritical carbon dioxide, Journal of Molecular Liquids 350 (2022).
[13] M.H. Zuknik et al., Solubility of virgin coconut oil in supercritical carbon dioxide, Journal of Food Engineering 168 (2016) 240–244.
[14] F. An et al., Machine learning model for prediction of drug solubility in supercritical solvent: Modeling and experimental validation, Journal of Molecular Liquids 363 (2022).
[15] H. Rguigui, Characterization Theorems for the Quantum White Noise Gross Laplacian and Applications, Complex Analysis and Operator Theory 12 (7) (2018) 1637–1656.
[16] H.A. Othman et al., Nanomaterial efficacy on freezing of PCM with involvement of numerical simulation, Journal of Molecular Liquids 362 (2022).
[17] Alpaydin, E., Introduction to machine learning. 2020: MIT press.
[18] S. González et al., A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities, Information Fusion 64 (2020) 205–237.
[19] L. Breiman, Bagging predictors, Machine learning 24 (2) (1996) 123–140.
[20] L. Breiman, Using iterated bagging to debias regressions, Machine Learning 45 (3) (2001) 261–277.
[21] Seyghaly, R., et al. Interference Recognition for Fog Enabled IoT Architecture using a Novel Tree-based Method. in 2022 IEEE International Conference on Omni-layer Intelligent Systems (COINS). 2022. IEEE Computer Society.
[22] R.E. Schapire, Y. Freund, Boosting: Foundations and algorithms, Kybernetes (2013).
[23] Y. Freund, R.E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of computer and system sciences 55 (1) (1997) 119–139.
[24] R.E. Schapire, The boosting approach to machine learning: An overview, Nonlinear estimation and classification (2003) 149–171.
[25] C. Ying et al., Advance and prospects of AdaBoost algorithm, Acta Automatica Sinica 39 (6) (2013) 745–758.
[26] Buitinck, L., et al., API design for machine learning software: experiences from the scikit-learn project. arXiv preprint arXiv:1309.0238, 2013.
[27] G. Lemaître, F. Nogueira, C.K. Aridas, Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning, The Journal of Machine Learning Research 18 (1) (2017) 559–563.
[28] Drucker, H. Improving regressors using boosting techniques. in ICML. 1997. Citeseer.
[29] Duan, T., et al. Ngboost: Natural gradient boosting for probabilistic prediction. in International Conference on Machine Learning. 2020. PMLR.
[30] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Frontiers in neurorobotics 7 (2013) 21.
[31] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Annals of statistics (2001) 1189–1232.
[32] L. Mason et al., Boosting algorithms as gradient descent, Advances in neural information processing systems 12 (1999).
[33] V.-H. Truong et al., A robust method for safety evaluation of steel trusses using Gradient Tree Boosting algorithm, Advances in Engineering Software 147 (2020).
[34] S. Alshehri et al., Design of predictive model to optimize the solubility of Oxaprozin as nonsteroidal anti-inflammatory drug, Scientific Reports 12 (1) (2022) 13106.
[35] Q. Xu et al., PDC-SGB: Prediction of effective drug combinations using a stochastic gradient boosting algorithm, Journal of theoretical biology 417 (2017) 1–7.
[36] W.K. Abdelbasset et al., Development of GBRT Model as a Novel and Robust Mathematical Model to Predict and Optimize the Solubility of Decitabine as an Anti-Cancer Drug, Molecules 27 (17) (2022) 5676.
[37] P. Geurts, D. Ernst, L. Wehenkel, Extremely randomized trees, Machine learning 63 (1) (2006) 3–42.
[38] S. Dutta, U. Mukherjee, S.K. Bandyopadhyay, Pharmacy Impact on Covid-19 Vaccination Progress Using Machine Learning Approach, Journal of Pharmaceutical Research International (2021) 202–217.
[39] Y.-Y. Song, L. Ying, Decision tree methods: applications for classification and prediction, Shanghai archives of psychiatry 27 (2) (2015) 130.
[40] M. Majrashi, A.S. Al-Shati, M. Grishina, S.M. Sarkar, M.T. Nguyen-Le, S. Shirazian, Experimental measurement and thermodynamic modeling of Chlorothiazide solubility in supercritical carbon dioxide, Case Studies in Thermal Engineering 41 (2023) 102621.