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

ABSTRACT

Nowadays, supercritical fluid technology (SFT) has been an interesting scientific subject in disparate industrial-based activities such as drug delivery, chromatography, and purification. In this technology, solubility plays an incontrovertible role. Therefore, achieving more knowledge about the development of promising numerical/computational methods of solubility prediction to validate the experimental data may be advantageous for increasing the quality of research and therefore, the efficacy of novel drugs. Decitabine with the chemical formula $C_8H_{12}N_4O_4$ is a chemotherapeutic agent applied for the treatment of disparate bone-marrow-related malignancies such as acute myeloid leukemia (AML) by preventing DNA methyltransferase and activation of silent genes. This study aims to predict the optimum value of decitabine solubility in CO_2SCF by employing different machine learning-based mathematical models. In this investigation, we used AdaBoost (Adaptive Boosting) to boost three base models including Linear Regression (LR), Decision Tree (DT), and GRNN. We used a dataset that has 32 sample points to make solubility models. One of the two input features is P (bar) and the other is T (k). ADA-DT (Adaboost Algorithm Decision Tree), ADA-LR (Adaboost Algorithm-Linear Regresion), and ADA-GRNN (Generative Regression Neural Network) models showed MAE of 6.54×10^{-5}, $4.66\ 10^{-5}$, and $8.35\ 10^{-5}$, respectively. Also, in terms of R-squared score, these models have 0.986, 0.983, and 0.911 scores, respectively. ADA-LR was selected as the primary model according to numerical and visual analysis. Finally, the optimal values are (P = 400 bar, T = 3.38 K 102, Y = $1.064\ 10^{-3}$ mol fraction) using this model.