



ORIGINAL ARTICLE

Solubility enhancement of decitabine as anticancer drug via green chemistry solvent: Novel computational prediction and optimization



Nader Ibrahim Namazi^a, Sameer Alshehri^{b,*}, Rawan Bafail^c, Bader Huwaimel^k, Amal M. Alsubaiyel^d, Ali H. Alamri^e, Ahmed D. Alatawi^f, Hossam Kotb^g, Mohd Sani Sarjadi^h, Md. Lutfor Rahman^{h,*}, Mohammed A.S. Abourehab^{i,j}

^a *Pharmaceutics and Pharmaceutical Technology Department, College of Pharmacy, Taibah University, Al Madinah Al Munawarah 30001, Saudi Arabia*

^b *Department of Pharmaceutics and Industrial Pharmacy, College of Pharmacy, Taif University, P.O. BOX 11099, Taif 21944, Saudi Arabia*

^c *Department of Pharmaceutics and Pharmaceutical Technology, College of Pharmacy, Taibah University, Medina, Saudi Arabia*

^d *Department of Pharmaceutics, College of Pharmacy, Qassim University, Buraidah 52571, Saudi Arabia*

^e *Department of Pharmaceutics, College of Pharmacy, King Khalid University, Abha 62529, Saudi Arabia*

^f *Department of Clinical Pharmacy, College of Pharmacy, Jouf University, Sakaka, Al-Jouf, Saudi Arabia*

^g *Department of Electrical Power and Machines, Faculty of Engineering, Alexandria University, Alexandria, Egypt*

^h *Faculty of Science and Natural Resources, Universiti Malaysia Sabah, Kota Kinabalu 88400, Sabah, Malaysia*

ⁱ *Department of Pharmaceutics, College of Pharmacy, Umm Al-Qura University, Makkah 21955, Saudi Arabia*

^j *Department of Pharmaceutics and Industrial Pharmacy, Faculty of Pharmacy, Minia University, Minia 61519, Egypt*

^k *Department of Pharmaceutical Chemistry, College of Pharmacy, University of Ha'il, Hail 81442, Saudi Arabia*

Received 14 June 2022; accepted 11 September 2022

Available online 16 September 2022

KEYWORDS

Supercritical fluid;
Validation;
Predictive models;
Simulation;
Solubility

Abstract Nowadays, supercritical fluid technology (SFT) has been an interesting scientific subject in disparate industrial-based activities such as drug delivery, chromatography, and purification. In this technology, solubility plays an incontrovertible role. Therefore, achieving more knowledge about the development of promising numerical/computational methods of solubility prediction to validate the experimental data may be advantageous for increasing the quality of research and therefore, the efficacy of novel drugs. Decitabine with the chemical formula $C_8H_{12}N_4O_4$ is a chemotherapeutic agent applied for the treatment of disparate bone-marrow-related malignancies such as acute myeloid leukemia (AML) by preventing DNA methyltransferase and activation of

* Corresponding authors.

E-mail addresses: s.alshehri@tu.edu.sa (S. Alshehri), lotfor@ums.edu.my (M.L. Rahman).

Peer review under responsibility of King Saud University.



Production and hosting by Elsevier

silent genes. This study aims to predict the optimum value of decitabine solubility in CO₂/SCF by employing different machine learning-based mathematical models. In this investigation, we used AdaBoost (Adaptive Boosting) to boost three base models including Linear Regression (LR), Decision Tree (DT), and GRNN. We used a dataset that has 32 sample points to make solubility models. One of the two input features is P (bar) and the other is T (K). ADA-DT (Adaboost Algorithm-Decision Tree), ADA-LR (Adaboost Algorithm-Linear Regression), and ADA-GRNN (Generative Regression Neural Network) models showed MAE of 6.54×10^{-5} , 4.66×10^{-5} , and 8.35×10^{-5} , respectively. Also, in terms of R-squared score, these models have 0.986, 0.983, and 0.911 scores, respectively. ADA-LR was selected as the primary model according to numerical and visual analysis. Finally, the optimal values are (P = 400 bar, T = 3.38 K × 10², Y = 1.064 × 10⁻³ mol fraction) using this model.

© 2022 The Author(s). Published by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Over the last twenty years, research, and developments (R&D) sections of modern pharmaceutical companies are putting numerous endeavors to discover novel therapeutic methodologies to modify the drug performance for the treatment of fatal and life-threatening diseases (Taberner et al., 2016; Yadav and Dewangan, 2021). Decitabine (Dacogen®) is known a commonly-used chemotherapeutic drug, which has an indication for the treatment of myelodysplastic syndromes acute myeloid leukemia (AML). This cytidine analog can slow the growth of cancerous cells by inhibiting the nucleic acid synthesis (National Center for Biotechnology Information, 2022). Ball-stick molecular structure following with other physicochemical properties of this drug is presented in Table 1.

Supercritical fluid technology (SCFT) has recently been identified as an attractive procedure for momentous industrial-based activities including reaction, purification, nanoparticles manufacturing and drug delivery because of its noteworthy benefits like very low detrimental effect on ecosystem, negligible toxicity and superior quality of the final products. Considering the abovementioned advantages, supercritical fluids (SCFs) are being extensively applied as a trustworthy alternative for many poisonous and eco-harmful solvents (Kankala et al., 2017; Tran and Park, 2021).

A SCF belongs to a novel class of fluid with the temperature and pressure greater than the critical point, where no recognizable difference between liquid and vapor phases exists. To put the issue into perspective view, in SCFs, no distinguishable boundary is present between vapor and liquid phases (McHugh and Krukoni, 2013; Erkey, 2011). Due to demonstrating the properties of both liquid and vapor, SCFs possess great ability to conveniently dissolve different substances. In comparison with disparate types of SCFs, CO₂ has more popularity of application owing to its mild critical points, ease of use, linear structure and nonexplosive characteristics (Behjati Rad et al., 2019; Sabet et al., 2012).

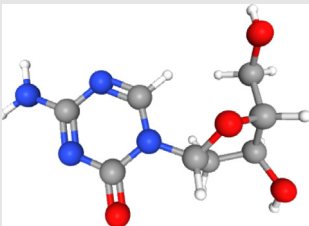
In a SCF system, the solubility of a material (as the capability of extraction at disparate temperature and pressure) plays a significant role in enhancing the acceleration of an extraction process at the initial levels and decreasing the duration of the process. Due to the incontrovertible effect of solubility on SFT processes, various mathematical/computational procedures like molecular simulation, artificial intelligence (AI) and equation of state-based approaches have been employed to estimate the solubility of different substances (i.e., drugs) in SCFs (Hossain et al., 2019; Coimbra et al., 2006; Bitencourt et al., 2016). Nowadays, AI technique has shown its potential of application in different fields of study such as chemical reaction, drug delivery, separation, and heated pipe (Zhu et al., 2021; Öztürk et al., 2018; Staszak, 2020; Wang et al., 2021).

AI approaches that allow machines to learn from data without having to be explicitly programmed are known as machine learning (ML). meta-programs that interpret experimental data and apply that to modify procedures are the objective of machine learning (El Naqa and Murphy, 2015; Wang et al., 2016). It is possible to use data-driven ML technology instead of traditional modeling methodologies to tackle with the complex, non-linear, and unpredictable substances of biodiesel process (Aghbashlo et al., 2021; Gupta et al., 2021).

Boosting is a category of ensemble method can integrate the results of numerous weak predictors to produce a robust predictor. Then, boosting employs a sequential logic to employ weak estimators, which means that the outcome of each weak estimator influences the next estimator. Especially AdaBoost (Freund and Schapire, 1997) is a representative boosting learning algorithm can obtain weak classifiers progressively using reweighted training data. This study used three base models (weak estimators) and boosted them through Adaboost approaches. These models include Linear Regression (LR), Decision Tree (DT), and GRNN.

The reaction of the output parameters to input parameters is modeled using linear regression. This relationship is mostly concerned with figuring out how input variables affect output variables.

Table 1 Characteristics of Loxoprofen (National Center for Biotechnology Information, 2022; Wikipedia Contributors., 2022).

Structure	Formulation	CAS number	Molecular weight	Route of administration
	C ₈ H ₁₂ N ₄ O ₄	2353-33-5	228.208 g.mol ⁻¹	Intravenous

Data can be segmented using basic criteria to generate an empirical tree in decision tree modeling. Repetitive splitting generates a set of rules that can be used to make predictions. Classification and regression trees (CART), C5.0, C4.5, and Chi-squared automatic interaction detection (CHAID) are among the most used tree approaches. It is possible to use C4.5 with a nominal target and inputs that are either nominal or interval (Rokach and Maimon, 2007; Song and Ying, 2015).

Regression function simulation problems may be modelled by using GRNN, which is an RBF-based neural network that simulates dependent variables using a probabilistic framework. It overcomes the problem of local minima that other neural networks have because of its probabilistic construction (Caloiero et al., 2022).

2. Experimental apparatus

Comparison of developed models' outcomes with experimental data obtained by Pishnamazi et al. was done in this paper for validating the results (Pishnamazi et al., 2021). In this paper, determination of decitabine solubility in the CO₂SCF was obtained experimentally via a pressure–volume–temperature (PVT) cell fabricated based on combining static approach with gravimetric technique (Pishnamazi et al., 2021). Decitabine solubility was achieved at wide ranges of pressures from 12 to 40 MPa and temperatures from 308 to 338 K. The significant purpose of current study is focusing on the optimization of solubility in various ranges of pressure and temperature.

3. Data set

A dataset has been used with 32 sample points similar to (Pishnamazi et al., 2021) to create solubility models. There are two input features, Pressure and Temperature, and one result feature, solubility. Table 2 displays the process parameters and output.

4. Methodology

4.1. Base models

In different scopes of AI, trees are among the most essential data structures. A decision tree (DT) is a ML method for analyzing data that is widely utilized. A decision tree can be used to do either regression or classification problems. A basic decision tree is composed of internal nodes (can perform a query on input data), edges (can return the outcome of the query and send it to the one of child nodes), and terminal or leaf nodes (can get decision on output) (Quinlan, 1996; Xu et al., 2005).

Each dataset feature is treated as a node or hub in the DT, with the root node. It has been started through a single node then down the tree to meet the requirements and make better decisions to demonstrate the efficiency of the tree approaches. This strategy will be refined since a leaf node is found. The terminal node could be the DT's results (Kushwah et al., 2021; Breiman et al., 2017; Mathuria, 2013). CART (Breiman et al., 2017), CHAID (Quinlan, 1996), C4.5, and C5.0 (Segal and Bloch, 1989) can be introduced as a comment decision tree induction algorithms (Abdelbasset et al., 2022).

Linear regression is the other primary regression technique employed in this work as a base model:

$$y = \beta_0 + \beta_1 x + \varepsilon$$

The output, represented by y , is the model's independent variable, illustrated by \times (Pombeiro et al., 2017; Kim et al., 2020).

$$\sum_{k=1}^n (y_k - \bar{y})^2 = \sum_{k=1}^n (y_k - \hat{y}_k)^2 + \sum_{k=1}^n (\hat{y}_k - \bar{y}_k)^2$$

Here, y_k stands for the actual value k , \bar{y}_k is the average of y_k , and k denotes the \hat{y}_k .

GRNN is a sort of brain network focused on RBF (radial basis function). To display the reliant factors in a relapse work recreation issue, RBF takes on a probabilistic construction. It conquers the disservice of nearby minima that regular brain networks experience because of its probabilistic design (Caloiero et al., 2022).

The quantity of neurons in the information and result is equivalent to the element of the information and result, separately, in a three-layer brain network like a counterfeit brain organization. In contrast to a counterfeit brain organization, be that as it may, the amount of neurons in the center layer is self-evident and equivalent to the quantity of noticed information used to align the model (Araghinejad, 2013).

A normal (Gaussian) performance function is implemented in the neural network's middle layer, as shown below: (Helali et al., 2022)

$$f(X_r, t) = e^{-[I(t)]^2}$$

$$I(t) = \|X_r - X_t\| \times 0.8326/h, \quad t = 1, 2, \dots, n$$

Here, $\|X_r - X_t\|$ is the Euclidean distance equation of an actual-valued carrier of predictors (X_r) and the actual carrier of estimator linked to the t^{th} neuron (X_t), and h is the spread value that shows the spread of radial basis function and shows the best value. Be that as it may, the spread ought to be chosen by the client inside the scope of ($h > 0$), as more prominent amounts lead to smoother work guess and lower sums lead to close wellness. The normal spread esteem is 1.0.

The output of GRNN model (Y_r) (estimated) for the carrier of (X_r) is produced utilizing a kernel equation of the results [$f(X_r, t)$]: (Khaldi et al., 2015)

$$Y_r = \frac{1}{\sum_{t=1}^n f(X_r, t)} \sum_{t=1}^n [f(X_r, t) \times T_t]$$

4.2. AdaBoost

It is feasible to build an ensemble model by grouping together multiple weak estimators to outperform a single estimator in terms of performance. It was proposed by Freund and Schapire (Freund and Schapire, 1997) as an ensemble procedure to enhance the precision through modifying the sample weight dispensation, which is implemented as the AdaBoost algorithm.

This method became recently more popular due to of the numerous applications it offers. This method, as the name suggests, enhances the abilities of simple models to deal with more complex circumstances. In order to address difficult problems, there are two types of models: simple models and complex models. As their structure is straightforward, simple models offer high generalization properties, which is one of their main draws. Due to a large bias inherent in their structure, they are incapable of solving complex problems in the actual world.

Table 2 Process parameters and output.

No	Pressure (bar)	Temperature (K)	Y (mole fraction)
1	120	3.08×10^2	5.04×10^{-5}
2	120	3.18×10^2	4.51×10^{-5}
3	120	3.28×10^2	3.69×10^{-5}
4	120	3.38×10^2	2.84×10^{-5}
5	160	3.08×10^2	8.23×10^{-5}
6	160	3.18×10^2	9.37×10^{-5}
7	160	3.28×10^2	9.11×10^{-5}
8	160	3.38×10^2	7.79×10^{-5}
9	200	3.08×10^2	1.18×10^{-4}
10	200	3.18×10^2	1.55×10^{-4}
11	200	3.28×10^2	1.77×10^{-4}
12	200	3.38×10^2	2.05×10^{-4}
13	240	3.08×10^2	1.37×10^{-4}
14	240	3.18×10^2	1.87×10^{-4}
15	240	3.28×10^2	2.82×10^{-4}
16	240	3.38×10^2	3.71×10^{-4}
17	280	3.08×10^2	1.76×10^{-4}
18	280	3.18×10^2	2.40×10^{-4}
19	280	3.28×10^2	3.42×10^{-4}
20	280	3.38×10^2	4.90×10^{-4}
21	320	3.08×10^2	1.97×10^{-4}
22	320	3.18×10^2	2.69×10^{-4}
23	320	3.28×10^2	4.27×10^{-4}
24	320	3.38×10^2	7.15×10^{-4}
25	360	3.08×10^2	2.18×10^{-4}
26	360	3.18×10^2	3.40×10^{-4}
27	360	3.28×10^2	5.60×10^{-4}
28	360	3.38×10^2	8.74×10^{-4}
29	400	3.08×10^2	2.83×10^{-4}
30	400	3.18×10^2	5.06×10^{-4}
31	400	3.28×10^2	7.88×10^{-4}
32	400	3.38×10^2	1.07×10^{-3}

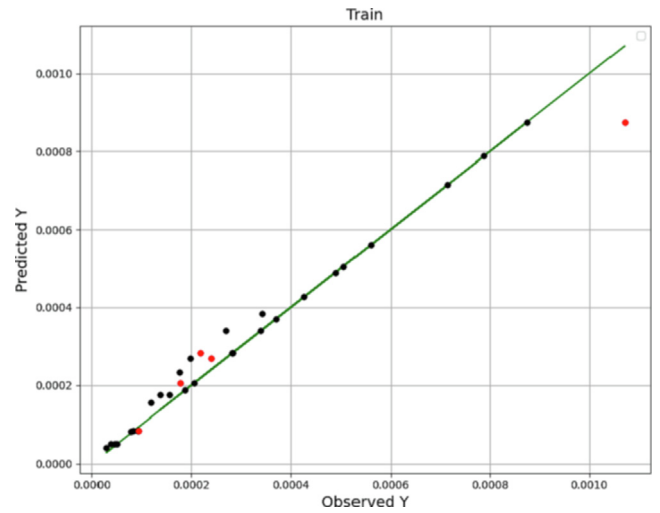
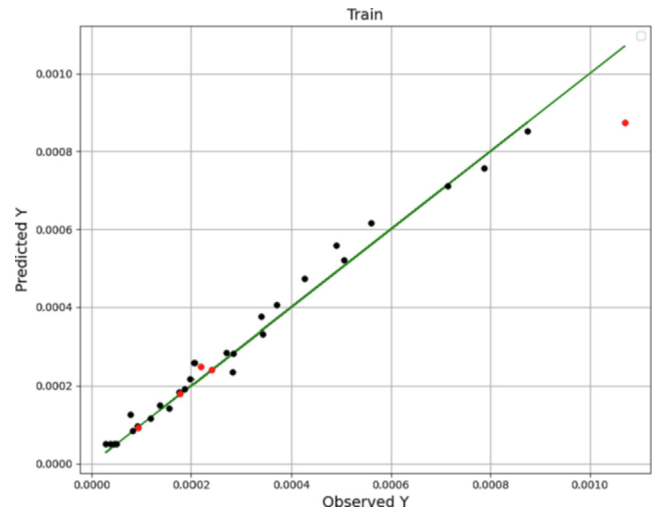
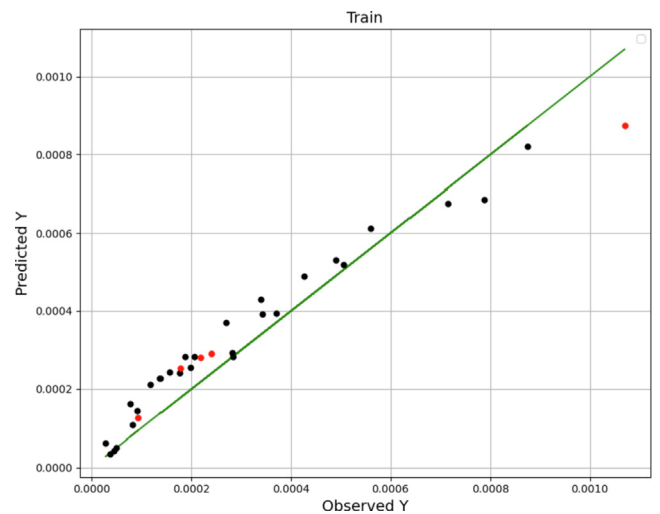
Over-fitting or having a complex structure, on the other hand, is more common in complicated models, and their application in reality is more difficult because of this (Buitinck et al., 1309). With these issues in mind, AdaBoost presents a solution: Using a weak predictor as a base model, another methods are merged progressively to produce a robust process be able to behave for complicated scenarios, as shown in this method (Lemaître et al., 2017).

5. Results

We used MAE, RMSE, and R-squared to validate the performance of regression models after modifying hyper-parameters of models by running different combinations of them. The horizontal spacing among any two continuous amounts, particularly the formations of observed and expected outputs, may be determined using the following equation, the Mean Absolute Error (MAE) (Botchkarev, 2018).

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

In this equation, n refers to the quantity of dataset and y_i stands for the actual observed value, and \hat{y} is also the estimated values.

**Fig. 1** ADA-DT model: observed-predicted chart.**Fig. 2** ADA-LR model: observed-predicted chart.**Fig. 3** ADA-GRNN model: observed-predicted chart.

Models	MAE	RMSE	R-Squared
ADA-DT	6.54×10^{-5}	9.41×10^{-5}	0.986
ADA-LR	4.66×10^{-5}	8.88×10^{-5}	0.983
ADA-GRNN	8.35×10^{-5}	1.01×10^{-4}	0.911

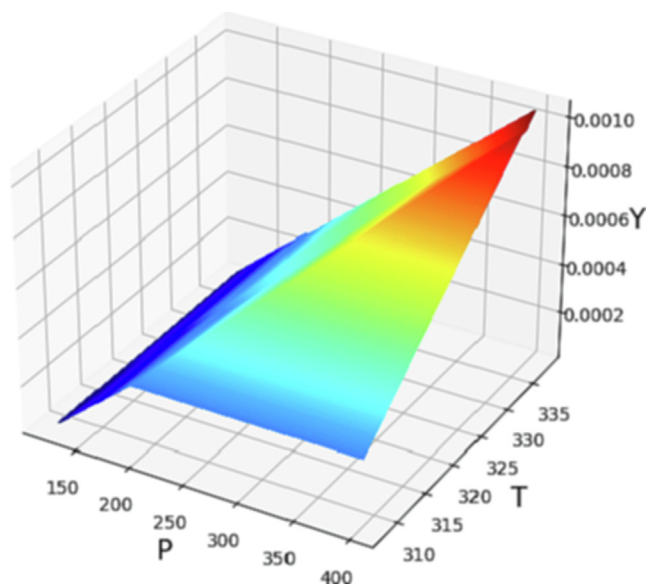


Fig. 4 3D illustration of inputs/outputs (ADA-LR Model).

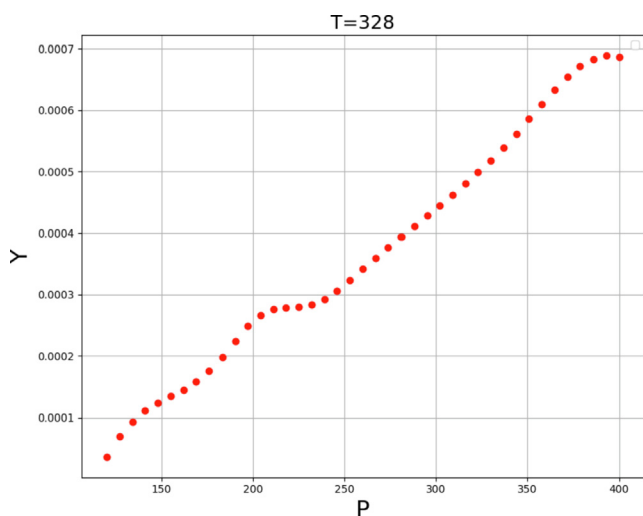


Fig. 5 Tendency of variable P.

The Root Mean Squared Error (RMSE) was the second examination pointer utilized (RMSE). A dataset's standard deviation is determined utilizing the normal qualities and noticed values given by this situation.

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

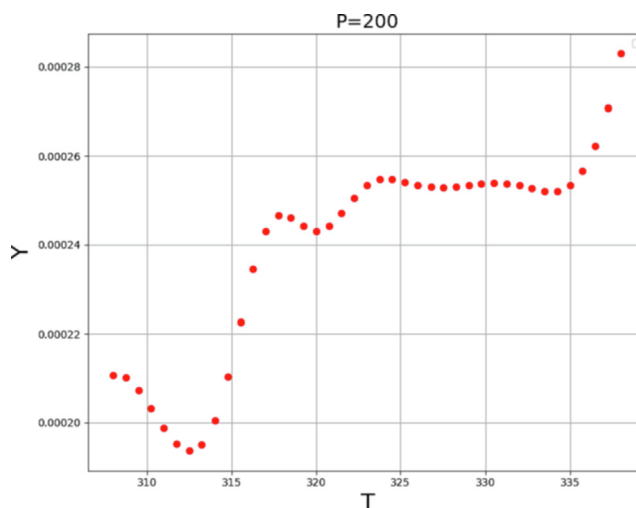


Fig. 6 Tendency of variable T.

Table 4 Optimized process parameters.

P (bar)	T (K)	Solubility (Y)
400	3.38×10^2	1.064×10^{-3}

the R²-Score (Botchkarev, 2018):

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \mu)^2}$$

μ refers the average of the real observed data (Botchkarev, 2018). Figs. 1, 2 and 3 schematically compare the experimental results (actual results) with the results achieved from ADA-DT, ADA-LR and ADA-GRNN predictive models. In these figures, black dots are the indicator of train and red dots are the indicator of test, respectively. Additionally, green line illustrates the actual values. Comparison of the obtained MAE, RMSE and R-squared values of three developed models based on Table 3 implies that the ADA-LR model is the most accurate model with better generality.

The solubilizing strength of SCFs profoundly depends on the pressure and temperature values. Near the critical point, variation of temperature and pressure may have great influence on the solubilizing power of SCFs. Fig. 4 shows the 3D scheme based on the ADA + LR computational model to simultaneously investigate the influence of pressure and temperature as input variables on the solubility amount of decitabine in SCCO₂. Moreover, the influence of pressure and temperature as individual input parameters on the solubility of decitabine is presented by Figs. 5 and 6, respectively. As mentioned before, through enhancing the pressure, the solubility of decitabine in SCCO₂ solvent due to molecular compression and consequently increasing the density amount of solvent. Fig. 6 shows that enhancing in the pressure value from 110 to 400 bar significantly improved the solubility of decitabine from about 0.00003 to 0.00068. About temperature, there is a trade-off between two competing parameters with different impacts. It means that enhance in temperature causes the enhancement of sublimation pressure (SP) and deterioration of solvent density (SD), simultaneously. Therefore, the

analysis of these two parameters for determining the encouraging or unfavorable effect of temperature on decitabine solubility is of great importance. At pressures superior to the cross-over pressure (CP), by increasing the temperature, the amount of SP increases while there is a great reduction in the value of SD. At these pressures, the influence of SP increment is more than the negative impact of SD reduction and thus, increase in the temperature causes improving the solubility of decitabine in SCCO₂ solvent. By dropping the pressure lower than the CP, the negative impact of SD reduction dominates the positive effect of SP improvement and therefore, by increasing the temperature the solubility of decitabine reduces substantially. The optimized value of decitabine solubility and its corresponding pressure and temperature are presented in Table 4.

6. Conclusion

Great advantages of CO₂SCF like mild critical points, simplicity of application, linear structure and nonexplosive characteristics has made this solvent more attractive than toxic organic solvents in industrial activities. Despite the indisputable role of experimental investigations on the achievement of various drugs' solubility, the existence of long operational duration and high cost has motivated the researchers to employ efficient and precise mathematical models based on artificial intelligence (AI) to predict this parameter. In this study, we used Ada-Boost (Adaptive Boosting) to improve three baseline models: Linear Regression (LR), Decision Tree (DT), and GRNN. To create solubility models, we used a dataset with 32 sample points. P (bar) is one of the two input features, while T is the other (k). MAE values for the ADA-DT, ADA-LR, and ADA-GRNN models were 6.54×10^{-5} , 4.66×10^{-5} , and 8.35×10^{-5} , respectively. These models also had R-squared scores of 0.986, 0.983, and 0.911. Based on numerical and visual examination, the ADA-LR model was chosen as the major model. Finally, using this model, the ideal parameters are ($P = 400$, $T = 3.38 \times 10^2$, $Y = 1.064 \times 10^{-3}$).

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

The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for supporting this work by Grant Code: (22UQU4290565DSR95).

References

- Abdelbasset, W.K., Elkholi, S.M., Ismail, K.A., Alshehri, S., Alobaida, A., Huwaimel, B., Alatawi, A.D., Alsubaiyel, A. M., Venkatesan, K., Abourehab, M.A.S., 2022. Development a novel robust method to enhance the solubility of Oxaprozin as nonsteroidal anti-inflammatory drug based on machine-learning. *Sci Rep* 12, 13138. <https://doi.org/10.1038/s41598-022-17440-4>.
- Aghbashlo, M., Peng, W., Tabatabaei, M., Kalogirou, S.A., Soltanian, S., Hosseinzadeh-Bandbafha, H., Mahian, O., Lam, S.S., 2021. Machine learning technology in biodiesel research: A review. *Prog. Energy Combust. Sci.* 85, 100904.
- Araghinejad, S., 2013. *Data-driven modeling: using MATLAB® in water resources and environmental engineering*. Springer Science & Business Media.
- Behjati Rad, H., Karimi Sabet, J., Varaminian, F., 2019. Effect of Stearic Acid as a Co-solvent on the Solubility Enhancement of Aspirin in Supercritical CO₂. *Chem. Eng. Technol.* 42, 1259–1267.
- Bitencourt, R.G., Cabral, F.A., Meirelles, A.J., 2016. Ferulic acid solubility in supercritical carbon dioxide, ethanol and water mixtures. *J. Chem. Thermodyn.* 103, 285–291.
- Botchkarev, A., 2018. Evaluating performance of regression machine learning models using multiple error metrics in azure machine learning studio. Available at SSRN 3177507.
- Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 2017. *Classification and regression trees*. Routledge.
- Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., Niculae, V., Prettenhofer, P., Gramfort, A., Grobler, J., 2013. API design for machine learning software: experiences from the scikit-learn project, arXiv preprint arXiv:1309.0238.
- Caloiero, T., Aristodemo, F., Ferraro, D.A., 2022. Annual and seasonal trend detection of significant wave height, energy period and wave power in the Mediterranean Sea. *Ocean Eng.* 243, 110322.
- Coimbra, P., Duarte, C., De Sousa, H., 2006. Cubic equation-of-state correlation of the solubility of some anti-inflammatory drugs in supercritical carbon dioxide. *Fluid Phase Equilib.* 239, 188–199.
- El Naqa, I., Murphy, M.J., 2015. What is machine learning?. In: *Machine Learning in Radiation Oncology*. Springer, pp. 3–11.
- Erkey, C., 2011. Fundamental aspects of supercritical fluids. In: *Supercritical Fluid Science and Technology*. Elsevier, pp. 11–19.
- Freund, Y., Schapire, R.E., 1997. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* 55, 119–139.
- Gupta, K., Kalita, K., Ghadai, R., Ramachandran, M., Gao, X., 2021. Machine Learning-Based Predictive Modelling of Biodiesel Production—A Comparative Perspective. *Energies* 2021, 14, 1122, in, s Note: MDPI stays neutral with regard to jurisdictional claims in published . . . , 2021.
- Helali, J., Ghaleni, M.M., Hosseini, S.A., Siraei, A.L., Saeidi, V., Safarpour, F., Mirzaei, M., Lotfi, M., 2022. Assessment of machine learning model performance for seasonal precipitation simulation based on teleconnection indices in Iran. *Arab J Geosci* 15, 1343. <https://doi.org/10.1007/s12517-022-10640-2>.
- Hossain, S., Kabehev, A., Parrow, A., Bergström, C.A., Larsson, P., 2019. Molecular simulation as a computational pharmaceuticals tool to predict drug solubility, solubilization processes and partitioning. *Eur. J. Pharm. Biopharm.* 137, 46–55.
- Kankala, R.K., Zhang, Y.S., Wang, S.B., Lee, C.H., Chen, A.Z., 2017. Supercritical fluid technology: an emphasis on drug delivery and related biomedical applications. *Adv. Healthcare Mater.* 6, 1700433.
- Kim, M.K., Kim, Y.-S., Srebric, J., 2020. Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression. *Sustainable Cities Soc.* 62, 102385.
- Khalidi, K., Boudraa, A.O., Torresani, B., Chonavel, T., 2015. HHT-based audio coding. *SIVIP* 9, 107–115. <https://doi.org/10.1007/s11760-013-0433-6>.
- Kushwah, J.S., Kumar, A., Patel, S., Soni, R., Gawande, A., Gupta, S., 2021. Comparative study of regressor and classifier with decision tree using modern tools. *Mater. Today: Proc.*
- Lemaître, G., Nogueira, F., Aridas, C.K., 2017. Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Machine Learn. Res.* 18, 559–563.
- Mathuria, M., 2013. Decision tree analysis on j48 algorithm for data mining. *Int. J. Adv. Res. Comput. Sci. Software Eng.* 3.
- McHugh, M., Krukonis, V., 2013. *Supercritical fluid extraction: principles and practice*. Elsevier.
- National Center for Biotechnology Information. PubChem Compound Summary for CID 451668, Decitabine. <https://pubchem.ncbi.nlm.nih.gov/compound/Decitabine>. Accessed May 22, 2022.

- Öztürk, A.A., Gündüz, A.B., Ozisik, O., 2018. Supervised machine learning algorithms for evaluation of solid lipid nanoparticles and particle size. *Comb. Chem. High Throughput Screening* 21, 693–699.
- Pishnamazi, M., Zabihi, S., Jamshidian, S., Borousan, F., Hezave, A. Z., Marjani, A., Shirazian, S., 2021. Experimental and thermodynamic modeling decitabine anti cancer drug solubility in supercritical carbon dioxide. *Sci. Rep.* 11, 1–8.
- Pombeiro, H., Santos, R., Carreira, P., Silva, C., Sousa, J.M., 2017. Comparative assessment of low-complexity models to predict electricity consumption in an institutional building: Linear regression vs. fuzzy modeling vs. neural networks. *Energy Build.* 146, 141–151.
- Quinlan, J.R., 1996. Learning decision tree classifiers. *ACM Computing Surveys (CSUR)* 28, 71–72.
- Rokach, L., Maimon, O.Z., 2007. *Data mining with decision trees: theory and applications*. World Scientific.
- Sabet, J.K., Ghotbi, C., Dorkoosh, F., Striolo, A., 2012. Solubilities of acetaminophen in supercritical carbon dioxide with and without menthol cosolvent: Measurement and correlation. *Scientia Iranica* 19, 619–625.
- Segal, M.R., Bloch, D.A., 1989. A comparison of estimated proportional hazards models and regression trees. *Stat. Med.* 8, 539–550.
- Song, Y.-Y., Ying, L., 2015. Decision tree methods: applications for classification and prediction. *Shanghai Arch. Psychiatry* 27, 130.
- Staszak, M., 2020. Artificial intelligence in the modeling of chemical reactions kinetics. *Phys. Sci. Rev.*
- Taberner, A., Gonzalez-Garcinuno, A., Galan, M.A., del Valle, E.M. M., 2016. Survey of supercritical fluid techniques for producing drug delivery systems for a potential use in cancer therapy. *Rev. Chem. Eng.* 32, 507–532.
- Tran, P., Park, J.-S., 2021. Application of supercritical fluid technology for solid dispersion to enhance solubility and bioavailability of poorly water-soluble drugs. *Int. J. Pharm.* 610, 121247.
- Wang, H., Lei, Z., Zhang, X., Zhou, B., Peng, J., 2016. *Machine learning basics*. Deep Learning, 98–164.
- Wang, X., Luo, L., Xiang, J., Zheng, S., Shittu, S., Wang, Z., Zhao, X., 2021. A comprehensive review on the application of nanofluid in heat pipe based on the machine learning: Theory, application and prediction. *Renew. Sustain. Energy Rev.* 150, 111434.
- Wikipedia contributors. (2022, February 18). Decitabine. In *Wikipedia, The Free Encyclopedia*. Retrieved 03:22, May 22, 2022, from <https://en.wikipedia.org/w/index.php?title=Decitabine&oldid=1072580011>.
- Xu, M., Watanachaturaporn, P., Varshney, P.K., Arora, M.K., 2005. Decision tree regression for soft classification of remote sensing data. *Remote Sens. Environ.* 97, 322–336.
- Yadav, D., Dewangan, H.K., 2021. PEGYLATION: an important approach for novel drug delivery system. *J. Biomater. Sci. Polym. Ed.* 32, 266–280.
- Zhu, H., Zhu, L., Sun, Z., Khan, A., 2021. Machine learning based simulation of an anti-cancer drug (busulfan) solubility in supercritical carbon dioxide: ANFIS model and experimental validation. *J. Mol. Liq.* 338, 116731.