Review Article

# Solubility of Pharmaceutical Compounds in Binary Solvents: A Tutorial Review on Numerical Methods

Elaheh Rahimpour, Nima Nazemieh, Tahir Suleymanov, William E. Acree, Jr., Abolghasem Jouyban

DOI: 10.34172/PS.026.43440

To appear in: Pharmaceutical Science (https://pstbzmed.com/)

Received date: 1 Nov 2025

Revised date: 24 Dec 2026

Accepted date: 3 Jan 2026

Please cite this article as: Rahimpour E, Nazemieh N, Tahir Suleymanov T, William E. Acree WE, Jouyban A. Solubility of pharmaceutical compounds in binary solvents: a tutorial review on numerical methods. Pharm Sci. 2026. doi: 10.34172/PS.026.43440

# Solubility of Pharmaceutical Compounds in Binary Solvents: A Tutorial Review on Numerical Methods

**Elaheh Rahimpour[1] https://orcid.org/0000-0002-4037-9436, Nima Nazemieh[1], Tahir Suleymanov[2], William E. Acree, Jr.[3], Abolghasem Jouyban[1,4,♦] https://orcid.org/0000-0002-4670-2783**

[1] *Pharmaceutical Analysis Research Center, Pharmaceutical Sciences Institute, Tabriz University of Medical Sciences, Tabriz, Iran*

[2] *Department of Pharmaceutical Chemistry, Azerbaijan Medical University, Baku, Azerbaijan*

[3] *Department of Chemistry, University of North Texas, Denton, TX 76203, USA*

[4] *Faculty of Pharmacy, Tabriz University of Medical Sciences, Tabriz, Iran*

♦ Corresponding author email: ajouyban@hotmail.com

**Abstract**

In this work the most widely used models applied for mathematical representation of the solubility data of pharmaceutical compounds in binary solvents were briefly reviewed. Then the utilizing SPSS software for solubility data modeling with some well-known mathematical models such as the van't Hoff, the Apelblat, the combined nearly ideal binary solvent/Redlich-Kister, Jouyban-Acree and the Jouyban-Acree-van't Hoff, the modified Wilson, and the $\lambda h$ is described as a video file. The video is prepared in two parts. The first part is a general explanation of solubility and the need for modeling these data. The second part is about how to use SPSS software for data modeling with the mentioned models.

**Key words**: Solubility, Data modeling, Cosolvency models, SPSS

## Introduction

Mathematical modeling is a concept that involves the use of mathematical techniques and tools to represent, analyze, validate and understand real-world systems, processes, or phenomena. It is an interdisciplinary field that combines elements of mathematics, experimental sciences, and engineering to develop models that help us gain insights, make predictions, and optimize various aspects.[1] Mathematical modeling philosophy comprises several key aspects, each contributing to the development of accurate and insightful models. These aspects are data validity, model simplicity, interpretability, and closeness to theoretical concepts. Data validity is essential for ensuring that the models are built upon reliable and accurate information. Techniques such as data cleaning, outlier detection, and statistical tests help to maintain data validity, which is crucial for developing models that can make accurate predictions and provide meaningful insights.[2-4] Model simplicity focuses on creating models that are easy to understand, interpret, and solve. Simplifying a model allows researchers to concentrate on the essential features and relationships between variables, particularly when dealing with complex systems. Interpretability is another crucial aspect of mathematical modeling, as it enables researchers to understand and explain the model's results and predictions in a clear and meaningful way. Interpretable models help to identify key factors influencing the system's behavior, understand relationships between variables, and make informed decisions based on these insights.[5]

Lastly, the closeness of a mathematical model to theoretical concepts ensures that it is grounded in relevant principles and ideas from the field it represents. This principle guarantees that the model incorporates essential features and relationships between variables, leading to more accurate and reliable predictions. By adhering to established theories and concepts, researchers

can develop models that reflect the true nature of the system being studied, facilitating the integration of new findings and advancements in the field.[6]

When developing a model, it is crucial to consider various factors to ensure its effectiveness and suitability for the given problem. One of the primary aspects to focus on is the availability and quality of valuable data. The data used should be representative of the real-world situation and collected through reliable sources, as it directly impacts the model's performance and accuracy.[7-11] Another essential consideration is the computational simplicity of the model. In scenarios where large datasets or real-time applications are involved, it is vital to minimize the computational requirements of the model. This can be achieved by employing simpler mathematical operations and algorithms that do not consume excessive computational resources.[12] The nature of the model, such as white box, grey box, or black box, also plays a significant role in its application. White box models provide complete transparency into the inner workings, making them suitable for applications where understanding the underlying processes is essential. Grey box models offer partial transparency, combining both known and unknown elements, and are useful when some parts of the system are well-understood while others are not. Black box models are entirely opaque, focusing on the input-output relationship, and are widely used in machine learning and artificial intelligence applications where the exact inner workings are not crucial.[13] It should be noted that applications of the artificial intelligence based methods on solubility data attracted more and more attentions in recent years.[14-18] These methods are proving highly effective at predicting chemical solubility, often surpassing traditional models. For instance, Cui et al.[19] used a deep neural network on 10,000 compounds to predict solubility in solvent mixtures. Other advancements include Transformer architectures

for SMILES string canonicalization [20] and a Bidirectional LSTM for physical property predictions.[21] Beyond these, artificial neural networks have been applied to pharmaceutical solubility,[22, 23] while studies comparing graph neural networks, random forests, and ensemble methods [24-28] underscore machine learning's versatility in providing accurate, industrially relevant solubility predictions. However, these models will not be discussed in this work.

Besides what has been mentioned, it is essential to determine whether the model should be linear or non-linear, depending on the nature of the relationship between input and output variables. Linear models assume a linear relationship, making them suitable for problems with clear cause-and-effect relationships. In contrast, non-linear models do not assume a linear relationship, allowing for more complex and diverse relationships, and making them ideal for real-world problems exhibiting intricate patterns and behaviors.[29] By considering these factors, one can create a model that accurately represents the system and is efficient in terms of computation and performance.

Mathematical modeling plays a crucial role in understanding complex processes and phenomena by providing a structured framework to represent and analyze real-world situations. By considering the mechanism analysis of a process, one can gain insights into the underlying principles governing the system, which helps in developing more accurate and reliable models. Mechanism analysis involves breaking down a process into its constituent components and understanding their interactions. This allows to identify key variables and relationships that influence the system's behavior.[30] Moreover, data fitting is an essential aspect of mathematical modeling, as it helps to validate the model's assumptions and parameters. It also assesses the quality and reliability of data by examining how well the chosen model fits the data and predicts

new observations. This process helps to identify potential issues such as outliers, missing values, or inconsistencies. To handle outliers, statistical tests like the t-test or F-test are used. For instance, the t-test can help determine whether a data point is significantly different from the rest of the data, and if it is, one may choose to remove it or repeat the data collection to ensure more precise results. In summary, by comparing the model's predictions with actual data, one can assess the model's accuracy and make necessary adjustments. This iterative process ensures that the model is a good fit for the observed data, thereby increasing its reliability for future predictions, which is the primary objective of mathematical modeling for un-measured data. The prediction process involves using the established model to extrapolate outcomes for variables that have not been directly measured or observed.[31] By doing so, one can gain valuable insights into the system's behavior under different conditions and make informed decisions based on these predictions. These techniques are particularly useful in understanding complex systems and processes, such as the solubility field. In the pharmaceutical industry, solubility is a major challenge in drug development.[32] According to estimates, more than half of newly discovered compounds are abandoned due to their low aqueous solubility. Oral administration is a common and popular method of drug delivery due to its ease of administration, high patient acceptance, affordability, and reduced need for sterile conditions. As a result, most pharmaceutical companies produce drugs in oral forms. However, designing oral pharmaceutical forms poses a significant challenge due to the low bioavailability of drugs. Bioavailability is influenced by several factors, including dissolution rate, solubility, drug permeability, and drug metabolism before entering the bloodstream. Among these factors, low aqueous solubility and low permeability are the primary reasons for the low bioavailability of drugs.[33] In order to optimize bioavailability, it is

essential to measure the solubility of drug candidates as early as possible in the discovery process using established methods for solubility determination including manual methods[34] and smart/synthetic methods.[35]

Beyond experimental determinations of solubility, computational models based on thermodynamic principles can be employed to predict solute solubility in the mixed solvents. These models can be classified into two distinct categories: (1) theoretical models and (2) semi-theoretical models.[36] Theoretical models are valuable tool for elucidating the underlying solubility behavior in the mixed solvents, while semi-theoretical models are more effective in interpolating experimental data. Theoretical models, also known as predictive models, are advantageous in that they can generate solubility predictions without requiring any prior experimental data. In contrast, correlative models, which involve semi-theoretical modeling, rely on curve-fitting parameters to establish relationships between experimental solubility data and related parameters.[37] Semi-theoretical models serve as the foundation for theoretical models. The empirical solubility models allowed for the establishment of correlations between solubility and some known parameters in solubility study procedures such as temperature and solvent composition. These correlations can then be used as a starting point for theoretical models, which aim to explain the underlying physical mechanisms that govern solubility behavior.

The main aim of this work is to explain some well-known mathematical models such as the van't Hoff, the Apelblat, the combined nearly ideal binary solvent/Redlich-Kister (CNIBS/R-K), Jouyban-Acree and the Jouyban-Acree-van't Hoff, the modified Wilson, and the $\lambda h$ which commonly used for solubility data correlation in the literature. Besides these explanations, a supplementary video file provides a comprehensive, step-by-step guide on how to utilize SPSS software to develop and

apply these models for solubility data analysis.

**Protocol**

***Principal of the mathematical linear models for solubility data***

*van't Hoff equation*

The van't Hoff equation is a mathematical model that relates the change in solubility of a solute to the change in temperature. It is widely used to understand the behavior of solutes in different solvents and to predict their solubility in a given solvent at various temperatures. By applying this equation, one can gain insights into the thermodynamic processes driving solute dissolution.[38] The general form of the model is as:

$$\ln x = A + \frac{B}{T} \quad (1)$$

The coefficients of the model are *A* and *B*.

*The Apelblat equation*

The non-linear relationship between ln $x_T$ and temperature was also modeled using the three-parameter Apelblat equation. This model provides a more accurate fit over wide temperature ranges than the van't Hoff equation.[38] This equation is written as:

$$\ln x = A + \frac{B}{T} + C \ln(T) \quad (2)$$

here *A*, *B*, and *C* are equation parameters.

*The CNIBS/R-K model*

The CNIBS/R-K model is a multi-linear cosolvency model to describe the solubility behavior of a solute in binary solvent mixtures at an isothermal condition. The model combines the concept of ideal binary mixing, which assumes that the solvent mixture behaves as a single solvent with ideal properties with the Redlich-Kister equation, which is a mathematical expression used to describe the non-ideal behavior of the solutions. The model is particularly useful for predicting the solubility of a solute in binary solvent mixtures at a specific temperature, as it takes into account the interactions between the solute and the solvent components. The CNIBS/R-K model equation can be given as:[39]

$$\ln x_m = w_1 \ln x_1 + w_2 \ln x_2 + w_1 w_2 \sum_{i=0}^{2} S_i (w_1 - w_2)^i \quad (3)$$

where $x_1$ and $x_2$ are mole fraction solubilities of a studied solute in mono-solvents 1 and 2, $x_m$ is the solute solubility in the binary mixtures, and $w_1$ and $w_2$ are mass fractions of solvent 1 and solvent 2 in the absence of the solute. The $S_i$ terms are the model parameters.

*Jouyban-Acree model*

The Jouyban-Acree model includes an additional term that take into account the effects of temperature on solubility. It is often referred to as a "temperature-dependent" model, as it incorporates temperature-dependent terms in addition to solvent composition terms that allow for more accurate predictions of solubility values over a range of temperatures and solvent compositions.[40] The model is:

$$\ln x_{m,T} = w_1 \ln x_{1,T} + w_2 \ln x_{2,T} + \frac{w_1 w_2}{T} \sum_{i=0}^{2} J_i (w_1 - w_2)^i \quad (4)$$

In which $x_{1,T}$ and $x_{2,T}$ are mole fraction drug solubilities in the mono-solvents 1 and 2, $x_{m,T}$ is the drug solubility in the solvent mixture at temperature *T*. The $J_i$ parameters are obtained by linear

regression analysis of $\ln x_{m,T} - w_1 \ln x_{1,T} - w_2 \ln x_{2,T}$ against $\frac{w_1 w_2}{T}$, $\frac{w_1 w_2 (w_1 - w_2)}{T}$, $\frac{w_1 w_2 (w_1 - w_2)^2}{T}$.

*Jouyban-Acree-van't Hoff model*

By integrating the Jouyban-Acree and van't Hoff models, the strengths of both models are combined and provide a more comprehensive understanding of the complex interactions between solutes and solvents, allowing for more precise predictions and correlations. The general form of the Jouyban-Acree-van't Hoff equation is given by:[40]

$$\ln x_{m,T} = w_1 \left(A_1 + \frac{B_1}{T}\right) + w_2 \left(A_2 + \frac{B_2}{T}\right) + \frac{w_1 w_2}{T} \sum_{i=0}^{2} J_i (w_1 - w_2)^i \quad (5)$$

$A_1$, $B_1$, $A_2$ and $B_2$ are the van't Hoff model's constants (intercept and slope) obtained by plotting $\ln x_{m,T}$ against $1/T$ in the mono-solvents at various temperatures. $J_i$ terms are computed using linear regression of $(\ln x_{m,T} - w_1 \left(A_1 + \frac{B_1}{T}\right) - w_2 \left(A_2 + \frac{B_2}{T}\right))$ vs $\frac{w_1 . w_2}{T}$, $\frac{w_1 . w_2 (w_1 - w_2)}{T}$, and $\frac{w_1 . w_2 (w_1 - w_2)^2}{T}$.

***Principal of the mathematical non-linear models for solubility data***

*The modified Wilson model*

The modified Wilson is a non-linear model used to describe the solubility of a solute in a binary mixture of solvents at a given temperature. It is an extension of the original Wilson equation. The modified Wilson equation takes into account the non-linear behavior of the solubility curve. This non-linearity is particularly important when dealing with binary mixtures of solvents, as the solubility of a solute can be affected by both the composition of the solvent and the temperature.[41] This model is as:

$$-\ln x_m = 1 - \frac{w_1 (1 + \ln x_1)}{w_1 + w_2 \lambda_{12}} - \frac{w_2 (1 + \ln x_2)}{w_1 \lambda_{21} + w_2} \quad (8)$$

$\lambda_{12}$ and $\lambda_{21}$ are the equation parameters.

*λh equation*

An additional non-linear model for fitting solubility data is the *λh* equation, also known as the Buchowski-Ksiazczak equation, which can be employed to predict solubility in binary mixed solvents at a constant temperature.[42] The equation like other solubility models take into account the non-ideal behavior of the solvent and the interactions between the solute and the solvent molecules. The model is given as:

$$\ln\left[1+\frac{\lambda(1-x)}{x}\right]=\lambda h\left[\frac{1}{T}-\frac{1}{T_m}\right] \quad (9)$$

where *λ* and *h* are the model parameters and $T_m$ is the solute' melting point.

**Discussion**

To model solubility data with a mathematical model, different software tools such as MATLAB, R, Excel, Python, SPSS, Mathcad, etc. can be used. Each of these software tools has its own strengths and weaknesses, and the choice of which one to use ultimately depends on the specific needs of the project and the user's familiarity with the software. Herein, we explain the details of the computations on the solubility of deferiprone in polyethylene glycol 400 + 2-propanol mixtures at various temperatures using SPSS software which was previously published by own research group.[43] To utilize SPSS software for modeling solubility data with various mathematical models, it needs to start by preparing the data. This includes importing the dataset into SPSS and ensuring that it includes the necessary variables, such as solubility as the dependent variable and temperature and solvent mass fraction as the independent variables. Additionally, other variables that could be relevant to the specific model must be added to a *.sav* file. A *.sav* file is a

file format used by SPSS to store data. It is also important to note that certain models may require additional steps or transformations of the data before fitting. For example, it may need to log-transform temperature or solubility values before fitting certain models. After selecting the model, the SPSS regression module is used to fit the model to the dataset. This involves entering the model equation into the "Model" field, specifying the dependent and independent variables, and running the analysis. After running the analysis in SPSS, an *output* file is created. The *output* file provides a summary of the results, including the estimated coefficients, R-squared value, F-statistic, residual plots and other descriptions which are defined in the corresponding *syntax* file. A *syntax* file (or *.SPS*) is a text file that contains SPSS commands or instructions that can be executed to perform specific tasks or analyses.

When evaluating the results of the analysis, one may want to examine the output to determine if the model is a good fit for the data. This includes checking the residual errors by computation of the mean relative deviation (*MRD* %) of back-calculated data for evaluating the accuracy of these models. The original form of the equation is as:

$$\%MRD = \frac{100}{N}\sum\left(\frac{|Calculated\ value - Observed\ value|}{Observed\ value}\right) \quad (10)$$

*N* is the number of data points.

In an ideal mode, the accuracy criterion for a correlative model should be close to zero, indicating that the data perfectly fit the model. However, in practice, this is rarely achieved due to experimental and measurement errors associated with the data, which typically result in a value greater than zero.[44] A more realistic approach may be to consider the accuracy criterion within the range of uncertainty values for repeated experiments (typically 1-10%).[36] However, in most of the works, *MRD*%<30% was considered as a relatively acceptable error for data modeling

procedures.

Once it is determined that the model is a good fit for investigated data, you can interpret the results to understand the relationship between the dependent variable (such as solubility in this work) and the independent variables (such as temperature and solvent mass fraction in this work). The estimated coefficients obtained from the running regression in SPSS can be used to provide a set of numerical values that represent the relationships between the variables in the studied model. These coefficients are used to quantify the magnitude and direction of the relationships between the independent variables and the dependent variable. For example, if the estimated coefficient for a particular independent variable is 0.5, it means that for every one-unit change in that independent variable, the dependent variable is expected to change by 0.5 units, while holding all other independent variables constant. This coefficient can be used to generate a predictive model that can be used to forecast the value of the dependent variable for new observations. Additionally, the estimated coefficients can be used to evaluate the relative importance of each independent variable in explaining the variation in the dependent variable, allowing us to identify which variables are most influential in shaping the outcome. The coefficient of a valid model could also be calculated using a minimum number of experimental solubility data which is a valuable strategy in early stages of drug discovery/development where small amount of the drug-candidate is available and lots of tests should be done.

By replacing the estimated coefficients in the general form of the studied equation, a predictive model can be created that can be used to estimate the value of the dependent variable for new observations. This model can be represented as a mathematical equation, where each independent variable is multiplied by its corresponding coefficient and added together to

produce a predicted value for the dependent variable.

**Conclusion**

The attached video provides a step-by-step guide to the entire explained process, from data preparation to model interpretation. The video proceeds to demonstrate the use of SPSS software, to perform the regression analysis, including the estimation of coefficients and the calculation of various diagnostic statistics. The video also provides a detailed explanation of how to interpret the results, including the meaning of the estimated coefficients and how to use them to make predictions. By following along with the video, viewers can gain a comprehensive understanding of the regression analysis process and how to apply it to their own solubility data.

**Funding**

This research was supported by the Pharmaceutical Analysis Research Center, Tabriz University of Medical Sciences, Tabriz, Iran under grant number 76902.

**Authors' Contribution**

**Conceptualization:** Abolghasem Jouyban

**Investigation:** Elaheh Rahimpour, Nima Nazemieh

**Supervision:** Abolghasem Jouyban

**Validation:** William E. Acree, Jr., Tahir Suleymanov

**Visualization:** William E. Acree, Jr., Tahir Suleymanov

**Writing–original draft:** Elaheh Rahimpour, Nima Nazemieh

**Writing–review & editing:** Abolghasem Jouyban, William E. Acree, Jr., Tahir Suleymanov

**Data availability**

Not applicable.

**Supplementary information**

The supplementary information containing a video clip demonstrates details of the computations

using SPSS available at https://ps.tbzmed.ac.ir/Article/ps-43440. The video in other languages are available in….

**Declarations**

**Ethical approval**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interest**

The author declares no conflict of interest.

**References**


1. Bear J. Conceptual and mathematical modeling. Seawater intrusion in coastal aquifers—concepts, methods and practices: Springer; 1999;127-61.
2. Moody DL, Shanks GG. Improving the quality of data models: empirical validation of a quality management framework. Inf Syst. 2003;28:619-50. doi: 10.1016/S0306-4379(02)00043-1
3. Jouyban A, Acree Jr. WE. Michael h. Abraham and his developed parameters: Various applications in medicine, chemistry and biology. Pharm Sci. 2022;28(2):170-3. doi: 10.34172/PS.2022.1
4. Jouyban A, Penã MÁP. Pilar bustamante and her advances in thermodynamic and solubility parameters. Pharm Sci. 2025;31(2):113-6. doi: 10.34172/PS.025.42150
5. Kaur H, Nori H, Jenkins S, Caruana R, Wallach H, Wortman Vaughan J, editors. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. Proceedings of the 2020 CHI conference on human factors in computing systems; 2020.
6. Wess R, Klock H, Siller H-S, Greefrath G, Wess R, Klock H, et al. Mathematical modelling. Measuring Professional competence for the teaching of mathematical modelling: A test instrument, Springer 2021;3-20. doi: 10.1007/978-3-030-78071-5
7. Simsion G, Witt G. Data modeling essentials: Elsevier; 2004.
8. Jafari P, Jouyban A. Prediction of drugs solubility in mono-solvent systems at different temperatures using a single determination. Pharm Sci. 2025;31(2):187-94. doi:

10.34172/PS.024.40474
9. Jouyban A. Solubility prediction of drugs in mono-solvents at various temperatures: Ciprofloxacin data simulation. Pharm Sci. 2024;30(3):391-7. doi: 10.34172/PS.2024.6
10. Mabhoot A, Jouyban A. Solubility of sodium phenytoin in propylene glycol + water mixtures in the presence of β-cyclodextrin. Pharm Sci. 2015;21(3):152-6. doi: 10.15171/PS.2015.30
11. Tinjacá DA, Martinez F, Almanza OA, Jouyban A, Acree WE. Solubility, correlation, dissolution thermodynamics and preferential solvation of meloxicam in aqueous mixtures of 2-propanol. Pharm Sci. 2022;28(1):130-44. doi: 10.34172/PS.2021.39
12. Rochefort-Maranda G. Simplicity and model selection. Eur J Philos Sci. 2016;6:261-79. doi: 10.1007/s13194-016-0137-1
13. Khan ME, Khan F. A comparative study of white box, black box and grey box testing techniques. IJACSA 2012; 3:12-5. doi: 10.14569/IJACSA.2012.0306038.
14. Jagdale S, Agarwal B, Karekar S, Rathi P, Jouyban-Gharamaleki V, Makarov DM, et al. Solubility, thermodynamic properties and predictive modeling of glimepiride in various solvents. J Mol Liq. 2025;431:127701. doi: 10.1016/j.molliq.2025.127701.
15. Jouyban A. Comments on “Artificial intelligence aided pharmaceutical engineering: Development of hybrid machine learning models for prediction of nanomedicine solubility in supercritical solvent”. J Mol Liq. 2025;423:126979. doi: 10.1016/j.molliq.2025.126979.
16. Jouyban A, Makarov DM, Acree WE. Some comments on the solubility of rivaroxaban and its numerical analysis methods using various algorithms. Phys Chem Liq. 2025. doi: 10.1080/00319104.2025.2523763.
17. Khayyati R, Almasi A, Eblaghi A, Naghavi Kalajahi SM, Ramezani M, Makarov DM, et al. Solubility of anti-epileptic drugs in mono-solvents at various temperatures: a data-driven simulation approach. Physics and Chemistry of Liquids 2025;63:874-83. doi: 10.1080/00319104.2025.2553893.
18. Makarov DM, Kalikin NN, Budkov YA, Gurikov P, Kruchinin SE, Jouyban A, et al. Improved Solubility Predictions in scCO2 Using Thermodynamics-Informed Machine Learning Models. J Chem Inf.Model. 2025;65:4043-56. doi: 10.1021/acs.jcim.5c00432.
19. Cui Q, Lu S, Ni B, Zeng X, Tan Y, Chen YD, et al. Improved prediction of aqueous solubility of novel compounds by going deeper with deep learning. Front oncol. 2020;10:121. doi: 10.3389/fonc.2020.00121
20. Karpov P, Godin G, Tetko IV. Transformer-CNN: Swiss knife for QSAR modeling and interpretation. J Cheminform 2020;12:17. doi: 10.1186/s13321-020-00423-w
21. Hou Y, Wang S, Bai B, Chan HS, Yuan S. Accurate physical property predictions via deep learning. Molecules 2022;27:1668. doi: 10.3390/molecules27051668
22. Nguyen HC, Alamray F, Kamal M, Diana T, Mohamed A, Algarni M, et al. Computational prediction of drug solubility in supercritical carbon dioxide: Thermodynamic and artificial intelligence modeling. J Mol Liq. 2022;354:118888. doi: 10.1016/j.molliq.2022.118888
23. Mac Fhionnlaoich N, Zeglinski J, Simon M, Wood B, Davin S, Glennon B. A hybrid approach to aqueous solubility prediction using COSMO-RS and machine learning. Chem Eng Res Des. 2024;209:67-71. doi: 10.1016/j.cherd.2024.07.050

24. Chen J-H, Tseng YJ. Different molecular enumeration influences in deep learning: an example using aqueous solubility. Brief Bioinform 2021;22:bbaa092. doi: 10.1093/bib/bbaa092
25. Žuvela P, Spataru A, Kern R, Wong MW. Machine learning in prediction of intrinsic aqueous solubility of drug-like compounds: Generalization, complexity, or predictive ability? J Chemom. 2021;35:e3349. doi: 10.1002/cem.3349
26. Conn JG, Carter JW, Conn JJ, Subramanian V, Baxter A, Engkvist O, et al. Blinded predictions and post hoc analysis of the second solubility challenge data: exploring training data and feature set selection for machine and deep learning models. J Chem Inf Model. 2023;63:1099-113. doi: 10.1021/acs.jcim.2c01189
27. Bao Z, Tom G, Cheng A, Watchorn J, Aspuru-Guzik A, Allen C. Towards the prediction of drug solubility in binary solvent mixtures at various temperatures using machine learning. J Chemom. 2024;16:117. doi: 10.1186/s13321-024-00911-3
28. Boobier S, Hose DR, Blacker AJ, Nguyen BN. Machine learning with physicochemical relationships: solubility prediction in organic solvents and water. Nat Commun 2020;11:5753. doi: 10.1038/s41467-020-19594-z
29. Sagnella G. Model fitting, parameter estimation, linear and non-linear regression. Trends Biochem Sci 1985;10:100-3. doi: 10.1016/0968-0004(85)90261-0
30. Fischer HP. Mathematical modeling of complex biological systems: From parts lists to understanding systems behavior. Alcohol Res Health. 2008;31:49-59. doi: 10.1007/978-3-8349-9870-5_4
31. Montesinos López OA, Montesinos López A, Crossa J. Overfitting, model tuning, and evaluation of prediction performance. Multivariate statistical machine learning methods for genomic prediction: Springer; 2022. p. 109-139.
32. Martínez F, Jouyban A, Acree Jr WE. Pharmaceuticals solubility is still nowadays widely studied everywhere. Pharm. Sci. 2017;23:1-2. doi: 10.15171/PS.2017.01
33. Pouton CW. Formulation of poorly water-soluble drugs for oral administration: physicochemical and physiological issues and the lipid formulation classification system. European J Pharm Sci. 2006;29:278-87. doi: 10.1016/j.ejps.2006.04.01628. Beg S, Swain S, Rizwan M, Irfanuddin M, Shobha Malini D. Bioavailability enhancement strategies: basics, formulation approaches and regulatory considerations. Curr. Drug Deliv. 2011;8:691-702. doi: 10.2174/156720111797635504
34. Zarghampour A, Jouyban K, Jouyban-Gharamaleki V, Jouyban A. Rahimpour E. A description on the shake-flask and laser monitoring-based techniques for determination of the drug's solubility. Pharm Sci. 2024; 30(2): 274-8. doi: 10.34172/PS.2024.2
35. Jouyban-Gharamaleki V, Jouyban A, Acree WE, Rahimpour E. Smart systems for determination of drug's solubility. Drug Dev Ind Pharm. 2019;45:177-87. doi: 10.1080/03639045.2018.1529786
36. Jouyban A, Chew NYK, Chan H-K, Sabour M, Acree Jr WE. A unified cosolvency model for calculating solute solubility in mixed solvents. Chem Pharm Bull. 2005; 53: 634-7. doi: 10.1248/cpb.53.634
37. Jouyban A. Review of the cosolvency models for predicting solubility of drugs in water-cosolvent mixtures. J Pharm Pharm Sci. 2008;11:32-58. doi: 10.18433/j3pp4k
38. Grant D, Mehdizadeh M, Chow A-L, Fairbrother J. Non-linear van't Hoff solubility-

temperature plots and their pharmaceutical interpretation. Int J Pharm. 1984;18:25-38. doi: 10.1016/0378-5173(84)90104-2

39. Acree Jr WE. Mathematical representation of thermodynamic properties: Part 2. Derivation of the combined nearly ideal binary solvent (NIBS)/Redlich-Kister mathematical representation from a two-body and three-body interactional mixing model. Thermochim Acta. 1992;198:71-9. doi: 10.1016/0040-6031(92)85059-5
40. Jouyban A, Acree Jr WE. Mathematical derivation of the Jouyban-Acree model to represent solute solubility data in mixed solvents at various temperatures. J Mol Liq. 2018;256:541-7. doi: 10.1016/j.molliq.2018.01.171
41. Jouyban-Gharamaleki A. The modified Wilson model and predicting drug solubility in water-cosolvent mixtures. Chem Pharm Bull. 1998;46:1058-61. doi: 10.1248/cpb.46.1058
42. Buchowski H, Khiat A. Solubility of solids in liquids: one-parameter solubility equation. Fluid Ph Equilib. 1986;25:273-8. doi: 10.1016/0378-3812(86)80003-6
43. Ahmadi M, Rezaei H, Zhao H, Rahimpour E, Jouyban A. Thermodynamic study of deferiprone dissolution in polyethylene glycol 400 and 2-propanol mixture. IJCCE 2023;42:2973-81. doi: 10.30492/ijcce.2023.1974301.5736.
44. Jouyban-Gharamaleki A, Valaee L, Barzegar-Jalali M, Clark B, Acree Jr. W. Comparison of various cosolvency models for calculating solute solubility in water–cosolvent mixtures. Int J Pharm. 1999;177:93-101. doi: https://doi.org/10.1016/s0378-5173(98)00333-0 .