



OPEN

## Design of predictive model to optimize the solubility of Oxaprozin as nonsteroidal anti-inflammatory drug

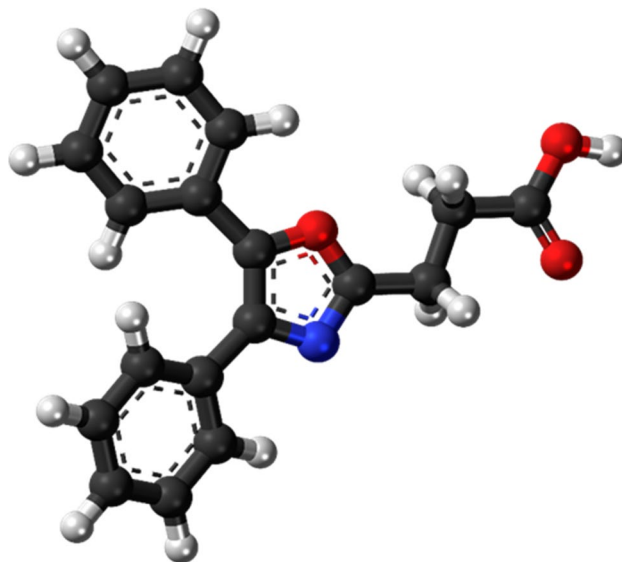
Sameer Alshehri<sup>1</sup>, Mohammed Alqarni<sup>2</sup>, Nader Ibrahim Namazi<sup>3</sup>, Ibrahim A. Naguib<sup>2</sup>, Kumar Venkatesan<sup>4</sup>, Yasser O. Mosaad<sup>5</sup>, Mahboubeh Pishnamazi<sup>6,7</sup>✉, Amal M. Alsubaiyel<sup>8</sup> & Mohammed A. S. Abourehab<sup>9,10</sup>

These days, many efforts have been made to increase and develop the solubility and bioavailability of novel therapeutic medicines. One of the most believable approaches is the operation of supercritical carbon dioxide fluid (SC-CO<sub>2</sub>). This operation has been used as a unique method in pharmacology due to the brilliant positive points such as colorless nature, cost-effectives, and environmentally friendly. This research project is aimed to mathematically calculate the solubility of Oxaprozin in SC-CO<sub>2</sub> through artificial intelligence. Oxaprozin is a nonsteroidal anti-inflammatory drug which is useful in arthritis disease to improve swelling and pain. Oxaprozin is a type of BCS class II (Biopharmaceutical Classification) drug with low solubility and bioavailability. Here in order to optimize and improve the solubility of Oxaprozin, three ensemble decision tree-based models including random forest (RF), Extremely random trees (ET), and gradient boosting (GB) are considered. 32 data vectors are used for this modeling, moreover, temperature and pressure as inputs, and drug solubility as output. Using the MSE metric, ET, RF, and GB illustrated error rates of 6.29E-09, 9.71E-09, and 3.78E-11. Then, using the R-squared metric, they demonstrated results including 0.999, 0.984, and 0.999, respectively. GB is selected as the best fitted model with the optimal values including 33.15 (K) for the temperature, 380.4 (bar) for the pressure and 0.001242 (mole fraction) as optimized value for the solubility.

### Abbreviations

SC-CO <sub>2</sub>	Supercritical carbon dioxide fluid
AI	Artificial intelligence
NSAID	Nonsteroidal anti-inflammatory
RF	Random forest
GB	Gradient boosting
ET	Extremely random trees
ML	Machine learning
API	Active pharmaceutical ingredient
SCFs	Supercritical fluids
DT	Decision tree
COP	Cross-over pressure

<sup>1</sup>Department of Pharmaceutics and Industrial Pharmacy, College of Pharmacy, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia. <sup>2</sup>Department of Pharmaceutical Chemistry, College of Pharmacy, Taif University, P. O. Box 11099, Taif 21944, Saudi Arabia. <sup>3</sup>Pharmaceutics and Pharmaceutical Technology Department, College of Pharmacy, Taibah University, Al Madinah Al Munawarah 30001, Saudi Arabia. <sup>4</sup>Department of Pharmaceutical Chemistry, College of Pharmacy, King Khalid University, Abha 62529, Kingdom of Saudi Arabia. <sup>5</sup>Department of Pharmacy Practice and Clinical Pharmacy, Faculty Pharmacy, Future Unibversity in Egypt, New Cairo 11835, Egypt. <sup>6</sup>Institute of Research and Development, Duy Tan University, Da Nang 550000, Viet Nam. <sup>7</sup>The Faculty of Pharmacy, Duy Tan University, Da Nang 550000, Viet Nam. <sup>8</sup>Department of Pharmaceutics, College of Pharmacy, Qassim University, Buraidah 52571, Saudi Arabia. <sup>9</sup>Department of Pharmaceutics, Faculty of Pharmacy, Umm Al-Qura University, Makkah 21955, Saudi Arabia. <sup>10</sup>Department of Pharmaceutics and Industrial Pharmacy, College of Pharmacy, Minia University, Minia 61519, Egypt. ✉email: mahboubbehpishnamazi@duytan.edu.vn



**Figure 1.** Schematic demonstration of Oxaprozin<sup>10</sup>.

Remarkable progression in pharmaceutical industry has paved the way towards creating novel therapeutic drugs for treating various challenging diseases<sup>1,2</sup>. Despite noteworthy development, poor solubility of active pharmaceutical ingredients (APIs) can be considered as the most prominent limitations for drug development<sup>3,4</sup>. Oxaprozin ( $C_{18}H_{15}NO_3$ ) can be recognized as one of the commonly-employed non-steroidal anti-inflammatory (NSAID) drug<sup>5,6</sup>. The analgesic and antipyretic characteristics of this propionic acid derivative has made it promising to appropriately alleviate the pain of acute/chronic disorders such as inflammation, swelling, osteoarthritis and rheumatoid arthritis<sup>7,8</sup>. Figure 1 presents the ball-stick demonstration of Oxaprozin. This NSAID drug possesses great ability to decline the formation of prostaglandin precursors from arachidonic acid via cyclo-oxygenase inhibition, which causes significant reduction in pain/inflammatory responses. Oxaprozin has shown superior efficacy compared to aspirin or piroxicam in the treatment of osteoarthritis<sup>9</sup>.

To improve the solubility of drugs, the indisputable role of solvents can't be ignored. These days, supercritical fluids (SCFs) are known as an innovative technique that demonstrates their efficiency for particle formation. This novel approach can overcome some disadvantages of conventional technologies such as crushing, crystallization and precipitation<sup>11,12</sup>. Supercritical carbon dioxide (SC-CO<sub>2</sub>) is being frequently applied to fractionate the precious components in pharmaceutical processes due to possessing noteworthy properties such as abundance, colorless nature, cost-effectiveness, and environmentally benign characteristic<sup>13</sup>. Due to the importance of solubility in SC-CO<sub>2</sub> for the design and development of novel drugs, the conduction of experimental investigation for evaluating the solubility of these drugs is of great importance<sup>14</sup>. Despite the great importance for obtaining the solubility of drugs, the existence of some economic/operational problems such as difficulty in solute-solvent interactions in SC-CO<sub>2</sub> system and high cost have limited the conduction of experimental investigation.

Therefore, development of mathematical modeling approaches to predict the solubility amount of disparate types of drugs can be an appropriate option to optimize the time and cost of processing. Nowadays, AI has been introduced as a promising predictive tool to measure the solubility of drugs, numerically. Apart from pharmacology, AI has found its indisputable role in disparate knowledge related to chemical engineering such as extraction, purification, separation, crystallization and chemical reactor engineering<sup>15</sup>. In most scientific fields, machine learning (ML) techniques are known as common computational procedures, including regression trees, neural networks, support vector machines. A variety of relationships between inputs and outputs are extracted by these models<sup>16-18</sup>.

The Decision Tree (DT) is one of the typically used learning models. A weak model is a simple predictor that is only likely to be better than a random estimator. The results of many base DT models are aggregated to form a stronger model in tree-based ensemble methods<sup>19,20</sup>.

Bagging and boosting are two of the most effective improvement strategies with Decision Trees. Bagging (Bootstrap Aggregating), developed by Breiman<sup>21</sup>, is one of the most basic and straightforward ensemble techniques, demonstrating outstanding performance while reducing variance and preventing overfitting. The Bagging algorithm is more diverse because of the bootstrap approach, which replicates and generates subsets of training data. All of the subsets are used to fit different basic estimators, and the final prediction results are compiled using a majority-vote method<sup>21,22</sup>.

One other ensemble method based on the Freund and Schapiro's study is boosting<sup>23</sup>. The aim of this research was optimization of Oxaprozin solubility within supercritical fluid by applying different machine learning models to find the best model for that.

By progressively reweighting the training data, this approach differs from Bagging in that it generates a diverse set of basic learners. A higher weight will be given to each sample whose estimation was weaker than

No	Temperature (K)	Pressure (bar)	Solubility (mole fraction)
1	$3.08 \times 10^2$	$1.20 \times 10^2$	$8.19 \times 10^{-5}$
2	$3.08 \times 10^2$	$1.60 \times 10^2$	$1.58 \times 10^{-4}$
3	$3.08 \times 10^2$	$2.00 \times 10^2$	$2.24 \times 10^{-4}$
4	$3.08 \times 10^2$	$2.40 \times 10^2$	$2.80 \times 10^{-4}$
5	$3.08 \times 10^2$	$2.80 \times 10^2$	$3.44 \times 10^{-4}$
6	$3.08 \times 10^2$	$3.20 \times 10^2$	$4.06 \times 10^{-4}$
7	$3.08 \times 10^2$	$3.60 \times 10^2$	$4.73 \times 10^{-4}$
8	$3.08 \times 10^2$	$4.00 \times 10^2$	$5.33 \times 10^{-4}$
9	$3.18 \times 10^2$	$1.20 \times 10^2$	$7.55 \times 10^{-5}$
10	$3.18 \times 10^2$	$1.60 \times 10^2$	$1.41 \times 10^{-4}$
11	$3.18 \times 10^2$	$2.00 \times 10^2$	$2.45 \times 10^{-4}$
12	$3.18 \times 10^2$	$2.40 \times 10^2$	$3.56 \times 10^{-4}$
13	$3.18 \times 10^2$	$2.80 \times 10^2$	$4.64 \times 10^{-4}$
14	$3.18 \times 10^2$	$3.20 \times 10^2$	$5.58 \times 10^{-4}$
15	$3.18 \times 10^2$	$3.60 \times 10^2$	$6.60 \times 10^{-4}$
16	$3.18 \times 10^2$	$4.00 \times 10^2$	$7.66 \times 10^{-4}$
17	$3.28 \times 10^2$	$1.20 \times 10^2$	$5.34 \times 10^{-5}$
18	$3.28 \times 10^2$	$1.60 \times 10^2$	$1.28 \times 10^{-4}$
19	$3.28 \times 10^2$	$2.00 \times 10^2$	$3.02 \times 10^{-4}$
20	$3.28 \times 10^2$	$2.40 \times 10^2$	$4.14 \times 10^{-4}$
21	$3.28 \times 10^2$	$2.80 \times 10^2$	$5.82 \times 10^{-4}$
22	$3.28 \times 10^2$	$3.20 \times 10^2$	$7.87 \times 10^{-4}$
23	$3.28 \times 10^2$	$3.60 \times 10^2$	$8.51 \times 10^{-4}$
24	$3.28 \times 10^2$	$4.00 \times 10^2$	$1.03 \times 10^{-3}$
25	$3.38 \times 10^2$	$1.20 \times 10^2$	$3.31 \times 10^{-5}$
26	$3.38 \times 10^2$	$1.60 \times 10^2$	$9.09 \times 10^{-5}$
27	$3.38 \times 10^2$	$2.00 \times 10^2$	$2.98 \times 10^{-4}$
28	$3.38 \times 10^2$	$2.40 \times 10^2$	$4.81 \times 10^{-4}$
29	$3.38 \times 10^2$	$2.80 \times 10^2$	$6.77 \times 10^{-4}$
30	$3.38 \times 10^2$	$3.20 \times 10^2$	$8.89 \times 10^{-4}$
31	$3.38 \times 10^2$	$3.60 \times 10^2$	$1.08 \times 10^{-3}$
32	$3.38 \times 10^2$	$4.00 \times 10^2$	$1.24 \times 10^{-3}$

**Table 1.** The whole dataset: 32 data vectors, where each vector has two input parameters (pressure and temperature) and one output (solubility).

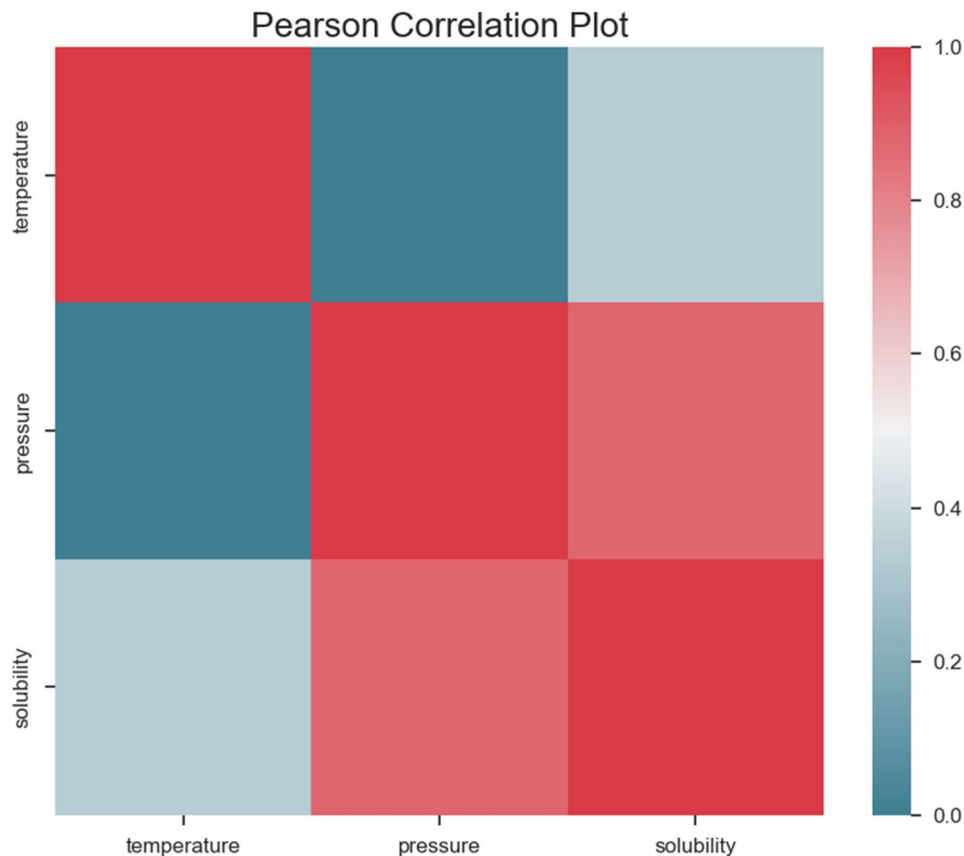
the previous estimator's in the subsequent training step. As a result, in subsequent bootstrap samples, it is more likely that training samples with weak estimates will appear, allowing bias to be effectively removed. Based on their prediction performance, the base estimators are weighted in the final Boosting algorithm model. A random forest model, Extra Trees, and Gradient Boosting model were all considered for inclusion in this research<sup>24–27</sup>.

## Experimental

Various predictive models in this research have been investigated and developed based on the experimental investigation of Khoshmaram et al. They experimentally measured the solubility of Oxaprozin using the combination of static and gravimetric techniques via a pressure–volume–temperature (PVT) cell<sup>14</sup>. This system can be filled with up to 0.4 L Oxaprozin and supercritical liquid. The adjustment of two momentous parameters for evaluating the solubility of drugs (temperature and pressure) in the PVT cell is an important advantage. In the PVT cell, increment of pressure causes the manufacturing of SC-CO<sub>2</sub> in the liquefaction unit. Then, the condensed solvent moves through the inline filter with the aim of purifying the solvent. In the next step, purified solvent enters a surge tank before the PVT cell. The controlling process of SC-CO<sub>2</sub> and Oxaprozin temperatures was implemented applying heating elements insulated by a PTFE layer.

## Data set

This study's dataset is derived from<sup>14</sup> that have just 32 data vectors. Each vector has two input parameters (pressure and temperature) and one output (solubility). The dataset is shown in Table 1 and Pearson correlation<sup>28</sup> of parameters are shown in Fig. 2.



**Figure 2.** Pearson correlation plot.

## Methodology

**Random forest and extra tree.** The random forest ensemble learning model is a tree-based technique that, like other ensemble learning methods, which is used to enhance the effectiveness of multiple base tree learners<sup>29</sup>. There will then be an unpruned regression tree built for every bootstrapped sample. This is what will happen next. Instead of using all the current predictors, a specified number of  $K$  base models are picked randomly to perform the function of split possibilities in this stage. This two-step operation will be iterated unto  $C$  decision trees with the above-mentioned characteristics are optimized, at which point unobserved data can be predicted by gathering the estimations of these  $C$  trees. Random forest uses a bagging strategy to boost tree diversity via constructing DTs using different training subsets, minimizing the model's total variance<sup>17</sup>. An RF regression predictor is expressed in the following equation:

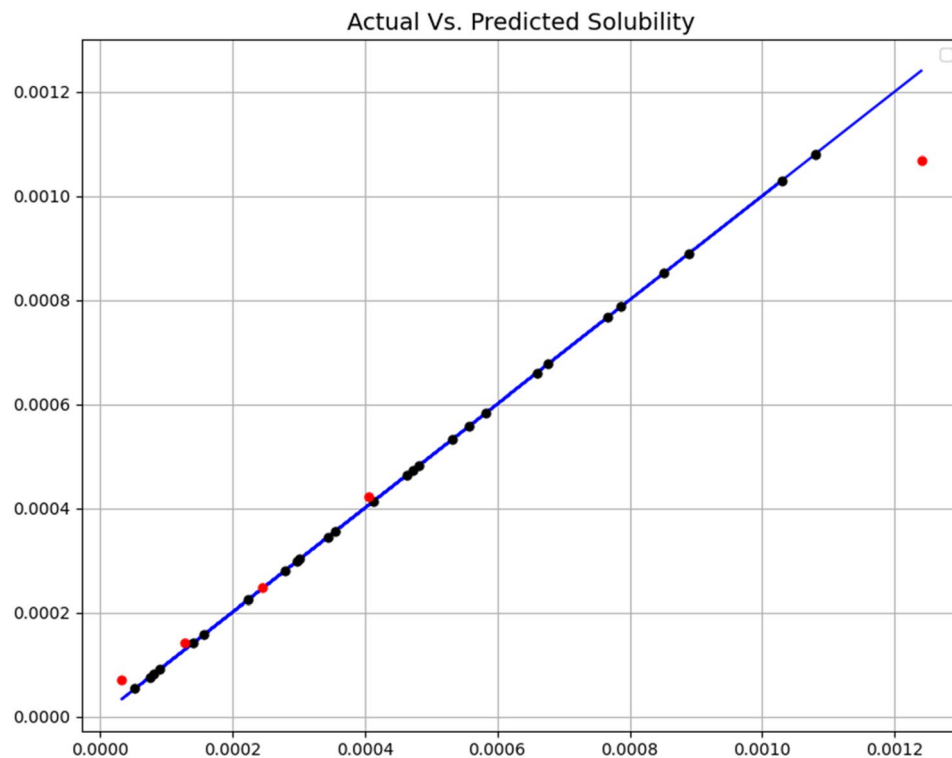
$$\hat{f}_{RF}^C(x) = \frac{1}{C} \sum_{i=1}^C T_i(x) \quad (1)$$

According to the previous equation,  $C$  refers to the count of decision trees,  $x$  identifies the data point, and  $T_i(x)$  refers to a unique DT built from bootstrap samples and a subset of entry variables. RF can predict out-of-bag error for the time being logging natively using samples which have not been selected in connection with the drive of this shaft during the bagging process. To compute an unbiased prediction of distribution error, this particular sub-association does not make use of any external data<sup>19,30</sup>. Assign substantial scores to each input variable. RF modifies one input variable while holding the others constant, and the model's average decrease is also assigned<sup>19</sup>.

Extra Trees (ET) are an overall tree-based approach like random forest. It strongly randomize both the cut point decision and the particularities of a tree node during its division Extra Tree becomes possible to categorize and regression tasks<sup>31,32</sup>.

As far as the differences are concerned, the two models are identical in that they develop multiple trees and divide nodes applying random subsets of functions, nevertheless, there are two major separations exist: Rather than using optimum splits, the ET uses randomized splits instead of bootstrap observations<sup>33</sup>.

**Gradient boosting.** Boosting is also an ensemble learning technique. Boosting comprises a sequence of base predictors rather than a single predictor to average them all together to improve prediction accuracy. In a stage-wise process, base estimators (decision trees here) are successively fitted to eliminate bias. At each phase, a



**Figure 3.** Expected and predicted solubility (ET model).

new learner is introduced to optimize the loss function. The first learner reduces the loss function to the smallest possible value using training data<sup>24,34,35</sup>. The residuals from the previous estimators are used by the following estimators. The gradient boosting method steps are depicted in the following Algorithm<sup>24,35,36</sup>:

Initialize  $F_0(x) = \operatorname{argmin}_p \sum_{i=1}^N L(y_i, P)$  (2)

For  $j \in \text{Error! Bookmark not defined.}$ :

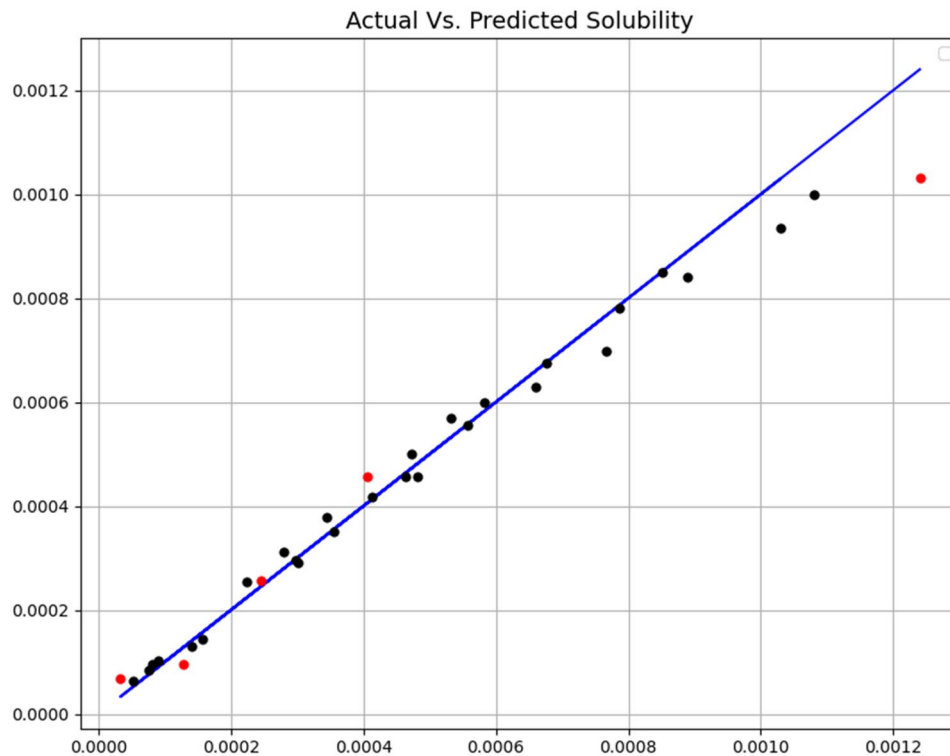
1. Compute the negative gradient
 
$$\bar{y}_i = - \left[ \frac{\partial L(y_i, F(x_i))}{\partial F_{x_i}} \right] \quad (3)$$
2. Create a model
 
$$a_m = \operatorname{argmin}_{a, \beta} \sum_{i=1}^N [\bar{y}_i - \beta h(x_i, a_j)]^2 \quad (4)$$
3. Select a step size for gradient descent as
 
$$p_k = \operatorname{argmin}_p \sum_{i=1}^N L(y_i, F_j - 1(x_i) + ph(x_i, a)) \quad (5)$$
4. Update the prediction of  $F(x)$ 

$$F_j(x) = F_j - 1(x) + p_k h(x, a_j) \quad (6)$$

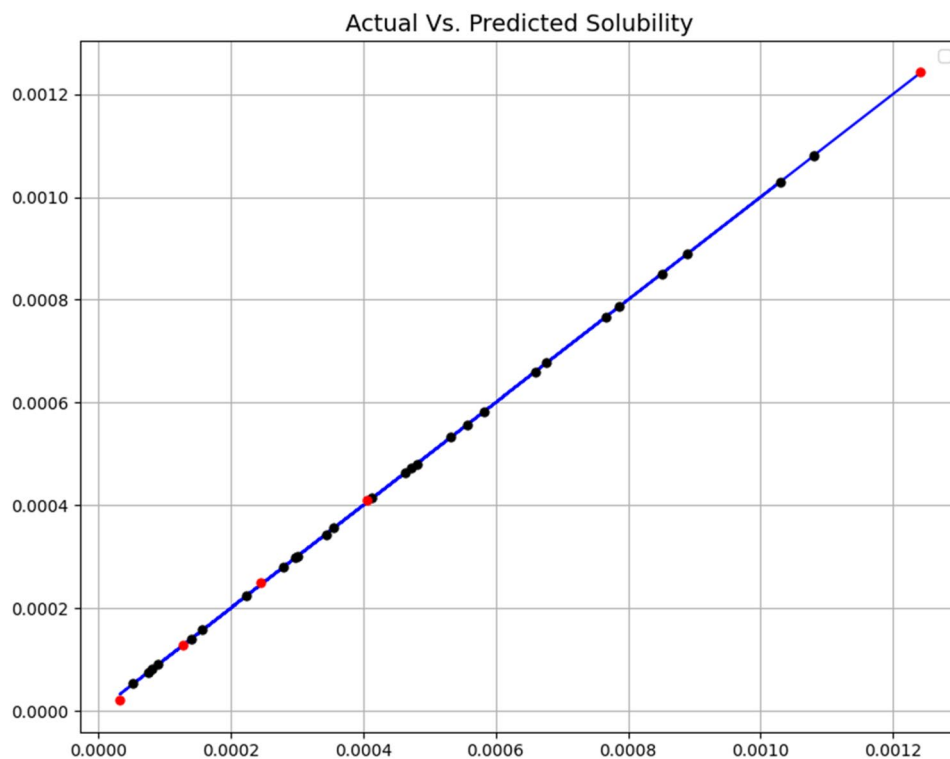
Output: the aggregated prediction function  $F_M(x)$

## Results

The tuning of the hyper-parameters of the mentioned models is based on a search grid. All three final models were evaluated by R-square and MSE criteria. Additionally, some visualization results were made, which will be discussed later. Figures 3, 4 and 5 show a comparison of expected values and predicted amounts. In the below figures, the blue line indicates the expected amounts and the points of the predicted values (red for the test data and black for the training data). In addition, Table 2 shows quantitative metrics to compare the three implemented



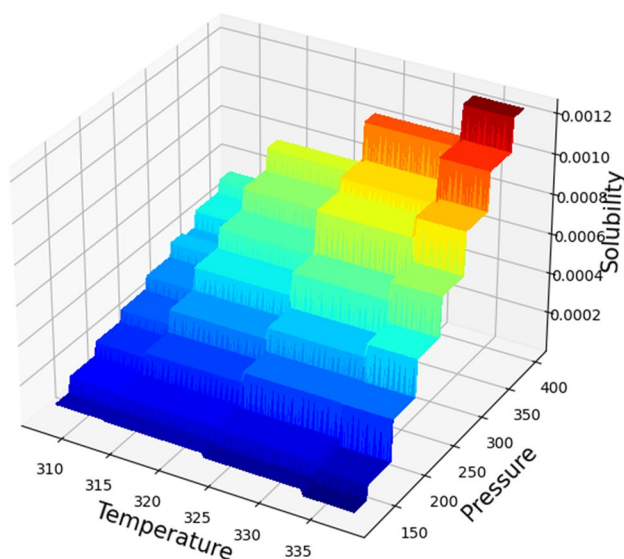
**Figure 4.** Expected and predicated solubility (RF model).



**Figure 5.** Expected and predicated solubility (GB model).

Models	MSE	R <sup>2</sup>
ET	$6.29 \times 10^{-9}$	0.999
RF	$9.71 \times 10^{-9}$	0.984
GB	$3.78 \times 10^{-11}$	0.999

**Table 2.** Final model results.



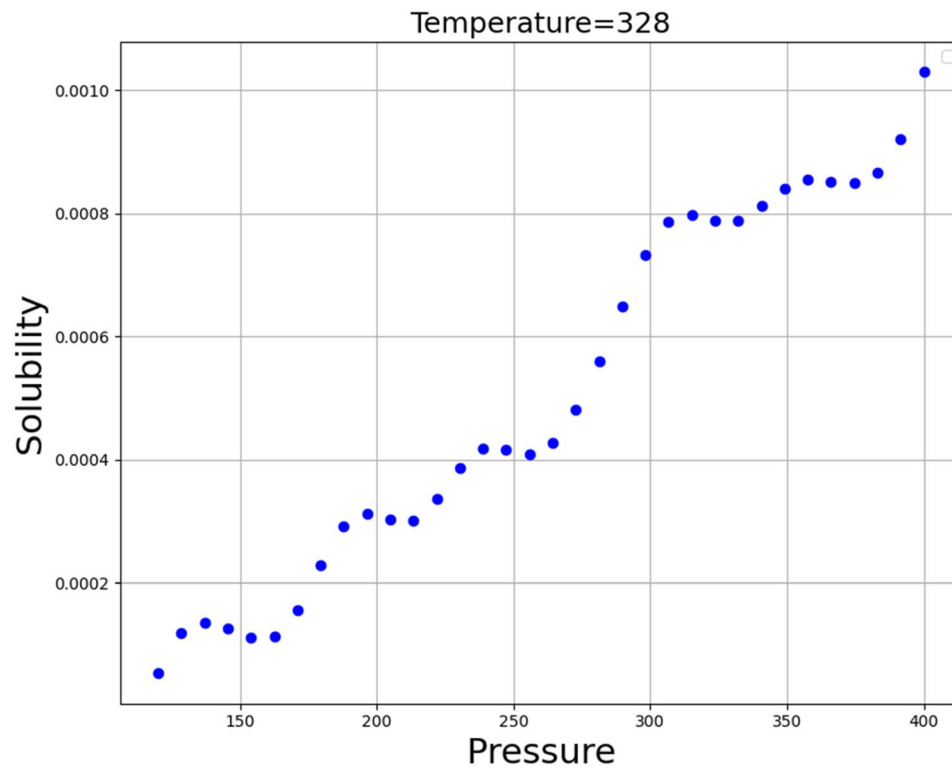
**Figure 6.** Input–output projection (GB).

models with the optimal hyper-parameters. Comparison of tabulated results in Table 2 has confirmed the fact that the GB is the most accurate and general model ( $R^2 = 0.999$  and  $MSE = 3.78E-11$ ), which has been used as the main model for the rest of the analysis.

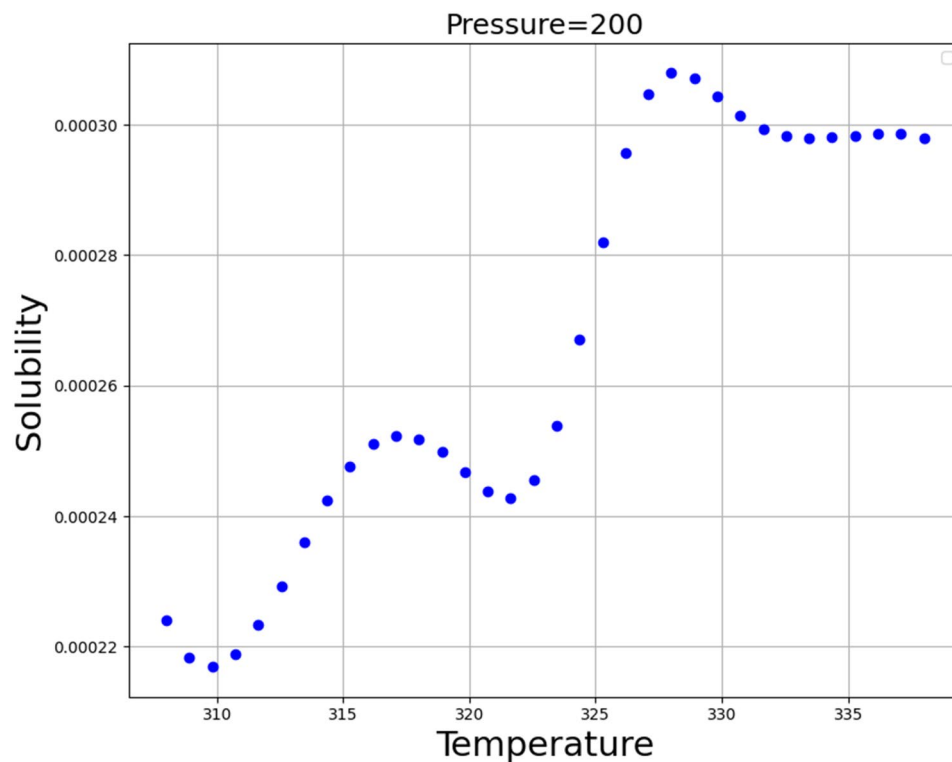
The simultaneous impacts of temperature and pressure as two prominent input parameters on the solubility as the only output is shown in 3D in Fig. 6. Furthermore, by holding each of the inputs fixed, the two-dimensional Figs. 7 and 8 are displayed. These figures correspond to the reality of the optimal values in Table 3. It can be perceived from the figures that the pressure of system has positive impact on the solubility of Oxaprozin in supercritical system. Indeed, increase in the pressure can improve the solvent density, which consequently intensifies the solvating power of the SC-CO<sub>2</sub> system. Although pressure has direct connection with the solubility of drug, the impact of temperature is entirely indirect. To evaluate the effect of temperature on drug solubility, the role of sublimation pressure and density above and below the cross-over pressure (COP) must be analyzed. At the pressures above the COP, the encouraging influence of sublimation pressure on solubility dominates the deteriorative impact of density reduction. Therefore, at these pressures, temperature increment significantly enhances the solubility in SC-CO<sub>2</sub> system. At pressures below the COP, the destructive impact of density decrement overcomes the positive effect of sublimation pressure. Therefore, at these amounts of pressures, increasing the temperature significantly reduces the solubility in SC-CO<sub>2</sub>. By concentrating on Table 3, it is recognized that the pressure and the temperature of 380.4 bar and 333.15 K are the optimum factors for reaching the greatest amount of Oxaprozin solubility.

## Conclusion

Now a days, numerous efforts have been made to develop green and efficient solvents to overcome the functional/operational detriments of organic solvents. Nowadays, SC-CO<sub>2</sub> has been introduced as a prevalently employed liquid solvent to fractionate the valuable components and increase the solubility of drugs in pharmaceutical processes because of its remarkable advantages (i.e., abundance, cost-effectiveness, and environmentally benign characteristic). In this paper, disparate types of numerical models were proposed via AI technique to anticipate the optimum value of Oxaprozin in SC-CO<sub>2</sub>. In this study, three ensemble decision tree-based models were used to model the problem: extremely random tree (ET), random forest (RF), and Gradient Tree Boosting (GB). This problem's available data consists of 32 data vectors with two inputs of temperature and pressure and an output of solubility. ET, RF, and GB had MSE error rates of  $6.29E-09$ ,  $9.71E-09$ , and  $3.78E-11$ . They also have R-squared scores of 0.999, 0.984, and 0.999, respectively. The final model chosen is GB, with the following optimal values:  $T = 33.15$ ,  $P = 380.4$ , and solubility = 0.001242, which shows the greatest amount of Oxaprozin solubility.



**Figure 7.** Solubility (mole fraction) based on pressure (bar), temperature (°K).



**Figure 8.** Solubility (mole fraction) base on temperature (°K), pressure (bar).



Temperature (K)	Pressure (bar)	Solubility (mole fraction)
333.15	380.4	0.001242

**Table 3.** Optimal values.

## Data availability

All data are available within the published paper.

Received: 31 May 2022; Accepted: 25 July 2022

Published online: 30 July 2022

## References

- Zeng, X., Tu, X., Liu, Y., Fu, X. & Su, Y. Toward better drug discovery with knowledge graph. *Curr. Opin. Struct. Biol.* **72**, 114–126 (2022).
- Zhuang, W., Hachem, K., Bokov, D., Ansari, M. J. & Nakhjiri, A. T. Ionic liquids in pharmaceutical industry: A systematic review on applications and future perspectives. *J. Mol. Liquids* **349**, 118145 (2021).
- Chakravarty, P., Famili, A., Nagapudi, K. & Al-Sayah, M. A. Using supercritical fluid technology as a green alternative during the preparation of drug delivery systems. *Pharmaceutics* **11**, 629 (2019).
- Savjani, K. T., Gajjar, A. K. & Savjani, J. K. Drug solubility: Importance and enhancement techniques. *Int. Sch. Res. Not.* **2012** (2012).
- Greenblatt, D. *et al.* Oxaprozin pharmacokinetics in the elderly. *Br. J. Clin. Pharmacol.* **19**, 373–378 (1985).
- Kean, W. F. *Oxaprozin: Kinetic and Dynamic Profile in the Treatment of Pain* 1275–1277 (Taylor & Francis, 2004).
- Fischer, J. & Ganellin, C. R. Analogue-based drug discovery. *Chem. Int. Newsmag. IUPAC* **32**, 12–15 (2010).
- Ganellin, C. R. *Analogue-Based Drug Discovery II* (Wiley, 2010).
- Miller, L. G. Oxaprozin: A once-daily nonsteroidal anti-inflammatory drug. *Clin. Pharm.* **11**, 591–603 (1992).
- File:Oxaprozin molecule ball.png. *Wikimedia Commons, the Free Media Repository* (2022). Retrieved 10:52, May 17, 2022 from [https://commons.wikimedia.org/w/index.php?title=File:Oxaprozin\\_molecule\\_ball.png&oldid=644050983](https://commons.wikimedia.org/w/index.php?title=File:Oxaprozin_molecule_ball.png&oldid=644050983).
- Baldelli, A., Boraey, M. A., Nobes, D. S. & Vehring, R. Analysis of the particle formation process of structured microparticles. *Mol. Pharm.* **12**, 2562–2573 (2015).
- Misra, S. K. & Pathak, K. Supercritical fluid technology for solubilization of poorly water soluble drugs via micro- and nanosized particle generation. *ADMET DMPK* **8**, 355–374 (2020).
- Bahramifar, N., Yamini, Y. & Shamsipur, M. Investigation on the supercritical carbon dioxide extraction of some polar drugs from spiked matrices and tablets. *J. Supercrit. Fluids* **35**, 205–211 (2005).
- Khoshmaram, A. *et al.* Supercritical process for preparation of nanomedicine: Oxaprozin case study. *Chem. Eng. Technol.* **44**, 208–212 (2021).
- Chinh Nguyen, H. *et al.* Computational prediction of drug solubility in supercritical carbon dioxide: Thermodynamic and artificial intelligence modeling. *J. Mol. Liquids* **354**, 118888 (2022).
- Bishop, C. M. Pattern recognition. *Mach. Learn.* **128** (2006).
- Rodriguez-Galiano, V., Sanchez-Castillo, M., Chica-Olmo, M. & Chica-Rivas, M. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. *Ore Geol. Rev.* **71**, 804–818 (2015).
- Goodfellow, I., Bengio, Y. & Courville, A. Machine learning basics. *Deep Learn.* **1**, 98–164 (2016).
- Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. *Classification and Regression Trees* (Routledge, 2017).
- Xue, M., Su, Y., Li, C., Wang, S. & Yao, H. Identification of potential type II diabetes in a large-scale Chinese population using a systematic machine learning framework. *J. Diabetes Res.* **2020** (2020).
- Breiman, L. Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996).
- Borra, S. & Di Ciaccio, A. Improving nonparametric regression methods by bagging and boosting. *Comput. Stat. Data Anal.* **38**, 407–420 (2002).
- Freund, Y. & Schapire, R. E. Experiments with a new boosting algorithm. In *ICML* 148–156 (Citeseer, 1996).
- Mason, L., Baxter, J., Bartlett, P. & Frean, M. Boosting algorithms as gradient descent. *Adv. Neural Inf. Process. Syst.* **12** (1999).
- Pardoe, D. & Stone, P. Boosting for regression transfer. In *ICML* (2010).
- Wu, Q., Burges, C. J., Svore, K. M. & Gao, J. Adapting boosting for information retrieval measures. *Inf. Retr.* **13**, 254–270 (2010).
- Ying, C., Qi-Guang, M., Jia-Chen, L. & Lin, G. Advance and prospects of AdaBoost algorithm. *Acta Autom. Sin.* **39**, 745–758 (2013).
- Benesty, J., Chen, J., Huang, Y. & Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing* 1–4 (Springer, 2009).
- Jiang, R., Tang, W., Wu, X. & Fu, W. A random forest approach to the detection of epistatic interactions in case-control studies. *BMC Bioinform.* **10**, 1–12 (2009).
- Zhang, J. *et al.* Rapid evaluation of texture parameters of Tan mutton using hyperspectral imaging with optimization algorithms. *Food Control* **135**, 108815 (2022).
- Geurts, P., Ernst, D. & Wehenkel, L. Extremely randomized trees. *Mach. Learn.* **63**, 3–42 (2006).
- Dutta, S., Mukherjee, U. & Bandyopadhyay, S. K. Pharmacy impact on Covid-19 vaccination progress using machine learning approach. *J. Pharm. Res. Int.* **33**, 202–217 (2021).
- Song, Y.-Y. & Ying, L. Decision tree methods: Applications for classification and prediction. *Shanghai Arch. Psychiatry* **27**, 130 (2015).
- Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **29**, 1189–1232 (2001).
- Truong, V.-H., Vu, Q.-V., Thai, H.-T. & Ha, M.-H. A robust method for safety evaluation of steel trusses using gradient tree boosting algorithm. *Adv. Eng. Softw.* **147**, 102825 (2020).
- Xu, Q. *et al.* PDC-SGB: Prediction of effective drug combinations using a stochastic gradient boosting algorithm. *J. Theor. Biol.* **417**, 1–7 (2017).

## Acknowledgements

The author would like to extend their science appreciation to Taif university Researchers Supporting Project Number (TURSP-2020/309), Taif University, Taif Saudi Arabia. The authors extend their appreciation to

the Deanship of Scientific Research at King Khalid University for funding this work through Large Groups Project under grant number (RGP2/50/43). The authors would like to thank the Deanship of Scientific Research at Umm Al-Qura University for Supporting this work by Grant code (22UQU4290565DSR45).

### Author contributions

S.A.: supervision, software, data analysis, validation, modeling, original draft writing, methodology. M.A.: investigation, original draft writing, data analysis, methodology. N.I.N.: writing and editing, data analysis, investigation, conceptualization. I.A.N.: writing and editing, methodology, data analysis, investigation, validation, conceptualization. K.V.: investigation, original draft writing, data analysis, methodology. Y.O.M.: writing and editing, modeling and validation, data analysis, investigation, conceptualization. M.P.: supervision, software, data analysis, validation, modeling, original draft writing, methodology. A.M.A.: investigation, writing, editing, data analysis, validation. M.A.S.A.: supervision, data analysis, validation, original draft writing, methodology.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to M.P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022