Natural Deep Eutectic Solvents Design Acceleration Using Variational Auto Encoder

Rachid Laref[1], Mehdi Oubahmane[1], Nicolas Rolland[2], Véronique Rataj[2], Adlane Sayede[1]
[1]Univ. Artois, UMR 8181, Unité de Catalyse et de Chimie du Solide (UCCS), Lens, F-62300, France
[2]Centrale Lille, UMR 8181, Unité de Catalyse et de Chimie du Solide (UCCS), Lille, F-59000, France
rachid.laref@univ-artois.fr

The growing demand for sustainable and eco-friendly alternatives in chemistry stems from environmental concerns associated with conventional organic solvents. Among emerging solutions, deep eutectic solvents (DES) have gained significant attention due to their tunable physicochemical properties and versatility in applications such as separations, gas capture, and biocatalysis. A subclass of DES, known as natural deep eutectic solvents (NADES), has emerged as a promising green alternative due to their non-toxic, biodegradable, and low-cost nature. However, the vast number of possible NADES combinations makes experimental characterization impractical. To address this challenge, machine learning (ML) models offer predictive capabilities for key properties such as density, viscosity, and melting temperature. In this study, we integrate ML techniques with supervised variational autoencoders (SVAE) [1] to design novel NADES systems and predict their physicochemical properties. We start by collecting molecular data represented as SMILES and computing relevant descriptors. ML models are then used to predict target properties and identify the most influential descriptors. After selecting the key descriptors, we use a variational autoencoder to project the data into a latent space. Based on desired property values, we generate descriptors for candidate molecules. We then search the COCONUT database for NADES whose descriptors closely match those generated by the supervised VAE. Finally, experimental validation is conducted to verify the predictions and assess the viability of the identified NADES.

References

[1] T. Ji, S. T. Vuppala, G. Chowdhary, et K.Driggs-Campbell, « Multi-Modal Anomaly Detection for Unstructured and Uncertain Environments », 15 décembre 2020, *arXiv*: arXiv:2012.08637. doi: 10.48550/arXiv.2012.08637.

Figures

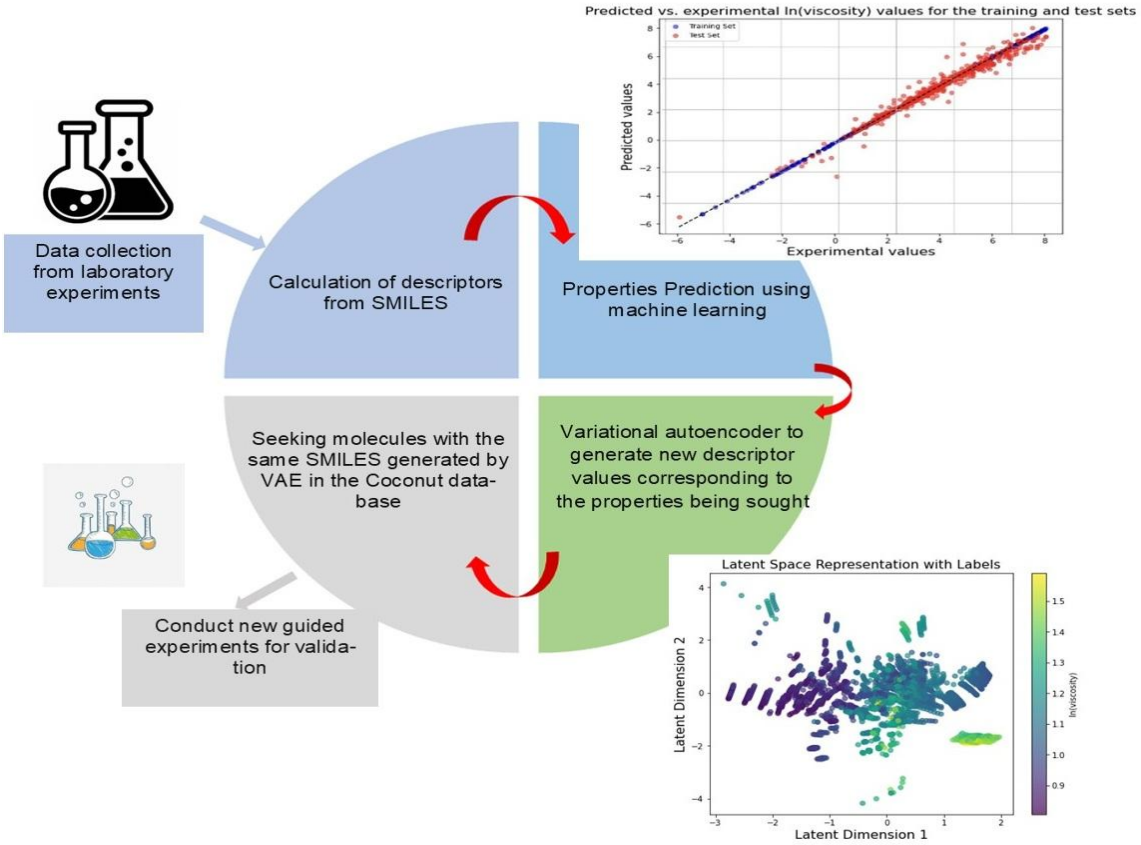

Figure 1. Schematic representation of the workflow for NADES design and property prediction. Molecular data represented as SMILES is used to compute descriptors, followed by feature selection using machine learning models. A supervised variational autoencoder projects the selected descriptors into a latent space, where candidate descriptors are generated based on desired property values. The COCONUT database is then searched for NADES with similar descriptors. Finally, experimental validation is conducted to confirm the predictions and assess the feasibility of the identified NADE