Rev. Colomb. Cienc. Quím. Farm., **54**(3), 804–820 (2025)

Scientific research article

Evaluation of AI models for predicting drug solubility in cosolvent systems

Mateo Vergara-Roa[1], Rogelio Manuel Alvarado-Martinez[2], Angelica Chappe-Chappe[2], Rossember Edén Cardenas-Torres[3], Cristian Rincón-Guio[4], María Paz Saavedra-Saavedra[5], Carlos Francisco Trujillo-Trujillo[5] & Julian David Aristizabal-Yusty[5]*

[1] Ingeniería de Sistemas, Universidad Cooperativa de Colombia Villavicencio, Colombia
[2] Grupo de Investigación: Ciencias e Ingeniería para las Tecnologías de la Información y las Comunicaciones - Bogotá, Escuela de Ciencias Básicas, Facultad de Ingeniería, Diseño e Innovación, Politécnico Grancolombiano, Sede Bogotá, Calle 57 No. 3 - 00 este, Bogotá 110231, Cundinamarca, Colombia
[3] Grupo de Energía Materiales y Diseño EnerDIMAT, Facultad de Ingeniería, Universidad de América, Av. Circunvalar No. 20-53, Bogotá 110321, Cundinamarca, Colombia
[4] Escuela OPINA, Institución Universitaria Politécnico Grancolombiano, Sede Bogotá: calle 57 3 - 00 Este, Bloque K, Bogotá D. C., Cundinamarca, Colombia
[5] Grupo de Investigación de Ingenierías UCC-Neiva, Facultad de Ingeniería, Universidad Cooperativa de Colombia, Sede Neiva, Calle 11 No. 1-51, Neiva 410001, Huila, Colombia

*Correspondent author E-mail: Julian.Aristizabal@campusucc.edu.co

Received: August 8, 2025
Corrected: August 20, 2025
Accepted: August 22, 2025

https://doi.org/10.15446/rcciquifa.v54n3.122716

SUMMARY

The low solubility of drugs presents a significant challenge in pharmaceutical development. This study evaluates the ability of five artificial intelligence models — Artificial Neural Network (ANN), Random Forest, K-Nearest Neighbors (KNN), Linear Regression, and XGBoost — to predict drug solubility in cosolvent systems. The study uses a dataset that integrates the physicochemical properties of the solute and solvent. The objective was to identify the most accurate model and analyze the relative importance of the predictor variables. The results demonstrate a clear hierarchy in model performance. Tree-based algorithms XGBoost and Random Forest showed exceptional predictive power with coefficients of determination (R^2) of 0.9737 and 0.9752, respectively, and the lowest mean squared errors. The ANN model also demonstrated robust performance (R^2 = 0.9585), whereas the linear regression model was unable to capture the nonlinear relationships inherent in the system (R^2 = 0.4463). Feature importance analysis revealed significant differences: tree models prioritized the drug's molar mass, while ANN attributed considerable importance to the cosolvent system's properties, such as the solvents' molar mass and volume. These differences highlight the complexity of solubility as a multifactorial phenomenon. We conclude that ensemble models, particularly XGBoost, are highly effective and accurate tools for *in silico* solubility prediction. Additionally, we find that a comparative modeling approach is crucial for gaining a holistic understanding of the physicochemical interactions that govern the process.

Keywords: Solubility; drugs; artificial intelligence (AI); predictive models; cosolvent systems; pharmacy.

RESUMEN

Evaluación de modelos de IA para predecir la solubilidad de fármacos en sistemas cosolventes

La baja solubilidad de los fármacos supone un reto importante en el desarrollo farmacéutico. En este estudio se evalúa la capacidad de cinco modelos de inteligencia artificial (red neuronal artificial, bosque aleatorio, K-vecinos más cercanos, regresión lineal y XGBoost) para predecir la solubilidad de los fármacos en sistemas cosolventes. El estudio utiliza un conjunto de datos que integra las propiedades fisicoquímicas del soluto y el solvente. El objetivo era identificar el modelo más preciso y analizar la importancia relativa de las variables predictivas. Los resultados muestran una clara jerarquía en el rendimiento de los modelos. Los algoritmos basados en árboles, XGBoost y Bosque Aleatorio, mostraron una capacidad predictiva excepcional, con coeficientes de determinación (R^2) de 0,9737 y 0,9752, respectivamente, y los errores cuadrados medios más bajos. El modelo ANN también demostró un buen rendimiento (R^2 = 0,9585), mientras que el modelo de regresión lineal no pudo captar las relaciones no lineales inherentes al sistema (R^2 = 0,4463). El análisis de importancia de las características reveló diferencias significativas: los modelos de árbol dieron prioridad a la masa molar del fármaco, mientras que la ANN atribuyó una importancia considerable a las propiedades del sistema cosolvente, como la masa molar y el volumen de los disolventes. Estas diferencias ponen de manifiesto la complejidad de la solubilidad como fenómeno multifactorial. En conclusión, los modelos de conjunto, en particular XGBoost, son herramientas muy eficaces y precisas para la predicción de la solubilidad *in silico.* Además, consideramos que un enfoque de modelización comparativa es fundamental para obtener una comprensión global de las interacciones fisicoquímicas que rigen el proceso.

Palabras clave: Solubilidad; medicamentos; inteligencia artificial (IA); modelos predictivos; sistemas cosolventes; farmacia.

RESUMO

Avaliação de modelos de IA para prever a solubilidade de medicamentos em sistemas de co-solventes

A baixa solubilidade dos fármacos representa um desafio importante no desenvolvimento farmacêutico. Este estudo avalia a capacidade de cinco modelos de inteligência artificial (rede neural artificial, floresta aleatória, K-vizinhos mais próximos, regressão linear e XGBoost) para prever a solubilidade dos fármacos em sistemas co-solventes. O estudo utiliza um conjunto de dados que integra as propriedades físico-químicas do soluto e do solvente. O objetivo era identificar o modelo mais preciso e analisar a importância relativa das variáveis preditivas. Os resultados mostram uma hierarquia clara no desempenho dos modelos. Os algoritmos baseados em árvores, XGBoost e Random Forest, mostraram uma capacidade preditiva excepcional, com coeficientes de determinação (R^2) de 0,9737 e 0,9752, respectivamente, e os menores erros quadráticos médios. O modelo ANN também demonstrou um bom desempenho (R^2 = 0,9585), enquanto o modelo de regressão linear não conseguiu captar as relações não lineares inerentes ao sistema (R^2 = 0,4463). A análise da importância das características revelou diferenças significativas: os modelos de árvore deram prioridade à massa molar do fármaco, enquanto a ANN atribuiu uma importância considerável às propriedades do sistema co-solvente, como a massa molar e o volume dos solventes. Estas diferenças evidenciam a complexidade da solubilidade como um fenómeno multifatorial. Em conclusão, os modelos de conjunto, em particular o XGBoost, são ferramentas muito eficazes e precisas para a previsão da solubilidade *in silico*. Além disso, consideramos que uma abordagem de modelagem comparativa é fundamental para obter uma compreensão global das interações físico-químicas que regem o processo.

Palavras-chave: Solubilidade; medicamentos; inteligência artificial (IA); modelos preditivos; sistemas co-solventes; farmácia.

1. INTRODUCTION

The prevalence of low solubility in pharmaceutical research and development (R&D) portfolios has reached alarming proportions. While previous studies indicated that more than 40% of new chemical entities (NCEs) were practically insoluble in water, more recent estimates place this figure between 70% and 90% of all compounds in development [1, 2]. This phenomenon is not random, but rather the result of a paradigm shift in drug discovery strategies. Modern tools such as high-throughput screening (HTS) and target-oriented drug design have revolutionized our ability to identify potent and selective compounds. However, this success has had an unintended consequence [3, 4].

These methodologies tend to select molecules that are structurally larger, more complex, and—crucially—more lipophilic (fat-loving) to achieve optimal binding to their biological targets [5]. There is a well-established inverse correlation between lipophilicity and aqueous solubility. Therefore, as discovery tools become more efficient at identifying potent ligands, they systematically generate a higher proportion of candidates with unfavorable physicochemical properties. This creates a self-perpetuating solubility crisis where success in the early stages of discovery fuels challenges in later stages of formulation. In addition to the drug's intrinsic chemical structure, other factors, such as its pKa (which determines its state of ionization based on the pH of the medium) and solid state (polymorphism), also significantly influence solubility [6]. Crystalline drug forms are generally less soluble than their amorphous counterparts, adding another layer of complexity to the formulation challenge [7].

Traditionally, determining solubility has been a laborious experimental process. The reference method, known as the "shake-flask" method, involves shaking an excess of the compound in a solvent until equilibrium is reached [8]. This process can take days and requires significant amounts of compound, solvents, and human resources. While this methodology is accurate, it is incompatible with the pace and scale of modern drug discovery, where thousands of compounds must be evaluated in the early stages [8-10].

The economic imperative to find alternatives is overwhelming. Bringing a new drug to market costs an estimated $1 billion, with more than 30% of that cost devoted to formulation [11]. Failure of a candidate in the advanced stages of development due to insurmountable formulation problems represents massive financial loss. *In silico* (computational) methods emerge as a strategic solution here [12, 13]. These approaches can predict a compound's solubility based solely on its molecular structure, eliminating the need for physical synthesis [12].

The shift toward silico prediction is a fundamental change in development strategy. Rather than taking a reactive approach where formulation scientists attempt to "rescue" a promising but poorly soluble compound in the late stages, this strategy is proactive and design-based. Computational tools enable solubility to be considered a key optimization parameter from the earliest stages of discovery, alongside potency and selectivity. Chemists can screen virtual libraries of millions of compounds to identify those with a favorable balance of activity and physicochemical properties. This approach dramatically accelerates the R&D cycle, reduces costs, decreases reliance on animal models, and increases the overall likelihood of clinical success. The goal is not to replace experimentation entirely, but rather to guide it, enabling laboratory resources to focus on the most promising drug candidates [13].

2. METHODOLOGY

The research design aimed to identify the machine learning model with the greatest predictive capacity for the solubility of certain drugs in pure solvents and solvent mixtures of industrial interest. To ensure reproducibility and prevent information leakage, the entire workflow was implemented, from data preprocessing to final validation.

Each implemented model is based on a distinct learning principle, enabling a comparative assessment of their ability to capture the relationships between drug properties, solvent characteristics, and operating conditions. **Linear Regression** was employed as the baseline model due to its simplicity and interpretability, assuming a linear relationship between variables; its primary role is to establish a minimum performance benchmark against which more complex methods can be evaluated. The **K-Nearest Neighbors (KNN)** algorithm, in contrast, bases its predictions on instance similarity, estimating the solubility of a compound from the values of its closest neighbors in the feature space, making it particularly useful for capturing local patterns.

Decision tree–based models, such as **Random Forest** and **XGBoost**, allow for the identification of complex and non-linear interactions between variables. Random Forest constructs multiple trees trained on random subsets of both the data and the features, averaging their outputs to reduce variance and enhance stability. XGBoost, on the other hand, optimizes sequentially by correcting the residual errors of previous predictions, incorporating regularization to improve generalization. Finally, a **Multilayer Perceptron (MLP) Artificial Neural Network** was evaluated, capable of approximating highly non-linear functions through interconnected hidden layers of neurons, making it well-suited for modeling complex physicochemical phenomena.

In this context, it is important to remember that the methodology for developing an artificial intelligence model capable of predicting drug solubility is based on a supervised machine learning workflow. Initially, a database was consolidated and curated that integrates some specified predictor variables, from the intrinsic properties of the drug such as its Molar Mass, Solubility Parameter, Melting Temperature (K), and Melting Enthalpy (kJ/mol), to the characteristics of the co-solvent system, including the properties of Solvents 1 and 2 and their respective molar fractions. This dataset underwent rigorous preprocessing, which included: 1) encoding categorical variables (solvent names) into a numerical format using techniques such as one-hot encoding; and 2) standardizing or normalizing all numerical features so that their different scales would not negatively affect the model's performance.

2.1. Data preprocessing and segmentation

The initial dataset (Tables 1 and 2) was preprocessed to adapt the variables to the modeling algorithms. Categorical variables were transformed using one-hot encoding, and numerical variables were standardized using *z*-scores to mitigate scale bias.

Then, the preprocessed dataset was stratified into training (80%) and testing (20%) subsets. A fixed random seed (random_state=42) was used to ensure reproducibility of the partition. We encapsulated the entire preprocessing and modeling flow in a Scikit-learn Pipeline structure to ensure that data transformers were adjusted exclusively on the training set.

Table 1. Physicochemical properties of some drugs[a,b]

Drugs	Molar mass	Solubility parameter	Molar volume	Melting temperature	Melting enthalpy	Solubility ref.
Hydroxytyrosol	154.16	37.05	131.2	333.55	94	[14, 15]
Indomethacin	357.79	24.5	230	433	55.1	[16-23]
Triclocarban	315.58	26.5	177.6	527.54	41.5	[24-32]
Progesterone	314.46	19.81	304.7	403.4	26.9	[33]
Meloxicam	351.4	30.95	189.1	536.7	43.9	[34-38]
Amygdalin	457.43	29.9	377.3	499.56	No disponible	[39]
Acetaminophen	151.16	28.2	111.2	441.2	27.6	[40-41]
Naproxen	230.26	23.4	189.4	428.8	34.2	[22, 42-44]
Methocarbamol	241.24	25.64	171.9	369.8	40.06	[45-47]
Benzoin	212.24	33.07	142.2	406	35.85	[48]
Sulfadiazine	250.28	28.89	150	532.65	44.3	[49-59]
Sulfamerazine	264.3	28.1	164.5	508.45	41.3	[49, 50, 60-66]
Sulfamethazine	278.33	27.42	179	468.95	39.2	[49, 50, 59, 62, 67-71]
Sulfacetamide	214.24	28.4	134.1	455.2	29.79008	[72]
Sulfapyridine	249.29	27.1	158.5	462.7	40.47	[73-75]
Sulfamethizole	270.33	23.65	172.3	480.8	46.4	[76-78]
Sulfanilamide	172.2	24.81	128.7	435.4	23.30488	[77, 78]
Methylparaben	152.15	28.1	110.5	399.2	25.3	[79]
Ethylparaben	166.17	26.97	126.5	389.2	27.9	[80, 81]
Propylparaben	180.2	26.1	142.5	369.25	31.3	[82]
Nimodipine	418.44	22.9	309.2	397.2	41.7	[83]

[a]See Delgado *et al.* [84]; [b]Units: Molar mass in g/mol; Solubility parameter in $MPa^{1/2}$, Molar volume in cm^3/mol; Melting temperature in K; Melting enthalpy in kJ/mol.

2.2. Model training and comparative evaluation:

We trained and evaluated a set of five regression algorithms [85], which were selected to cover different modeling approaches: artificial neural network (ANN), a random forest, a K-nearest neighbors (KNN) model, a linear regression model, and an XGBoost model.

The performance of each model was quantified using a five-fold cross-validation protocol applied only to the training set. The evaluation was based on three standard regression task metrics: the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). During this stage, we optimized the hyperparameter k for the KNN model and identified k = 11 as the value offering the optimal balance between bias and variance.

Table 2. Physicochemical properties of some solvents of interest to the pharmaceutical industry

Solvents	Molar mass (g/mol)	Solubility parameter ($MPa^{1/2}$)	Molar volume (cm^3/mol)	Solubility Ref.
Water	18.01528	47.9	18	[86, 87]
Methanol	32.04	29.6	41	
Ethanol	46.068	26.6	59	
1-Propanol	60.0952	24.5	75	
1-octanol	130.23	20.9	158	
Acetonitrile	41.05	24.8	53	
1,4-Dioxane	88.11	23.2	86	
Ethylene glycol	62.07	34.9	56	
Cyclohexane	84.16	16.8	108	
Propylene glycol	76.09	30.7	73.4	
PEG 400	400	21.6	354.45	
PEG 200	200	24.3	177.65	
PEG 300	300	22.5	265.98	
Ethyl acetate	88.11	18.2	98	
N-methyl-2-pyrrolidone	99.13	23	96.24	

3. RESULTS AND DISCUSSION

The solubility of a solid drug in a solvent mixture is fundamentally governed by the balance between the energy required to break down the crystalline lattice of the solute and the energy released by the solvation of the drug molecules [6]. In this context, the melting temperature (K) and melting enthalpy (kJ/mol) directly reflect the stability of the crystal lattice [7]. A high melting temperature and enthalpy indicate an energetically stable crystal, which hinders dissolution and decreases solubility at a given temperature. This dissolution process is only favorable if the drug-solvent interactions offset the energy penalty. Solubility parameters (δ), based on Hildebrand-Scatchard theory, are crucial here because affinity is maximized and solubility is enhanced when the drug solubility parameter is very close to that of the solvent medium [87, 88]. In a cosolvent system consisting of Solvent 1 and Solvent 2, the parameters are not evaluated in isolation. Instead, an effective solubility parameter for the mixture is calculated [89]. This parameter is a weighted function of the Solvent 1 Solubility Parameter and the Solvent 2 Solubility Parameter [90, 91]. It is adjusted according to the molar fractions of Solvent 1 and Solvent 2. In this way, by manipulating the molar fractions, we can "tune" the solubility parameter of the mixture to optimally match that of the drug, thereby maximizing solubility. Finally, the molar volume of the drug and the molar volumes of solvents 1 and 2, together with their respective molar masses, are fundamental properties that, although sometimes omitted in simple approximations, are indispensable for rigorous calculations of the free energy of mixing, as they influence the entropic and energetic contributions of the process, allowing for a precise quantitative description of the system at the molar level, which is the foundation of all these thermodynamic relationships [6].

Figure 1 provides a detailed analysis of the five machine learning models that were evaluated to predict the solubility of certain drugs compared to the experimental solubility values. The solid line represents the identity line, which indicates a perfect prediction where the calculated solubility (x_3^{Cal}) equals the experimental solubility (x_3^{Exp}). The dotted line shows the trend of the model's predictions. Five models were analyzed: an artificial neural network

(ANN), a random forest, a K-nearest neighbors (KNN) model, a linear regression model, and an XGBoost model.

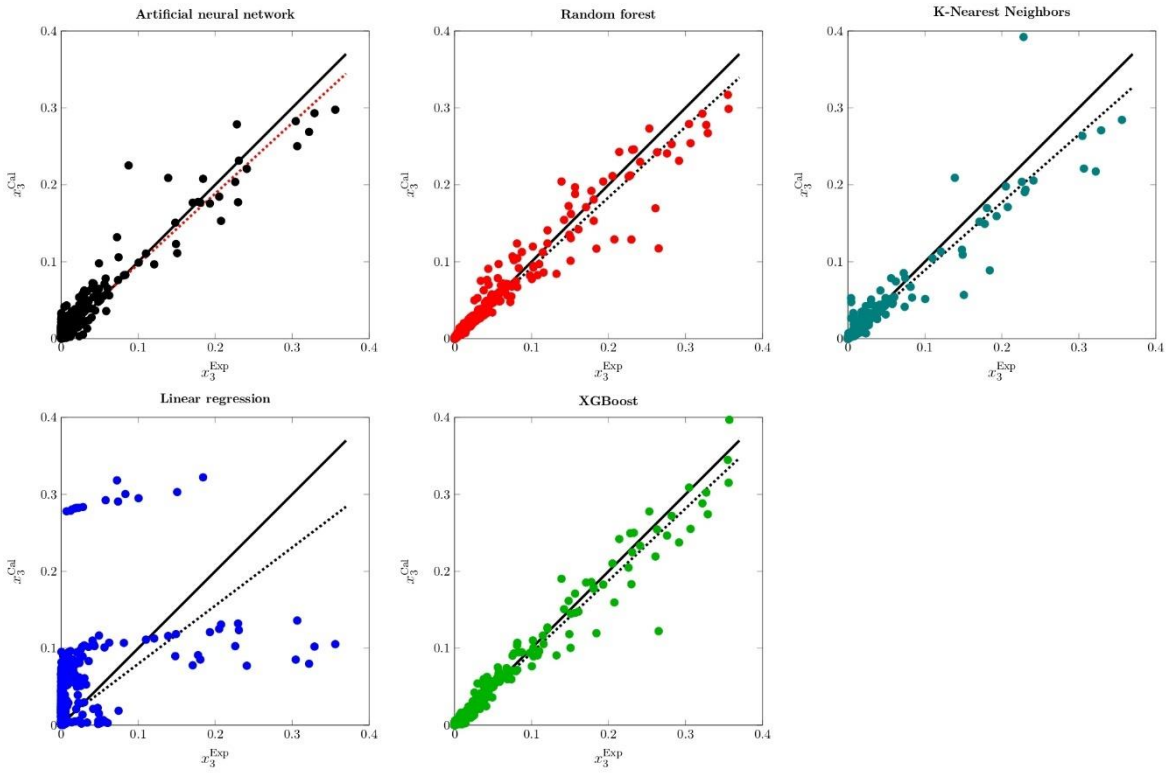

Figure 1. Predictions from different machine learning models for the calculated solubility variable vs. experimental solubility.

In relation to the first model, the artificial neural network demonstrates high performance. As shown in the graph, there is a tight grouping of points around the identity line, indicating high accuracy and an almost perfect correspondence between the predicted and experimental values. This is supported by the performance metrics, which include a coefficient of determination (R^2) of 0.9585 and a mean squared error (MSE) of 0.00012522. These values suggest that the model explains 95.85% of the variability in the experimental data, with very low residual error. The dotted regression line nearly overlaps with the identity line, indicating that the model is unbiased and highly effective for prediction. The Random Forest model also exhibits outstanding performance when evaluated. Its predictions cluster tightly around the identity line with minimal point dispersion. The metrics confirm this: an R^2 of 0.9752, surpassing even the ANN. The mean square error (MSE) is 0.00008593, the lowest of all the analyzed models, indicating exceptional accuracy and minimal prediction error. The proximity of the regression line to the identity line further supports the idea of a robust and highly reliable model. The K-Nearest Neighbors model shows moderate performance compared to tree and neural network models. The graph shows greater dispersion of points around the identity line, indicating lower prediction accuracy. This is confirmed by the metrics: an R^2 of 0.8996 and an MSE of 0.00030587. While the model still has acceptable explanatory power, the dispersion is significantly greater, particularly within the lower range of values. The dotted regression line follows the general trend but deviates more than it does in high-performance models, suggesting lower generalization ability. The Linear Regression model shows poor performance for this task. The graph shows wide dispersion and clear deviation of the regression line from the

identity line. The metrics support this observation: an R^2 of 0.4463 indicates that the model explains less than 45% of the variability in the data. The MSE is 0.003471, which is much higher than in the other models and implies substantial prediction error. The presence of outliers and clustering of points in the lower left of the graph suggest that the linear model fails to capture the underlying nonlinear relationship between the variables. Finally, the XGBoost model's performance is exceptional and comparable to that of the Random Forest model. The compact clustering of points around the identity line indicates high accuracy. This observation is confirmed by the metrics, which show an R^2 of 0.9737, like the R^2 of the Random Forest model. The MSE is 0.00008629, nearly identical to Random Forest's, which reinforces the conclusion that this model is accurate. The dotted regression line nearly overlaps with the identity line, demonstrating zero bias and optimal predictive power.

Upon comparing the five models, three clear performance categories emerged. XGBoost and Random Forest are the highest performing models. These models are based on decision trees and demonstrate a superior ability to capture complex, nonlinear relationships in the data. This ability is confirmed by their high R^2 values (0.9737 and 0.9752, respectively) and low MSEs. These results suggest that ensemble learning techniques are particularly effective for this dataset. The artificial neural network falls into the second category. Although it performs excellently with an R^2 of 0.9585, it is slightly inferior to boosted tree models. However, its ability to model nonlinear relationships makes it a solid option for prediction with an excellent fit. The K-Nearest Neighbors model is intermediate. Its proximity-based approach is less accurate than ensemble methods or neural networks for this problem, resulting in greater data dispersion. Finally, linear regression is the model with the lowest performance. Its inability to handle the data's inherent nonlinearity results in a poor fit and high error, rendering it unsuitable for predicting solubility. In summary, the most appropriate model depends on the priority given to maximum accuracy versus interpretability. Thus, XGBoost and Random Forest are the best options in terms of accuracy.

When analyzing the residual plots (Figure 2) for the five machine learning models, a clear performance hierarchy emerges. The XGBoost model shows the best performance, with residuals that are densely and homogeneously clustered around the zero line. This demonstrates high accuracy and ideal homoscedasticity. The Random Forest model also ranks highly, showing a consistent and random distribution of residuals across the entire prediction range, indicating robust performance. The artificial neural network (ANN) also performs well, though it shows slight heteroscedasticity for higher predicted values. In contrast, the K-Nearest Neighbors model shows moderate performance, with greater dispersion and more evident heteroscedasticity patterns, resulting in lower accuracy. Linear Regression is clearly the least effective model, with a pattern of residuals that deviates dramatically from the zero line. This reveals a systematic bias and an inability to capture the nonlinear nature of the data.

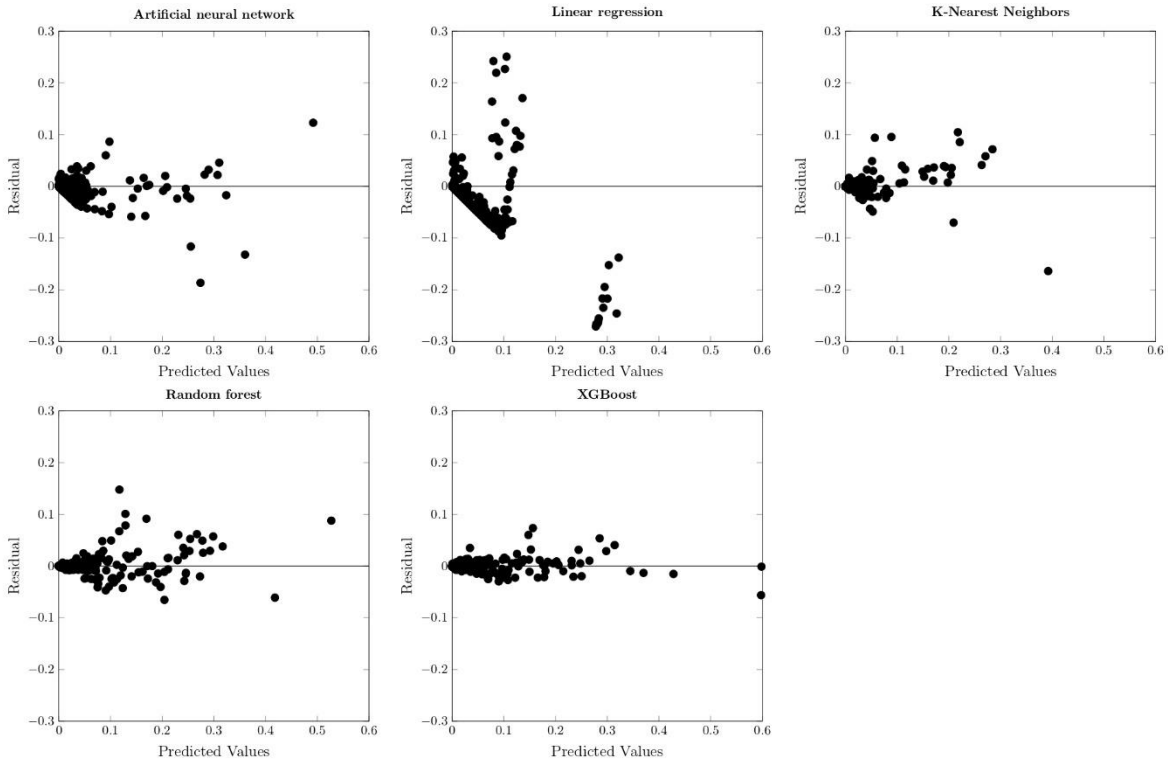

Figure 2. Analysis of residuals from machine learning models implemented in the prediction of solubility of certain drugs.

3.1. Comparison of the importance of predictor variables determined by permutation of AI models

Table 3 summarizes the importance values for each feature across the five analyzed models. The features are ranked according to their importance in the artificial neural network (ANN) model, which was found to be the most sensitive to all variables.

Important note on metrics: Each model uses a different metric to calculate importance.

a) ANN and KNN: Measure the average decrease in performance (R^2) per permutation. A higher value means that the feature is more important.
b) Linear Regression: It measures the absolute magnitude of the coefficient. A larger coefficient indicates greater influence on the prediction.
c) Random Forest and XGBoost: They use internal importance metrics, such as Gini impurity reduction or gain. A higher value means greater importance.

Table 3. The importance of predictor variables determined by permutation of AI models

Variable	ANN (R^2 Drop)	KNN (R^2 Drop)	Random Forest (Importance)	XGBoost (Importance)	Linear regression (Coefficient)
Molar mass solvent 1	~1.350	~0.490	0	0	~0.220
Molar volume solvent 1	~0.780	~0.040	0	0	~0.290
Molar volume of drug	~0.600	~0.015	0	0	~0.095
Mole fraction of most polar solvent (in cosolvent mixture)	~0.530	~0.140	~0.150	0	~0.008
Mole fraction of least polar solvent (in cosolvent mixture)	~0.330	~0.130	~0.160	~0.080	~0.007
Molar mass of drug	~0.300	~0.015	~0.670	~0.940	~0.025
Enthalpy of fusion (kJ/mol)	~0.280	~0.025	0	0	~0.012
Drug solubility parameter	~0.150	~0.005	~0.010	~0.010	~0.020
Solvent 1 solubility parameter	~0.140	~0.001	0	0	~0.035
Molar volume of solvent 2	~0.140	0	0	0	~0.100
Solvent 2 solubility parameter	~0.130	0	0	0	~0.040
Melting temperature of drug (K)	~0.120	~0.010	0	~0.002	~0.015
Molar mass of solvent 2	~0.080	0	0	0	~0.090
Study temperature	~0.040	~0.035	~0.020	~0.010	~0.005

A comparative analysis of the models reveals that solubility does not depend on a single dominant characteristic, but rather on a complex interaction of multiple factors. This conclusion is evident in the clear divergence between the approaches of each algorithm. Tree models, such as XGBoost and Random Forest, emphasize the properties of the solute and show a preference for drug molar mass, probably because it is a strong indicator of molecular size and intermolecular forces. However, other models paint a different picture. The artificial neural network (ANN), K-nearest neighbors (KNN), and linear regression models give significant weight to the properties of the solvent mixture, such as its molar masses and volumes. This finding has profound physicochemical significance since solvation (the solute-solvent interaction) is the true thermodynamic driver of dissolution. This comparison ultimately demonstrates the risk of relying on a single model. A limited analysis of XGBoost could mistakenly conclude that solvent properties are irrelevant. This is because both XGBoost and Random Forest show zero values for the importance of physicochemical parameters, such as solubility parameters. This is theoretically unlikely and highlights the limitations of these AI models in explaining solubility behavior. Therefore, contrasting different algorithms is crucial to obtaining a multifaceted view, which validates that both the drug and the solvent system are indispensable pieces in the complex puzzle of solubility.

4. CONCLUSION

The study's comparative analysis conclusively demonstrates that tree-based models are superior for predicting solubility. The XGBoost and Random Forest algorithms achieved the highest performance, with coefficients of determination (R^2) of 0.9737 and 0.9752, respectively. Linear Regression, on the other hand, proved inadequate for this complex problem. Its inability to capture the nonlinear relationships inherent in the solubility system resulted in poor performance ($R^2 = 0.4463$) and high prediction error. This disparity underscores the complexity of

solubility as a multifactorial phenomenon, a reality reinforced by the models' differing interpretations of variable importance. Significant differences in weighting were found: while the tree models prioritized the drug's molar mass, the artificial neural network (ANN) attributed considerable importance to the cosolvent system's properties, such as the solvents' molar mass and volume. For these reasons, the study concludes that model comparison is crucial for a comprehensive understanding. Using a comparative modeling approach is essential to gaining a holistic understanding of the physicochemical interactions that govern the process and avoiding biased or incomplete conclusions that could arise from relying on a single algorithm.

CONFLICTS OF INTEREST

All authors declare that there are no conflicts of interest.

REFERENCES

1. S. Stegemann, C. Moreton, S. Svanbäck, K. Box, G. Motte & A. Paudel. Trends in oral small-molecule drug discovery and product development based on product launches before and after the Rule of Five. *Drug Discov. Today*, **28**(2), 103344 (2023). Doi: https://doi.org/10.1016/j.drudis.2022.103344
2. B. Xie, Y. Liu, X. Li, P. Yang & W. He. Solubilization techniques used for poorly water-soluble drugs. *Acta Pharm. Sinica B*, **14**(11), 4683–4716 (2024). Doi: https://doi.org/10.1016/j.apsb.2024.08.027
3. K.T. Savjani, A.K. Gajjar & J.K. Savjani. Drug solubility: Importance and enhancement techniques. *ISRN Pharm.*, **2012**, 195727 (2012). Doi: https://doi.org/10.5402/2012/195727
4. M. Khoubnasabjafari, D.R. Delgado, F. Martinez, A. Jouyban & W.E. Acree. Predicting the solubility, thermodynamic properties and preferential solvation of sulphamethazine in acetonitrile + water mixtures using a minimum number of experimental data points. *Phys. Chem. Liq.*, **59**(3) 400–411 (2021). Doi: https://doi.org/10.1080/00319104.2020.1731812
5. G. Krishna, K.J. Chen, C.C. Lin & A.A. Nomeir. Permeability of lipophilic compounds in drug discovery using in-vitro human absorption model, Caco-2. *Int. J. Pharm.*, **222**(1), 77–89 (2001). Doi: https://doi.org/10.1016/S0378-5173(01)00698-6
6. Y. Pal, P.K. Deb, S. Bandopadhyay, N. Bandyopadhyay & R.K. Tekade. Role of physicochemical parameters on drug absorption and their implications in pharmaceutical product development. In: R.K. Tekade (editor). *Dosage Form Design Considerations: Volume I: A volume in Advances in Pharmaceutical Product Development and Research.* Academic Press, 2018; pp. 85–116. Doi: https://doi.org/10.1016/B978-0-12-814423-7.00003-4
7. S.B. Murdande, M.J. Pikal, R.M. Shanker & R.H. Bogner. Aqueous solubility of crystalline and amorphous drugs: Challenges in measurement. *Pharm. Dev. Technol.*, **16**(3), 187–200 (2011). Doi: https://doi.org/10.3109/10837451003774377
8. W.J. Mader & T. Higuchi. Phase solubility analysis. *C R C Crit. Rev. Anal. Chem.*, **1**(2), 193–215 (1970). Doi: https://doi.org/10.1080/10408347008542734
9. J.J. Lingane. Advances in Analytical Chemistry and Instrumentation. Volume I (Book Review). *J. Am. Chem. Soc.*, **83**(10), 2404–2404 (1961). Doi: https://doi.org/10.1021/ja01471a053
10. L.W. Dittert, T. Higuchi & D.R. Reese. Phase solubility technique in studying the formation of complex salts of triamterene. *J. Pharm. Sci.*, **53**(11), 1325–1328 (1964). Doi: https://doi.org/10.1002/jps.2600531108
11. Current model for financing drug development: From concept through approval. In: *Breakthrough Business Models: Drug Development for Rare and Neglected Diseases and Individualized Therapies: Workshop Summary.* Institute of Medicine (US) Forum on Drug Discovery, Development, and Translation. National Academies Press, Washington (DC), 2009; pp. 7–11. URL: https://www.ncbi.nlm.nih.gov/books/NBK50977/pdf/Bookshelf_NBK50977.pdf, accessed August 7, 2025.

12. L. Morici, O. Jordan, E. Allémann & C. Rodríguez-Nogales. Recent advances in nanocrystals for arthritis drug delivery. *Expert Opin. Drug Deliv.*, **22**(7), 1031–1042 (2025). Doi: https://doi.org/10.1080/17425247.2025.2505758
13. N.Y. Echeagaray-Solorza, Y. Díaz-Romero, L.G. Tobón-Galicia & S. Tejada-Paniagua. Predicting drug solubility in cosolvent systems using artificial intelligence algorithms. *Rev. Colomb. Cienc. Quím. Farm.*, **54**(1), 131–144 (2025). Doi: https://doi.org/10.15446/rcciquifa.v54n1.119553
14. D.R. Delgado, A. Sosnik & F. Martínez. Transfer thermodynamics of triclosan from water to organic solvents with different hydrogen bonding capability. *Lat. Am. J. Pharm.*, **30**(3), 459–466 (2011). URL: https://www.latamjpharm.org/resumenes/30/3/LAJOP_30_3_1_7.pdf
15. A. Aydi, I. Dali, K. Ghachem, A.Z. Al-Khazaal, D.R. Delgado & L. Kolsi. Solubility of Hydroxytyrosol in binary mixture of ethanol + water from (293.15 to 318.15) K: Measurement, correlation, dissolution thermodynamics and preferential solvation. *Alexandria Eng. J.*, **60**(1) 905–914 (2021). Doi: https://doi.org/10.1016/j.aej.2020.10.019
16. A.R. Holguín, G.A. Rodríguez, D.M. Cristancho, D.R. Delgado & F. Martínez. Solution thermodynamics of indomethacin in propylene glycol+water mixtures. *Fluid Phase Equilib.*, **314**, 134–139 (2012). Doi: https://doi.org/10.1016/j.fluid.2011.11.001
17. E.A. Cantillo, D.R. Delgado & F. Martinez, Solution thermodynamics of indomethacin in ethanol + propylene glycol mixtures. *J. Mol. Liq.*, **181**, 62–67 (2013). Doi: https://doi.org/10.1016/j.molliq.2013.02.008
18. M.Á. Peña, D.R. Delgado & F. Martínez. Preferential solvation of indomethacin in some aqueous co-solvent mixtures. *Chem. Eng. Commun.*, **203**(5), 619–627 (2016). Doi: https://doi.org/10.1080/00986445.2015.1074898
19. M.A. Ruidiaz, D.R. Delgado, F. Martínez & Y. Marcus. Solubility and preferential solvation of indomethacin in 1,4-dioxane+water solvent mixtures. *Fluid Phase Equilib.*, **299**(2), 259–265 (2010). Doi: https://doi.org/10.1016/j.fluid.2010.09.027
20. A.R. Holguín, D.R. Delgado & F. Martínez. Indomethacin solubility in propylene glycol + water mixtures according to the Extended Hildebrand Solubility Approach. *Lat. Am. J. Pharm.*, **31**(5), 720–726 (2012). URL: https://www.latamjpharm.org/resumenes/31/5/LAJOP_31_5_1_12.pdf
21. M.A. Ruidiaz, D.R. Delgado & F. Martínez. Indomethacin solubility estimation in 1,4-dioxane + water mixtures by the extended Hildebrand solubility approach. *Quim. Nova*, **34**(9), 1569–1574 (2011). Doi: https://doi.org/10.1590/S0100-40422011000900016
22. G.A. Rodríguez, D.R. Delgado & F. Martínez. Preferential solvation of indomethacin and naproxen in ethyl acetate + ethanol mixtures according to the IKBI method. *Phys. Chem. Liq.*, **52**(4), 533–545 (2014). Doi: https://doi.org/10.1080/00319104.2013.842474
23. M.Á. Peña, D.R. Delgado & F. Martínez. Preferential solvation of indomethacin in 1,4-dioxane + water mixtures according to the inverse Kirkwood–Buff integrals method. *Phys. Chem. Liq.*, 54(4) 462–474 (2016). Doi: https://doi.org/10.1080/00319104.2015.1115329
24. D.R. Delgado, A.R. Holguin & F. Martínez. Solution thermodynamics of triclosan and triclocarban in some volatile organic solvents. *Vitae*, **19**(1), 79–92 (2012). Doi: https://doi.org/10.17533/udea.vitae.10838
25. A.R. Holguín, D.R. Delgado & F. Martínez. Thermodynamic study of the solubility of triclocarban in ethanol + propylene glycol mixtures. *Quim. Nova*, **35**(2), 280–285 (2012). Doi: https://doi.org/10.1590/S0100-40422012000200009
26. D.I. Caviedes-Rubio, C.P. Ortiz, F. Martinez & D.R. Delgado. Thermodynamic assessment of triclocarban dissolution process in *N*-methyl-2-pyrrolidone + water cosolvent mixtures. *Molecules*, **28**(20), 7216 (2023). Doi: https://doi.org/10.3390/molecules28207216
27. A.M. Cruz-González, M.S. Vargas-Santana, S.d.J. Polania-Orozco, C.P. Ortiz, N.E. Cerquera, F. Martínez, D.R. Delgado, A. Jouyban & W.E. Acree, Jr. Thermodynamic analysis of the solubility of triclocarban in ethylene glycol + water mixtures. *J. Mol. Liq.*, **325**, 115222 (2021). Doi: https://doi.org/10.1016/j.molliq.2020.115222

28. J.J. Agredo-Collazos, C.P. Ortiz, N.E. Cerquera, R.E. Cardenas-Torres, D.R. Delgado, M.Á. Peña & F. Martínez. Equilibrium solubility of triclocarban in (cyclohexane + 1,4-dioxane) mixtures: Determination, correlation, thermodynamics and preferential solvation. *J. Solution Chem.*, **51**(12), 1603–1625 (2022). Doi: https://doi.org/10.1007/s10953-022-01209-4
29. V. Puentes-Lozada, D.I. Caviedes-Rubio, C. Rincón-Guio, N.E. Cerquera, R.E. Cardenas-Torres, C.P. Ortiz, F. Martinez & D.R. Delgado, Thermodynamic study of the solubility of triclocarban in polyethylene glycol 200 + water cosolvent mixtures at different temperatures. *Molecules*, **30**(12), 2631 (2025). Doi: https://doi.org/10.3390/molecules30122631
30. A.C. Gaviria-Castillo, J.D. Artunduaga-Tole, J.D. Rodríguez-Rubiano, J.A. Zuñiga-Andrade, D.R. Delgado, A. Jouyban & F. Martínez. Solution thermodynamics and preferential solvation of triclocarban in 1,4-dioxane (1) + water (2) mixtures at 298.15 K, *Phys. Chem. Liq.*, **57**(1), 55–66 (2019). Doi: https://doi.org/10.1080/00319104.2017.1416613
31. D.R. Delgado, E.M. Mogollon-Waltero, C.P. Ortiz, M.Á. Peña, O.A. Almanza, F. Martínez & A. Jouyban. Enthalpy-entropy compensation analysis of the triclocarban dissolution process in some 1,4-dioxane (1) + water (2) mixtures. *J. Mol. Liq.*, **271**, 522–529 (2018). Doi: https://doi.org/10.1016/j.molliq.2018.09.026
32. C.A. Munoz-Ortiz, N.E. Cerquera, J.K.C. Camacho, J. Osorio-Gallego, R.E. Cardenas-Torres, M. Herrera & D.R. Delgado. Preferential solvation of triclocarban in *N*-methyl-2-pyrrolidone + water cosolvent mixtures according to the Inverse Kirkwood-Buff Integrals (IKBI) method and correlation of solubility by means of some mathematical models. *Rev. Colomb. Cienc. Quim. Farm.*, **53**(1), 219–243 (2024). Doi: https://doi.org/10.15446/rcciquifa.v53n1.111422
33. A. Torres-Cardozo, N.E. Cerquera, C.P. Ortiz, J. Osorio-Gallego, R.E. Cardenas-Torres, F. Angarita-Reina, F. Martinez & D.R. Delgado. Thermodynamic analysis of the solubility of progesterone in 1-octanol + ethanol cosolvent mixtures at different temperatures. *Alexandria Eng. J.*, **64**, 219–235 (2023). Doi: https://doi.org/10.1016/j.aej.2022.08.035
34. D.R. Delgado, A.R. Holguín, O.A. Almanza, F. Martínez & Y. Marcus. Solubility and preferential solvation of meloxicam in ethanol + water mixtures. *Fluid Phase Equilib.*, **305**(1), 88–95 (2011). Doi: https://doi.org/10.1016/j.fluid.2011.03.012
35. D.M. Cristancho, D.R. Delgado & F. Martínez. Meloxicam solubility in ethanol + water mixtures according to the extended Hildebrand solubility approach. *J. Solution Chem.*, **42**(8), 1706–1716 (2013). Doi: https://doi.org/10.1007/s10953-013-0058-y
36. D.M. Jiménez, Z.J. Cárdenas, D.R. Delgado, A. Jouyban & F. Martínez. Solubility and solution thermodynamics of meloxicam in 1,4-dioxane and water mixtures. *Ind. Eng. Chem. Res.*, **53**(42), 16550–16558 (2014). Doi: https://doi.org/10.1021/ie503101h
37. A.R. Holguín, D.R. Delgado, F. Martínez & Y. Marcus. Solution thermodynamics and preferential solvation of meloxicam in propylene glycol + water mixtures. *J. Solution Chem.*, **40**(12), 1987–1999 (2011). Doi: https://doi.org/10.1007/s10953-011-9769-0
38. D.R. Delgado, A. Jouyban & F. Martínez. Solubility and preferential solvation of meloxicam in methanol + water mixtures at 298.15 K. *J. Mol. Liq.*, **197**, 368–373 (2014). Doi: https://doi.org/10.1016/j.molliq.2014.06.006
39. A. Aydi, C. Ayadi, K. Ghachem, A.Z. Al-Khazaal, D.R. Delgado, M. Alnaief & L. Kolsi. Solubility, solution thermodynamics, and preferential solvation of amygdalin in ethanol + water solvent mixtures. *Pharmaceuticals*, **13**(11), 395 (2020). Doi: https://doi.org/10.3390/ph13110395
40. E.A. Ahumada, D.R. Delgado & F. Martínez. Solution thermodynamics of acetaminophen in some PEG 400 + water mixtures. *Fluid Phase Equilib.*, **332**, 120–127 (2012). Doi: https://doi.org/10.1016/j.fluid.2012.07.004
41. E.A. Ahumada, D.R. Delgado & F. Martínez. Corrigendum to "Solution thermodynamics of acetaminophen in some PEG 400 + water mixtures" [Fluid Phase Equilibr. 332 (2012) 120-127]. *Fluid Phase Equilib.*, **334**, 204 (2012). Doi: https://doi.org/10.1016/j.fluid.2012.08.017
42. A.R. Holguín, D.R. Delgado, M.A. Ruidiaz, E.F. Vargas & F. Martínez. Apparent molar volumes of sodium naproxen in water at several concentrations and temperatures. *Lat. Am. J. Pharm.*, **30**(3), 619–623 (2011). URL: https://www.latamjpharm.org/resumenes/30/3/LAJOP_30_3_2_17.pdf

43. D.R. Delgado, M.A. Ruidiaz, S.M. Gómez, M. Gantiva & F. Martínez. Thermodynamic study of the solubility of sodium naproxen in some ethanol + water mixtures. *Quim. Nova*, **33**(9), 1923–1927 (2010). Doi: https://doi.org/10.1590/S0100-40422010000900019
44. G.A. Rodríguez, D.R. Delgado, F. Martínez, A. Jouyban & W.E. Acree, Jr. Solubility of naproxen in ethyl acetate + ethanol mixtures at several temperatures and correlation with the Jouyban-Acree model. *Fluid Phase Equilib.*, **320**, 49–55 (2012). Doi: https://doi.org/10.1016/j.fluid.2012.02.009
45. K.C. Mercado, G.A. Rodríguez, D.R. Delgado, F. Martínez & A. Romdhani. Solution thermodynamics of methocarbamol in some ethanol + water mixtures. *Quim. Nova*, **35**(10), 1967–1972 (2012). Doi: https://doi.org/10.1590/S0100-40422012001000015
46. D.M. Jiménez, Z.J. Cárdenas, D.R. Delgado, F. Martínez & A. Jouyban. Preferential solvation of methocarbamol in aqueous binary co-solvent mixtures at 298.15 K. *Phys. Chem. Liq.*, **52**(6), 726–737 (2014). Doi: https://doi.org/10.1080/00319104.2014.915755
47. Z.J. Cárdenas, D.M. Jiménez, G.A. Rodríguez, D.R. Delgado, F. Martínez, M. Khoubnasabjafari & A. Jouyban. Solubility of methocarbamol in some cosolvent + water mixtures at 298.15 K and correlation with the Jouyban-Acree model. *J. Mol. Liq.*, **188**, 162–166 (2013). Doi: https://doi.org/10.1016/j.molliq.2013.10.012
48. A.E. Garrido-Romero, J.L. Aroca-Trujillo, R. Rodriguez-Serrezuela, D.I. Caviedes-Rubio & D.R. Delgado. Thermodynamic study of the solubility of benzoin in ethyl acetate (1) + ethanol (2) mixtures. *J. Eng. Appl. Sci.*, **14**(14), 4951–4960 (2019). URL: https://account.makhillpublications.co/files/published-files/mak-jeas/2019/14-4951-4960.pdf
49. M.d.M. Muñoz, D.R. Delgado, M.Á. Peña, A. Jouyban & F. Martínez. Solubility and preferential solvation of sulfadiazine, sulfamerazine and sulfamethazine in propylene glycol + water mixtures at 298.15 K. *J. Mol. Liq.*, **204**, 132–136 (2015). Doi: https://doi.org/10.1016/j.molliq.2015.01.047
50. D.R. Delgado & F. Martínez. Preferential solvation of sulfadiazine, sulfamerazine and sulfamethazine in ethanol + water solvent mixtures according to the IKBI method. *J. Mol. Liq.*, **193**, 152–159 (2014). Doi: https://doi.org/10.1016/j.molliq.2013.12.021
51. D.R. Delgado & F. Martínez. Solution thermodynamics of sulfadiazine in some ethanol + water mixtures. *J. Mol. Liq.*, **187**, 99–105 (2013). Doi: https://doi.org/10.1016/j.molliq.2013.06.011
52. D.R. Delgado & F. Martínez. Solubility and preferential solvation of sulfadiazine in methanol + water mixtures at several temperatures. *Fluid Phase Equilib.*, **379**, 128–138 (2014). Doi: https://doi.org/10.1016/j.fluid.2014.07.013
53. D.R. Delgado, O. Bahamón-Hernandez, N.E. Cerquera, C.P. Ortiz, F. Martínez, E. Rahimpour, A. Jouyban & W.E. Acree, Jr. Solubility of sulfadiazine in (acetonitrile + methanol) mixtures: Determination, correlation, dissolution thermodynamics and preferential solvation. *J. Mol. Liq.*, **322**, 114979 (2021). Doi: https://doi.org/10.1016/j.molliq.2020.114979
54. D.R. Delgado & F. Martínez. Thermodynamic study of the solubility of sodium sulfadiazine in some ethanol + water cosolvent mixtures. *Vitae*, **17**(2), 191–198 (2010). Doi: https://doi.org/10.17533/udea.vitae.6344
55. A.M. Cruz-González, M.S. Vargas-Santana, C.P. Ortiz, N.E. Cerquera, D.R. Delgado, F. Martínez, A. Jouyban & W.E. Acree, Jr. Solubility of sulfadiazine in (ethylene glycol + water) mixtures: Measurement, correlation, thermodynamics and preferential solvation. *J. Mol. Liq.*, **323**, 115058 (2021). Doi: https://doi.org/10.1016/j.molliq.2020.115058
56. D.M. Jiménez, Z.J. Cárdenas, D.R. Delgado, M.Á. Peña & F. Martínez. Solubility temperature dependence and preferential solvation of sulfadiazine in 1,4-dioxane+water co-solvent mixtures. *Fluid Phase Equilib.*, **397**, 26–36 (2015). Doi: https://doi.org/10.1016/j.fluid.2015.03.046
57. D.R. Delgado, C.P. Ortiz, F. Martínez & A. Jouyban. Equilibrium solubility of sulfadiazine in (acetonitrile + ethanol) mixtures: Determination, correlation, dissolution thermodynamics, and preferential solvation. *Int. J. Thermophys.*, **45**, 129 (2024). Doi: https://doi.org/10.1007/s10765-024-03405-4
58. C.F. Trujillo-Trujillo, F. Angarita-Reina, M. Herrera, C.P. Ortiz, R.E. Cardenas-Torres, F. Martinez & D.R. Delgado. Thermodynamic analysis of the solubility of sulfadiazine in (acetonitrile 1-propanol) cosolvent mixtures from 278.15 K to 318.15 K. *Liquids*, **3**(1), 7–18 (2023). Doi: https://doi.org/10.3390/liquids3010002

59. M.S. Vargas-Santana, A.M. Cruz-González, N.E. Cerquera, A.S. Escobar-Rodriguez, R.E. Cardenas, O. Calderón-Losada, C.P. Ortiz & D.R. Delgado. Extended Hildebrand solubility approach and Yalkowsky-Roseman model for estimating the solubility of sulfadiazine and sulfamethazine in some {ethylene glycol (1) + water (2)} mixtures at several temperatures. *Rev. Colomb. Cienc. Quím. Farm.*, **50**(3), 812–836 (2022). Doi: https://doi.org/10.15446/rcciquifa.v50n3.100240
60. D.R. Delgado & F. Martínez. Solution thermodynamics and preferential solvation of sulfamerazine in methanol + water mixtures. *J. Solution Chem.*, **44**(2), 360–377 (2015). Doi: https://doi.org/10.1007/s10953-015-0317-1
61. C.P. Ortiz, D.I. Caviedes-Rubio, F. Martinez & D.R. Delgado. Solubility of sulfamerazine in acetonitrile + ethanol cosolvent mixtures: Thermodynamics and modeling. *Molecules*, **29**(22), 5294 (2024). Doi: https://doi.org/10.3390/molecules29225294
62. D.R. Delgado & F. Martínez. Solubility and solution thermodynamics of sulfamerazine and sulfamethazine in some ethanol + water mixtures. *Fluid Phase Equilib.*, **360**, 88–96 (360). Doi: https://doi.org/10.1016/j.fluid.2013.09.018
63. R.E. Cárdenas-Torres, C.P. Ortiz, W.E. Acree, Jr., A. Jouyban, F. Martínez & D.R. Delgado. Thermodynamic study and preferential solvation of sulfamerazine in acetonitrile + methanol cosolvent mixtures at different temperatures. *J. Mol. Liq.*, **349**, 118172 (2022). Doi: https://doi.org/10.1016/j.molliq.2021.118172
64. M.S. Vargas-Santana, A.M. Cruz-González, C.P. Ortiz, D.R. Delgado, F. Martínez, M.Á. Peña, W.E. Acree & A. Jouyban. Solubility of sulfamerazine in (ethylene glycol + water) mixtures: Measurement, correlation, dissolution thermodynamics and preferential solvation. *J. Mol. Liq.*, **337**, 116330 (2021). Doi: https://doi.org/10.1016/j.molliq.2021.116330
65. C.P. Ortiz, R.E. Cardenas-Torres, M. Herrera & D.R. Delgado. Numerical analysis of sulfamerazine solubility in acetonitrile + 1-propanol cosolvent mixtures at different temperatures. *Sustainability*, **15**(8), 6596 (2023). Doi: https://doi.org/10.3390/su15086596
66. J.H. Blanco-Márquez, C.P. Ortiz, N.E. Cerquera, F. Martínez, A. Jouyban & D.R. Delgado. Thermodynamic analysis of the solubility and preferential solvation of sulfamerazine in (acetonitrile + water) cosolvent mixtures at different temperatures. *J. Mol. Liq.*, **293**, 111507 (2019). https://doi.org/10.1016/j.molliq.2019.111507
67. D.R. Delgado, O.A. Almanza, F. Martínez, M.A. Peña, A. Jouyban & W.E. Acree, Jr. Solution thermodynamics and preferential solvation of sulfamethazine in (methanol + water) mixtures. *J. Chem. Thermodyn.*, **97**, 264–276 (2016). Doi: https://doi.org/10.1016/j.jct.2016.02.002
68. A. Aydi, C.P. Ortiz, D.I. Caviedes-Rubio, C. Ayadi, S. Hbaieb & D.R. Delgado. Solution thermodynamics and preferential solvation of sulfamethazine in ethylene glycol + water mixtures. *J. Taiwan Inst. Chem. Eng.*, **118**, 68–77 (2021). Doi: https://doi.org/10.1016/j.jtice.2020.12.031
69. J.H. Blanco-Márquez, D.I. Caviedes Rubio, C.P. Ortiz, N.E. Cerquera, F. Martínez & D.R. Delgado. Thermodynamic analysis and preferential solvation of sulfamethazine in acetonitrile + water cosolvent mixtures. *Fluid Phase Equilib.*, **505**, 112361 (2020). Doi: https://doi.org/10.1016/j.fluid.2019.112361
70. D.R. Delgado, J.K. Castro-Camacho, C.P. Ortiz, D.I. Caviedes-Rubio & F. Martinez. Dissolution thermodynamics of the solubility of sulfamethazine in (acetonitrile + 1-propanol) mixtures. *Pharmaceuticals*, **17**(12), 1594 (2024). Doi: https://doi.org/10.3390/ph17121594
71. C.P. Ortiz, R.E. Cardenas-Torres, F. Martínez & D.R. Delgado. Solubility of sulfamethazine in the binary mixture of acetonitrile + methanol from 278.15 to 318.15 K: Measurement, dissolution thermodynamics, preferential solvation, and correlation. *Molecules*, **26**(24), 7588 (2021). Doi: https://doi.org/10.3390/molecules26247588
72. I.P. Osorio, F. Martínez, D.R. Delgado, A. Jouyban & W.E. Acree, Jr. Solubility of sulfacetamide in aqueous propylene glycol mixtures: Measurement, correlation, dissolution thermodynamics, preferential solvation and solute volumetric contribution at saturation. *J. Mol. Liq.*, **297**, 111889 (2020). Doi: https://doi.org/10.1016/j.molliq.2019.111889

73. D.R. Delgado, G.A. Rodríguez & F. Martínez. Thermodynamic study of the solubility of sulfapyridine in some ethanol + water mixtures. *J. Mol. Liq.*, **177**, 156–161 (2013). Doi: https://doi.org/10.1016/j.molliq.2012.11.001
74. D.R. Delgado, M.A. Peña, F. Martínez, A. Jouyban & W.E. Acree, Jr. Further numerical analyses on the solubility of sulfapyridine in ethanol + water mixtures. *Pharm. Sci.* (Tabriz), **22**(3), 143–152 (2016). Doi: https://doi.org/10.15171/PS.2016.24
75. D.R. Delgado, G.A. Rodríguez, A.R. Holguín, F. Martínez & A. Jouyban. Solubility of sulfapyridine in propylene glycol + water mixtures and correlation with the Jouyban-Acree model. *Fluid Phase Equilib.*, **341**, 86–95 (2013). Doi: https://doi.org/10.1016/j.fluid.2012.12.017
76. D.R. Delgado, A. Romdhani & F. Martínez. Solubility of sulfamethizole in some propylene glycol + water mixtures at several temperatures. *Fluid Phase Equilib.*, **322–323**, 113–119 (2012). Doi: https://doi.org/10.1016/j.fluid.2012.03.014
77. D.R. Delgado, A. Romdhani & F. Martínez. Thermodynamics of sulfanilamide solubility in propylene glycol + water mixtures. *Lat. Am. J. Pharm.*, **30**(10), 2024–2030 (2011). URL: https://www.latamjpharm.org/resumenes/30/10/LAJOP_30_10_1_23.pdf
78. C.P. Ortíz, R.E. Cardenas-Torres, D.I. Caviedes-Rubio, S.D.J. Polania-Orozco & D.R. Delgado. Thermodynamic analysis and preferential solvation of sulfanilamide in different cosolvent mixtures. *Phys. Chem. Liq.*, **60**(1), 9–24 (2022). Doi: https://doi.org/10.1080/00319104.2021.1888382
79. A.M. Romero-Nieto, D.I. Caviedes-Rubio, J. Polania-Orozco, N.E. Cerquera & D.R. Delgado. Temperature and cosolvent composition effects in the solubility of methylparaben in acetonitrile + water mixtures. *Phys. Chem. Liq.*, **58**(6), 722–735 (2020). Doi: https://doi.org/10.1080/00319104.2019.1636379
80. A.M. Romero-Nieto, N.E. Cerquera, F. Martínez & D.R. Delgado. Thermodynamic study of the solubility of ethylparaben in acetonitrile + water cosolvent mixtures at different temperatures. *J. Mol. Liq.*, **287**, 110894 (2019). Doi: https://doi.org/10.1016/j.molliq.2019.110894
81. A.M. Romero-Nieto, N.E. Cerquera & D.R. Delgado. Measurement and correlation of solubility of ethylparaben in pure and binary solvents and thermodynamic properties of solution. *Rev. Colomb. Cienc. Quim. Farm.*, **48**(2), 332–347 (2019). Doi: https://doi.org/10.15446/rcciquifa.v48n2.82702
82. C.P. Ortiz, R.E. Cardenas-Torres, M. Herrera & D.R. Delgado. Thermodynamic analysis of the solubility of propylparaben in acetonitrile–water cosolvent mixtures. *Sustainability*, **15**(6), 4795 (2023). Doi: https://doi.org/10.3390/SU15064795
83. J.L. Gómez, G.A. Rodríguez, D.M. Cristancho, D.R. Delgado & F. Martínez. Solution thermodynamics of nimodipine in some PEG 400 + ethanol mixtures. *Phys. Chem. Liq.*, **51**(5), 651–662 (2013). Doi: https://doi.org/10.1080/00319104.2013.771262
84. D.R. Delgado, M. Vergara & R.E. Cardenas-Torres. Database of drug solubility in cosolvent mixtures at different temperatures. Mendeley Data, V1, 2025. Doi: https://doi.org/10.17632/g686s23sy5.1
85. M. Vergara-Roa, R.E. Cardenas-Torres & D.R. Delgado. Artificial neural network modeling of drug solubility in cosolvent systems. Mendeley Data, V1, 2025. Doi: https://doi.org/10.17632/3ztcsxvjyb.1
86. Y. Marcus. *The Properties of Solvents.* John Wiley & Sons Ltd, Chichester (UK), 1998.
87. A.F.M. Barton. *CRC Handbook of Solubility Parameters and Other Cohesion Parameters.* 2nd ed. CRC Press, Boca Raton (FL), 1991.
88. G. Scatchard. Solutions of nonelectrolytes. *Annu. Rev. Phys. Chem.*, **3**, 259–274 (1952). Doi: https://doi.org/10.1146/annurev.pc.03.100152.001355
89. Y.-G. Li, T. Teng, J.-F. Lu, G. Chen & J.-D. Li. A study of Scatchard-Hildebrand solution theory for metal solvent extraction systems. *Fluid Phase Equilib.*, **30**, 297–306 (1986). Doi: https://doi.org/10.1016/0378-3812(86)80063-2
90. J.T. Rubino & S.H. Yalkowsky. Cosolvency and cosolvent polarity. *Pharm. Res.*, **4**(3), 220–230 (1987). Doi: https://doi.org/10.1023/A:1016456127893
91. C.M. Hansen. The universality of the solubility parameter. *Ind. Eng. Chem. Product Res. Dev.*, 8(1), 2–11 (1969). Doi: https://doi.org/10.1021/i360029a002

HOW TO CITE THIS ARTICLE

M. Vergara-Roa, R.M. Alvarado-Martinez, A. Chappe-Chappe, R.E. Cárdenas-Torres, C. Rincón-Guio, M.P. Saavedra-Saavedra, C.F. Trujillo-Trujillo & J.D. Aristizabal-Yusty. Evaluation of AI models for predicting drug solubility in cosolvent systems. *Rev. Colomb. Cienc. Quim. Farm.*, **54**(3), 804–820 (2025). Doi: https://doi.org/10.15446/rcciquifa.v54n3.122716