Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEKI)
Vol. 9, No. 1, March 2023, pp. 49-57
ISSN: 2338-3070, DOI: 10.26555/jiteki.v9i1.25608

Development of Novel Machine Learning to Optimize the Solubility of Azathioprine as Anticancer Drug in Supercritical Carbon Dioxide

Arya Adhyaksa Waskita, Stevry Yushady CH Bissa, Ika Atman Satya, Ratna Surya Alwi
Research Centre for Computing, National Research and Innovation Agency (BRIN), Jl. Raya Jakarta-Bogor KM 46 Cibinong, Indonesia

ARTICLE INFO

Article history:

Received January 9, 2023
Revised January 29, 2023
Published February 02, 2023

Keywords:

Solubility;
Machine learning;
Supercritical carbon dioxide;
Optimization;
Azathioprine

ABSTRACT

Supercritical carbon dioxide (Sc-CO_2) has thus been proposed as an appropriate solvent for diluting the pharmaceuticals to increase particle size. The use of supercritical fluids (SCFs) in various industrial applications, such as extraction, chromatography, and particle engineering, has attracted considerable interest. Recognizing the solubility behavior of various drugs is an essential step in the pharmaceutical industry's pursuit of the most effective supercritical approach. In this work, four models were used to predict the solubility of Azathioprine in supercritical carbon dioxide, including Ridge regression (RR), Huber regression (HR), Random forest (RF), and Gaussian process regression (GPR). The R-squared scores of all four models are 0.974, 0.6518, 0.966, and 1.0 for Ridge regression (RR), Huber regression (HR), Random forest (RF), and Gaussian process regression (GPR) models, respectively. The RMSE error rates of 2.843 $\times 10^{-13}$, 7.036 $\times 10^{-12}$, 5.673 $\times 10^{-13}$, and 1.054 $\times 10^{-30}$ for the RR, HR, RF, and GPR models, respectively. MAE metrics of 1.205 $\times 10^{-6}$, 2.151 $\times 10^{-6}$, 5.997 $\times 10^{-7}$ and 9.419 $\times 10^{-16}$ errors were also found in the RR, HR, RF, and GPR models, respectively. It was found that Ridge regression (RR), Random forest (RF), and Gaussian process regression (GPR) models can be used to predict any compound's solubility in supercritical carbon dioxide.

Corresponding Author:

Ratna Surya Alwi, Research Centre for Computing, National Research and Innovation Agency (BRIN), Jl. Raya Jakarta-Bogor KM 46 Cibinong, Indonesia
Email: ratn017@brin.go.id

1. INTRODUCTION

In the field of pharmaceutical manufacturing, invention of new medications and innovation of favorable therapeutic approachesare the most significant obstacles [1]. An essential step in the pharmaceutical industry's quest for the most effective supercritical approach is the recognition of the solubility behavior of various therapeutic drugs [2]. In fact, the development of methods of particle engineering that are suitableto regulate particle size is of the utmost importance due to the significance of aspects to take into account include solubility and bioavailability [3][4][5][6][7][8]. Sc-CO_2 is well-liked because of features such as low price, low critical pressure, high diffusivity, low viscosity, and high inertness or toxicity. Sc-CO_2 has thus been proposed as a suitable solvent for increasing the particle size of pharmaceuticals. The use of supercritical fluids (SCFs) in numerous applications within the industrial sector, such as chromatography, extractions, and particle engineering, has attracted considerable interest [9], [10][11]–[16]. There have been a great deal of scientific investigations, both experimental and theoretical, carried out to comprehend the properties of Sc-CO_2 systems, particularly the interactions between individual molecules in supercritical fluid solutions [17][3], [10], [18]. Additionally, advancements have been made in the use of Sc-CO_2 as an alternative solvent system for the processing of materials [19], [20].

Azathioprine is a crystalline solid mercaptopurine derivative with the chemical name 6-[(1-methyl-4-nitro-1H-imidazol-5-yl)thio]-9H-purine. It's a drug used mainly to stop the body from rejecting a transplant by

lowering the immune system's T-lymphocyte borne delayed immune responses [21][22][23]. Rheumatoid arthritis, organ transplantation, Crohn's disease, chronic active hepatitis, systemic lupus erythematosus, polyarteritis nodosa, and other autoimmune disorders are all treated with azathioprine [22][23]. Additionally, Azathioprine is authorized as a medicine for the treatment of particular forms of cancer and inflammatory bowel diseases [24], [25]. Only a very small amount of azathioprine can be dissolved in water and other aqueous solutions, but it is highly soluble in organic solvents such as Dimethyl sulfoxide [21].

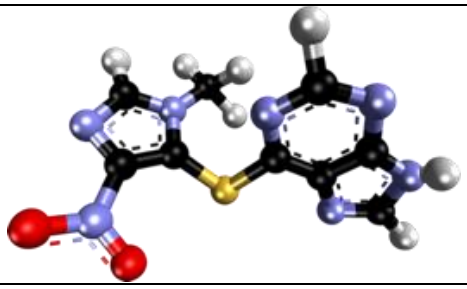
Recently, machine learning (ML) method has become a powerful tool in the scientific disciplines [26]–[34]. In the present work, four models were used to predict the solubility of Azathioprine in supercritical carbon dioxide, including Ridge regression (RR), Huber regression (HR), Random forest (RF), and Gaussian process regression (GPR). Moreover, R^2, RMSE, and MAE were utilized to evaluate the models used. As mentioned previously, the novelty of this study is the application of machine learning to four distinct new models in order to optimize their configurations (hyperparameters) to improve and predict how well a drug will dissolve in water. Thus, the pharmaceutical industry benefits from research, new drugs are developed, and promising therapeutic approaches are advanced.

2. METHODS

2.1. Data Sheet

In this study, 32 data points of solubility on Azathioprine in Sc-CO_2 were used [35]. The chemical structure, formula, molecular weight, and melting temperature of Azathioprine are presented in Table 1. Fig. 1 shows the research diagrams of this study. Y is the solubility output, which has two inputs (temperature = X_1, and pressure = X_2), and it is displayed in Table 2.

Table 1. Azathioprine's molecular form, melting point, and other physical characteristics

Solute	Chemical Form	Structure	Molecular Weight (g/mol)	T_m (K)
Azathioprine	$C_9H_7N_7O_2S$		277.263	526 [36]

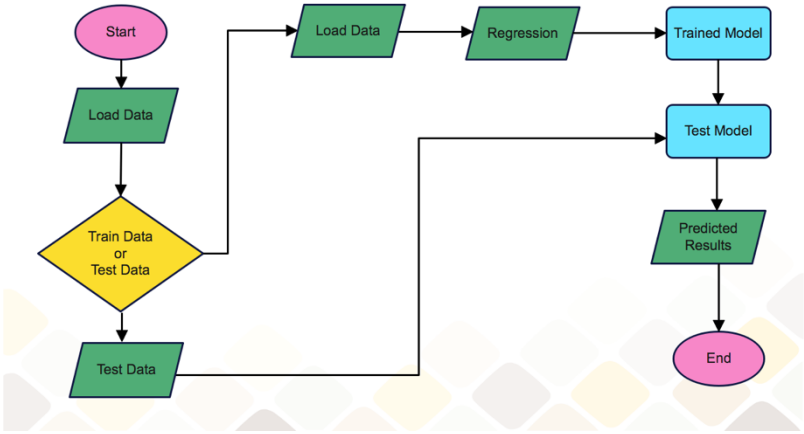


Fig. 1. Research diagrams of Azathioprine predictive solubility in Sc-CO_2

2.2. Ridge Regression (RR)

Ridge regression is a well-known parameter estimation technique that can be applied in multiple linear regression in order to address the frequent collinearity problem. The following is the standard model for performing multiple linear regressions.

$$y = \boldsymbol{x}\beta + \varepsilon, \tag{1}$$

where $E(\varepsilon) = 0$, $E(\varepsilon\varepsilon') = \sigma^2 I_n$, and x is $(n \times p)$ and by bold symbols. The matrix I_n is the identity matrix with dimension n by n [37].

Table 2. Data sheet used in this work

No.	Temperature (K) (X_1)	Pressure (MPa) (X_2)	Y (10^6)
1	308	12	5.1
2	308	15	6.4
3	308	18	6.7
4	308	21	7.4
5	308	24	8.1
6	308	27	8.6
7	318	12	4.0
8	318	15	7.1
9	318	18	7.7
10	318	21	9.4
11	318	24	11.2
12	318	27	12.5
13	328	12	3.4
14	328	15	8.1
15	328	18	9.6
16	328	21	11.9
17	328	24	13.4
18	328	27	15.6
19	338	12	2.7
20	338	15	9.0
21	338	18	12.0
22	338	21	15.0
23	338	24	17.1
24	338	27	18.3
25	308	12	5.1
26	308	15	6.4
27	308	18	6.7
28	308	21	7.4
29	308	24	8.1
30	308	27	8.6
31	318	12	4.0
32	318	15	7.1

2.3. Huber Regression (HR)

Huber regression is an outlier-tolerant regression technique. It is to use a different loss function as opposed to the standard least-squares formula [38]. Definition of the Huber loss as

$$l_\tau(x) = \begin{cases} x^2/2, & \text{if} |x| \leq \tau, \\ \tau|x| - x^2/2, & \text{if} |x| > \tau, \end{cases} \tag{2}$$

where $\tau > 0$ is the robustification parameter that achieves a satisfactory compromise between bias and robustness. $l_\tau(x)$ is the quadratic form of the loss function of x, and when x exceeds some threshold, the graph linearizes τ in magnitude. The τ presides over the blending of quadraticand l_τ losses, which can be considered to be the two polar opposites of the Huber loss with $\tau = \infty$ and $\tau \to 0$, respectively.

2.4. Random Forest (RF)

The steps of bootstrapping and bagging need to be completed before a regression problem can be solved using the random forest (RF) method. A random subset of the training dataset is used in the first step of the process, which generates a set of decision trees based on the growth of each individual tree. After achieving the ensemble, the second stage disassembles the nodes of the decision tree by selecting random subsets of training samples during the initial bagging procedure.The decision is made by selecting the optimal subdivision and its value. The random forest (RF) model can be seen as a collection of decision trees, $G(x, \theta_r)$ is the G^{th} predicting tree, and θ provides a distribution vector that is independent and uniform and that was assigned

before the tree grew [39]. The Breiman (3) is used to construct the forest by combining and average. The RF model is comparable to a collection of decision trees, with G at the cantering the whole trees [39].

$$G(x, \theta_1, \dots \dots \theta_r) = \frac{1}{R}\sum_{r=1}^{R} G(x, \theta_r) \tag{3}$$

2.5. Gaussian Process Regression (GPR)

GPR is a type of nonlinear regression that doesn't use parametric models but uses a probabilistic regression framework [40].

The result variable y in this method can be presented as follows:

$$y = f\big(x(k)\big) + \varepsilon \tag{4}$$

Here (x) is a calculation of data results, f represents the lack of clarity regarding the functional dependence, and ε refers to Gaussian noise (σ_n^2) is the variation that is present in Gaussian noise. Mean and standard deviation are both Gaussian $p\ (y_*|X, y, x_*)$ $c$an be calculated by the following formulas [41].

$$\hat{y}_* = m(x_*) + k_*^T (K + \sigma_n^2 I)^{-1}\big(y - m(x_*)\big), \tag{5}$$

$$\sigma_{y_*}^2 = k_* + \sigma_n^2 - k_*^T (K + \sigma_n^2 I)^{-1} k_*, \tag{6}$$

Here *K* is matrix covariance by using the elements $k_{i,j} = cov\ \big(xi, x_j\big)$, vector *k* as follows

$$[k_*]i = cov\ (xi, x_*)\ and\ k_* = \ cov\ (x_*, x_*) \tag{7}$$

To make reliable predictions, the dataset is used to figure out the mean and covariance function attributes. Because of how the predictive possible distribution functions, the attributes are shown as hyper attributes. The hyper-attributes are made by maximizing $\log p(y|X)$.

$$\log p(y|X) = -\frac{1}{2} y^T (K + \sigma_n^2 I)^{-1} y - \frac{1}{2}\log\left(|K + \sigma_n^2 I|\right) - \frac{n}{2}\log\left(2\pi\right) \tag{8}$$

Here *n* is the quantity of training subset.

2.6. Model evaluation metrics

To evaluate the models selected performance, we used three metrics: the root mean square error, also known as RMSE, the mean absolute error, also known as MAE, and the coefficient of determination, R2, are calculated as follows:

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (y_i - \mu)^2} \tag{9}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{10}$$

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n} (\hat{y}_i - y_i)^2\right]^{1/2} \tag{11}$$

where y_i is the measured solubility, $\hat{y}_i$ is the predicted solubility, and n is the quantity of data.

3. RESULTS AND DISCUSSION

In this study, 32 data points on Azathioprine solubility in Sc-CO_2 were used. Table 1 presents the chemical structure, formula, molecular weight, and melting temperature of Azathioprine. *Y* is the solubility output, which has two input (temperature = X_1, and pressure = X_2) and it is displayed in Table 2. Scikit-learn [42] is a widely used Python package for conventional machine learning algorithms on which we train all of the models.

Fig. 2 shows the comparison of predicted solubility and real measured of Azathioprine using the Ridge Regression (RR). Fig. 3 presents the comparison of predicted solubility and real measured of Azathioprine

using the Huber Regression (HR). Moreover, the comparison of predicted solubility and real measured of Azathioprine using the Random Forest (RF) are presented in Fig. 4. Additionally, Fig. 5 shows the comparison of predicted solubility and real measured of Azathioprine using the Gaussian Process (GPR). Table 3 summarized performance of Ridge regression (RR), Huber regression (HR), Random forest (RF), and Gaussian process regression (GPR) models for the prediction of Azathioprine solubility in Sc-CO_2, respectively. It was found that the Gaussian Process (GPR) model prediction accuracy was better than three other developed regression machines as presented in Table 3 and Fig. 2- Fig. 5. Fig. 6 displays a 3D results of the input to the single output.

Table 3. Performance of various models (RR, HR, RF, and GPR) for Solubility prediction of Azathioprine in Sc-CO2

Models	R^2	RMSE	MAE
Ridge Regression (RR)	0.974	2.843×10^{-13}	1.205×10^{-6}
Huber Regression (HR)	0.6518	7.036×10^{-12}	2.151×10^{-6}
Random Forest (RF)	0.966	5.673×10^{-13}	5.997×10^{-7}
Gaussian Process (GPR)	1.0	1.054×10^{-30}	9.419×10^{-16}

The Ridge Regression, Huber Regression, Random Forest, and Gaussian Process models each have a RMSE error of 2.843×10^{-13}, 7.036×10^{-12}, 5.673×10^{-13}, and 1.054×10^{-30}, respectively. The MAE values for the RR, HR, RF, and GPR were also found to have 1.205×10^{-6}, 2.151×10^{-6}, $5.997 \times 10^{-7}$and 9.419×10^{-16}, respectively. Additionally, the R^2 values of the RR, HR, RF, and GPR were found 0.974, 0.6518, 0.966, and 1.0, respectively.

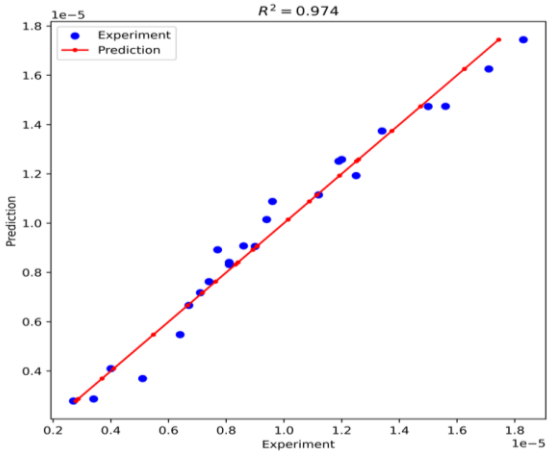


Fig. 2. Ridge Regression plot of predicted solubility of Azathioprine versus experiment

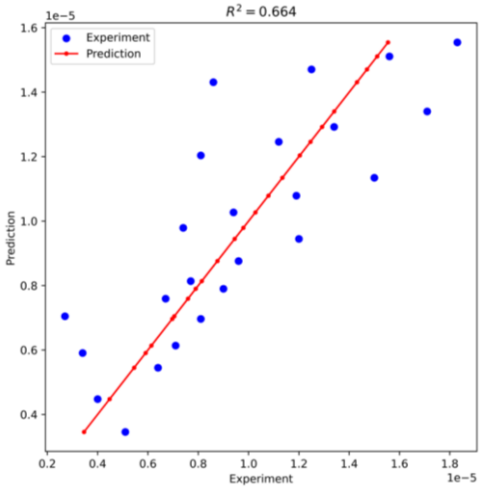


Fig. 3. Huber Regression plot of predicted solubility of Azathioprine versus experiment

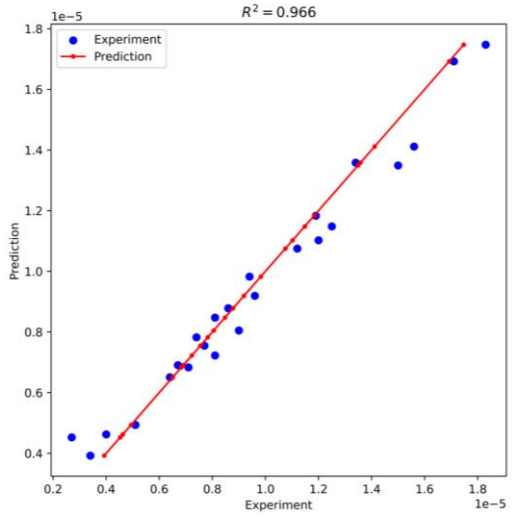


Fig. 4. Random Forest Regression plot of predicted solubility of Azathioprine versus experiment

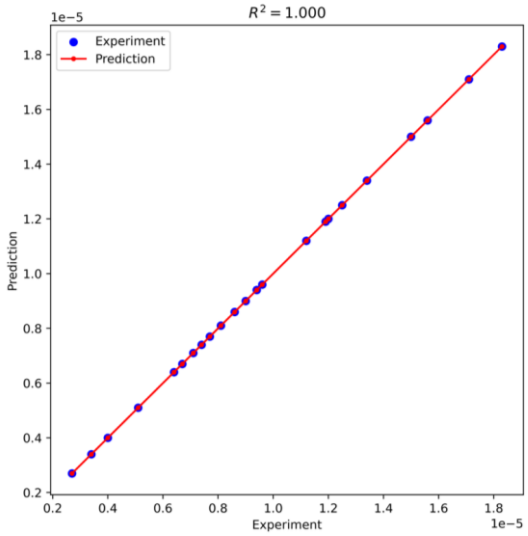


Fig. 5. Gaussian Regression plot of predicted solubility of Azathioprine versus experimental data

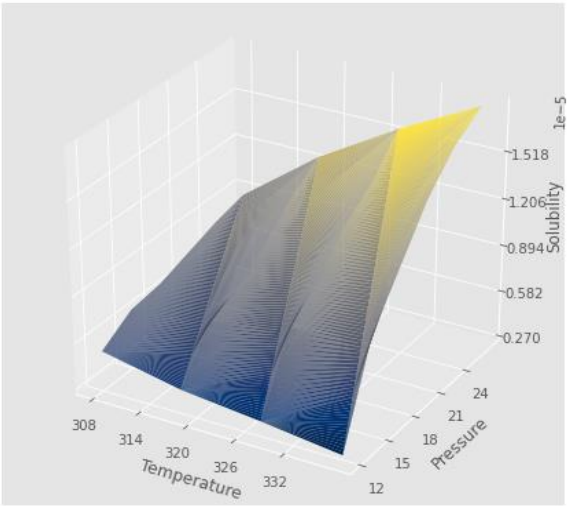


Fig. 6. Gaussian Regression 3D for the solubility of Azathioprine

4. CONCLUSION

In the pharmaceutical industry, optimizing the solubility of various drugs in Sc-CO_2 over a broad temperature and pressure range is a desirable endeavor. The primary objective of this study is to use four machine learning regression algorithms to predict the optimal solubility of anticancer and an immunosuppressive drug in Sc-CO_2. In this regard, four machine learning regression algorithms methods were used in this study to look at the data of solubility of Azathioprine in Sc-CO_2: Ridge regression (RR), Huber regression (HR), Random forest (RF), and Gaussian process regression (GPR). The RR, HR, RF, and GPR models each have a RMSE error rate of 2.843×10^{-13}, 7.036×10^{-12}, 5.673×10^{-13}, and 1.054×10^{-30}. The MAE metrics for the RR, HR, RF, and GPR were also found to have 1.205×10^{-6}, 2.151×10^{-6}, 5.997×10^{-7} and 9.419×10^{-16}. Additionally, the R^2 values of the RR, HR, RF, and GPR were found 0.974, 0.6518, 0.966, and 1.0. It was found that Ridge regression (RR), Random forest (RF), and Gaussian process regression (GPR) models can be used to predict the solubility of any compounds in supercritical carbon dioxide. Finally, this work can be used to optimize and predict the drug solubility. So, research helps the pharmaceutical industry, leads to the creation of new drugs, and moves forward promising therapeutic approaches.

Acknowledgments

The authors are thankful to the Head of Research Center for Computing, National Research and Innovation Agency (BRIN), Indonesia for providing the required computer.

REFERENCES

[1] R. K. Z. Kankala Yu Shrike; Wang, Shi-Bin; Lee, Chia-Hung; Chen, Ai-Zheng, "Supercritical Fluid Technology: An Emphasis on Drug Delivery and Related Biomedical Applications.," *Adv. Healthc. Mater.*, vol. 6, no. 16, pp. 1700433-NA, 2017, https://doi.org/10.1002/adhm.201700433.

[2] R. S. Alwi, T. Tanaka, and K. Tamura, "Measurement and correlation of solubility of anthraquinone dyestuffs in supercritical carbon dioxide," *J. Chem. Thermodyn.*, vol. 74, pp. 119–125, 2014, https://doi.org/10.1016/j.jct.2014.01.015.

[3] A. Tabernero, E. M. Martín del Valle, and M. A. Galán, "Supercritical fluids for pharmaceutical particle engineering: Methods, basic fundamentals and modelling," *Chem. Eng. Process. Process Intensif.*, vol. 60, pp. 9–25, 2012, https://doi.org/10.1016/j.cep.2012.06.004.

[4] M. Elveny, A. Khan, A. T. Nakhjiri, and A. B. Albadarin, "A state-of-the-art review on the application of various pharmaceutical nanoparticles as a promising technology in cancer treatment," *Arab. J. Chem.*, vol. 14, no. 10, p. 103352, 2021, https://doi.org/10.1016/j.arabjc.2021.103352.

[5] G. Sodeifian, S. A. Sajadian, and N. S. Ardestani, "Determination of solubility of Aprepitant (an antiemetic drug for chemotherapy) in supercritical carbon dioxide: Empirical and thermodynamic models," *J. Supercrit. Fluids*, vol. 128, pp. 102–111, 2017, https://doi.org/10.1016/j.supflu.2017.05.019.

[6] G. Sodeifian and S. A. Sajadian, "Solubility measurement and preparation of nanoparticles of an anticancer drug (Letrozole) using rapid expansion of supercritical solutions with solid cosolvent (RESS-SC)," *J. Supercrit. Fluids*, vol. 133, pp. 239–252, 2018, https://doi.org/10.1016/j.supflu.2017.10.015.

[7] G. Sodeifian, F. Razmimanesh, and S. A. Sajadian, "Prediction of solubility of sunitinib malate (an anti-cancer drug) in supercritical carbon dioxide (SC–CO2): Experimental correlations and thermodynamic modeling," *J. Mol. Liq.*, vol. 297, p. 111740, 2020, https://doi.org/10.1016/j.molliq.2019.111740.

[8] G. Sodeifian, F. Razmimanesh, S. A. Sajadian, and S. M. Hazaveie, "Experimental data and thermodynamic modeling of solubility of Sorafenib tosylate, as an anti-cancer drug, in supercritical carbon dioxide: Evaluation of Wong-Sandler mixing rule," *J. Chem. Thermodyn.*, vol. 142, p. 105998, 2020, https://doi.org/10.1016/j.jct.2019.105998.

[9] T. Rezaei *et al.*, "A universal methodology for reliable predicting the non-steroidal anti-inflammatory drug solubility in supercritical carbon dioxide," *Sci. Rep.*, vol. 12, no. 1, p. 1043, 2022, https://doi.org/10.1038/s41598-022-04942-4.

[10] G. Sodeifian, L. Nasri, F. Razmimanesh, and M. Abadian, "CO2 utilization for determining solubility of teriflunomide (immunomodulatory agent) in supercritical carbon dioxide: Experimental investigation and thermodynamic modeling," *J. CO2 Util.*, vol. 58, p. 101931, 2022, https://doi.org/10.1016/j.jcou.2022.101931.

[11] G. Sodeifian, S. A. Sajadian, and S. Daneshyan, "Preparation of Aprepitant nanoparticles (efficient drug for coping with the effects of cancer treatment) by rapid expansion of supercritical solution with solid cosolvent (RESS-SC)," *J. Supercrit. Fluids*, vol. 140, pp. 72–84, 2018, https://doi.org/10.1016/j.supflu.2018.06.009.

[12] G. Sodeifian, S. A. Sajadian, N. Saadati Ardestani, and F. Razmimanesh, "Production of Loratadine drug nanoparticles using ultrasonic-assisted Rapid expansion of supercritical solution into aqueous solution (US-RESSAS)," *J. Supercrit. Fluids*, vol. 147, pp. 241–253, 2019, https://doi.org/10.1016/j.supflu.2018.11.007.

[13] G. Sodeifian and S. A. Sajadian, "Utilization of ultrasonic-assisted RESOLV (US-RESOLV) with polymeric stabilizers for production of amiodarone hydrochloride nanoparticles: Optimization of the process parameters," *Chem. Eng. Res. Des.*, vol. 142, pp. 268–284, 2019, https://doi.org/10.1016/j.cherd.2018.12.020.

[14] G. Sodeifian, N. Saadati Ardestani, S. A. Sajadian, and H. Soltani Panah, "Experimental measurements and thermodynamic modeling of Coumarin-7 solid solubility in supercritical carbon dioxide: Production of nanoparticles

via RESS method," *Fluid Phase Equilib.*, vol. 483, pp. 122–143, 2019, https://doi.org/10.1016/j.fluid.2018.11.006.
[15] N. Saadati Ardestani, G. Sodeifian, and S. A. Sajadian, "Preparation of phthalocyanine green nano pigment using supercritical CO2 gas antisolvent (GAS): experimental and modeling," *Heliyon*, vol. 6, no. 9, p. e04947, 2020, https://doi.org/10.1016/j.heliyon.2020.e04947.
[16] F. Razmimanesh, G. Sodeifian, and S. A. Sajadian, "An investigation into Sunitinib malate nanoparticle production by US- RESOLV method: Effect of type of polymer on dissolution rate and particle size distribution," *J. Supercrit. Fluids*, vol. 170, p. 105163, 2021, https://doi.org/10.1016/j.supflu.2021.105163.
[17] S. K. P. Misra Kamla, "Supercritical fluid technology for solubilization of poorly water soluble drugs via micro- and naonosized particle generation," *ADMET DMPK*, vol. 8, no. 4, pp. 355-NA, 2020, https://doi.org/10.1016/j.molliq.2022.119689.
[18] G. Sodeifian, R. Surya Alwi, F. Razmimanesh, and F. Sodeifian, "Solubility of prazosin hydrochloride (alpha blocker antihypertensive drug) in supercritical CO2: Experimental and thermodynamic modelling," *J. Mol. Liq.*, p. 119689, 2022, https://doi.org/10.1016/j.molliq.2022.119689.
[19] C. Y. Ji Yang; Zhang, Min; Yu, Wenyan; Dong, Shuo, "Loxoprofen Sodium Alleviates Oxidative Stress and Apoptosis Induced by Angiotensin II in Human Umbilical Vein Endothelial Cells (HUVECs).," *Drug Des. Devel. Ther.*, vol. 14, no. NA, pp. 5087–5096, 2020, https://doi.org/10.2147/DDDT.S266175.
[20] N. S. Yamakawa Shintaro; Kimoto, Ayumi; Arai, Yasuhiro; Ishihara, Tomoaki; Yokomizo, Kazumi; Okamoto, Yoshinari; Otsuka, Masami; Tanaka, Ken Ichiro; Mizushima, Tohru, "Low direct cytotoxicity of loxoprofen on gastric mucosal cells," *Biol. Pharm. Bull.*, vol. 33, no. 3, pp. 398–403, 2010, https://doi.org/10.1248/bpb.33.398.
[21] D. W. Newton, S. Ratanamaneichatara, and W. J. Murray, "Dissociation, solubility and lipophilicity of azathioprine," *Int. J. Pharm.*, vol. 11, no. 3, pp. 209–213, 1982, https://doi.org/10.1016/0378-5173(82)90039-4.
[22] O. H. Nielsen, B. Vainer, and J. Rask-Madsen, "the treatment of inflammatory bowel disease with 6-mercaptopurine or azathioprine," *Aliment. Pharmacol. Ther.*, vol. 15, no. 11, pp. 1699–1708, 2001.
[23] G. B. Elion and G. H. Hitchings, "Azathioprine," in *Antineoplastic and immunosuppressive agents*, Springer, pp. 404–425, 1975, https://doi.org/10.1007/978-3-642-65806-8_19.
[24] C. T. Dollery, *Therapeutic drugs*, vol. 2. Churchill Livingstone, 1991.
[25] G. B. Appel, "New and future therapies for lupus nephritis.," *Cleve. Clin. J. Med.*, vol. 79, no. 2, pp. 134–140, 2012, https://doi.org/10.3949/ccjm.78gr.11004.
[26] Y. Cao, A. Khan, S. Zabihi, and A. B. Albadarin, "Neural simulation and experimental investigation of Chloroquine solubility in supercritical solvent," *J. Mol. Liq.*, vol. 333, p. 115942, 2021, https://doi.org/10.1016/j.molliq.2021.115942.
[27] I. Euldji, C. Si-Moussa, M. Hamadache, and O. Benkortbi, "QSPR Modelling of the Solubility of Drug and Drug-like Compounds in Supercritical Carbon Dioxide," *Mol. Inform.*, vol. 41, no. 10, p. 2200026, 2022, https://doi.org/10.1002/minf.202200026.
[28] H. Zhu, L. Zhu, Z. Sun, and A. Khan, "Machine learning based simulation of an anti-cancer drug (busulfan) solubility in supercritical carbon dioxide: ANFIS model and experimental validation," *J. Mol. Liq.*, vol. 338, p. 116731, 2021, https://doi.org/10.1016/j.molliq.2021.116731.
[29] A. Baghban, J. Sasanipour, and Z. Zhang, "A new chemical structure-based model to estimate solid compound solubility in supercritical CO2," *J. CO2 Util.*, vol. 26, pp. 262–270, 2018, https://doi.org/10.1016/j.jcou.2018.05.009.
[30] J. M. Schmidt Mário R. G.; Botti, Silvana; Marques, Miguel A. L., "Recent advances and applications of machine learning in solid-state materials science," *npj Comput. Mater.*, vol. 5, no. 1, pp. 1–36, 2019, https://doi.org/10.1038/s41524-019-0221-0.
[31] Y. H. Liu Weixiang; Cao, Bing-Yang, "Machine learning for predicting thermodynamic properties of pure fluids and their mixtures," *Energy*, vol. 188, no. NA, pp. 116091-NA, 2019, https://doi.org/10.1016/j.energy.2019.116091.
[32] N.-D. P. Hoang Anh-Duc; Nguyen, Quoc-Lam; Pham, Quang-Nhat, "Estimating Compressive Strength of High Performance Concrete with Gaussian Process Regression Model," *Adv. Civ. Eng.*, vol. 2016, no. 2016, pp. 1–8, 2016, https://doi.org/10.1155/2016/2861380.
[33] H. Z. Chen Chao; Jia, Ninghong; Duncan, Ian J.; Yang, Shenglai; Yang, YongZhi, "A machine learning model for predicting the minimum miscibility pressure of CO2 and crude oil system based on a support vector machine algorithm approach," *Fuel*, vol. 290, no. NA, pp. 120048-NA, 2021, https://doi.org/10.1016/j.fuel.2020.120048.
[34] M. Najmi *et al.*, "Estimating the Dissolution of Anticancer Drugs in Supercritical Carbon Dioxide with a Stacked Machine Learning Model," *Pharmaceutics*, vol. 14, no. 8, p. 1632, 2022, https://doi.org/10.3390/pharmaceutics14081632.
[35] G. Sodeifian, F. Razmimanesh, N. Saadati Ardestani, and S. A. Sajadian, "Experimental data and thermodynamic modeling of solubility of Azathioprine, as an immunosuppressive and anti-cancer drug, in supercritical carbon dioxide," *J. Mol. Liq.*, vol. 299, p. 112179, 2020, https://doi.org/10.1016/j.molliq.2019.112179.
[36] B. N. Singh and K. H. Kim, "Characterization and relevance of physicochemical interactions among components of a novel multiparticulate formulation for colonic delivery," *Int. J. Pharm.*, vol. 341, no. 1–2, pp. 143–151, 2007, https://doi.org/10.1016/j.ijpharm.2007.04.005.
[37] G. C. McDonald, "Ridge regression," *WIREs Comput. Stat.*, vol. 1, no. 1, pp. 93–100, 2009, https://doi.org/10.1002/wics.14.
[38] Q. Sun, W. X. Zhou, and J. Fan, "Adaptive Huber Regression," *J. Am. Stat. Assoc.*, vol. 115, no. 529, pp. 254–265, 2020, https://doi.org/10.1080/01621459.2018.1543124.

[39] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, https://doi.org/10.1023/A:1010933404324.
[40] R. Grbić, D. Kurtagić, and D. Slišković, "Stream water temperature prediction based on Gaussian process regression," *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7407–7414, 2013, https://doi.org/10.1016/j.eswa.2013.06.077.
[41] O. M. Claveria Enric; Torra, Salvador, "Modelling cross-dependencies between Spain's regional tourism markets with an extension of the Gaussian process regression model," *SERIEs*, vol. 7, no. 3, pp. 341–357, 2016, https://doi.org/10.1007/s13209-016-0144-7.
[42] F. Pedregosa *et al.*, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011, https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf?ref=https:/.

BIOGRAPHY OF AUTHORS

Arya Adhyaksa Waskita, received a bachelor degree from Department of Physics, Universitas Indonesia in 2002, a master degree from Department of Computer Science, Bogor Institute of Agriculture in 2007 and Doctoral degree from Faculty of Computer Science, Universitas Indonesia in 2016. His research interests include data science and HPC. Email: arya003@brin.go.id. Orcid: https://orcid.org/0000-0001-9552-1658.

Stevry Yushady CH Bissa, received the bachelor's degree from the Handayani Education Foundation College of Informatics Management and Computer Science, in 2013. His research interests include computer science, software engineering, and artificial intelligence. Email: stevry.yushady.ch.bissa@brin.go.id. Orcid ID: https://orcid.org/0000-0002-7710-388X.

Ika Atman Satya, received the B.S. degree in electrical engineering from the University of Brawijaya, Malang, Indonesia, in 1991. He is currently pursuing the degree in environmental management with Pakuan University, Bogor, Indonesia. He is also a Researcher/Engineer at the Research Center Computation, National Research and Innovation Agency (BRIN). His research interests include the IoT, microalgae bio sequestration using photobioreactor technology, and online environmental monitoring. E-mail: atmansatya@gmail.com. Orcid ID: https://orcid.org/0000-0002-5225-3881; Scopus ID: https://www.scopus.com/authid/detail.uri?authorId=57218935528.

Ratna Surya Alwi received the bachelor's degree from the Department of Chemical Engineering, Malang Institute of Technology, Indonesia, in 1999, the M. Sc. degree from the Department of Environmental Engineering, Hasanuddin Universitas, Indonesia, in 2010, and the doctoral degree in material sciences from Kanazawa University, in 2014. Her research interests include modelling, solubility, and supercritical fluids. Email: ratn017@brin.go.id. Orcid ID: https://orcid.org/0000-0002-2930-8223.