Can Machine Learning help to Design Task Specific Ionic Liquid?

Gyanendra Sharma,[a,*] Mengyang Qu,[a] Naoki Wada[a] and Kenji Takahashi [a,*]

[a]*Faculty of Biological Science and Technology, Institute of Science and Engineering, Kanazawa University, Kakuma-machi, Kanazawa, 920-1192, Japan*
**Email: sharmag-19@se.kanazawa-u.ac.jp, ktkenji@staff.kanazawa-u.ac.jp*

A new organic ion generator was developed using recurrent neural networks (RNN) and Monte Carlo Tree Search (MCTS) to explore the wide, unexploited chemical space of ionic liquids (ILs).[1] The process of ion generation is divided into two stages: (I) training an RNN on a dataset of existing ion structures and (II) employing MCTS to guide the RNN in generating new organic ion structures. The initial dataset consisted of 117,700,516 SMILES strings retrieved from PubChem. The RNN was trained to predict the next possible symbol in the SMILES sequence, generating new SMILES strings for potential IL candidates. MCTS employs a search tree to generate new symbols and simulates multiple SMILES strings to identify chemically valid and novel structures, using RDKit for verification. The generator produced billions of IL candidates, comprising 961,859 de novo cations and 957,070 de novo anions. To visualize the diversity of the generated ions, their chemical structures were encoded using the extended connectivity fingerprint (ECFP4) and mapped to 2D space using the Uniform Manifold Approximation and Projection (UMAP) technique Fig 1. The generated novel ionic liquid can be filter for task specific applications. Here in this study ILs are subjected to screen through a machine learning model, which has been pre-trained for cellulose solubility prediction, and further evaluating using a quantum chemistry model. In the next step we have synthesized few ILs to validate the machine learning and quantum chemistry screening method. Here it is important to bring in notice that synthesized ILs are screened from existing ILs data base to test the model because the generated ILs structure are challenging to synthesize however, not impossible.

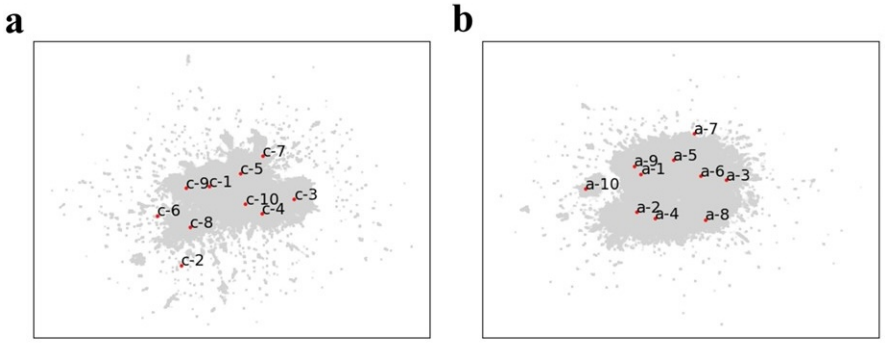
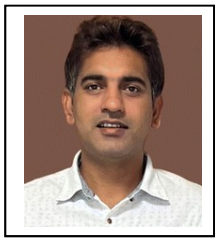


Fig 1. (a) The UMAP plot of generated *de novo* cations is depicted as gray dots. Randomly selected cations are highlighted in red for structural display. (b) The UMAP plot of generated *de novo* anions is shown as gray dots. Randomly selected anions are highlighted in red for structural display.

References

1. Qu, M.; Sharma G.*; Wada N.; Ikebata H.; Matsunami S.; Takahashi K* *J. Cheminform.* **2025,** Just Accepted.



Gyanendra Sharma received his PhD in 2017 from the Indian Institute of Technology Madras. Since 2023, he has been serving as an Assistant Professor at Kanazawa University, where his research focuses on the modification of polysaccharides in ionic liquid media. He employs computational chemistry and machine learning as tools to support his experimental work. With over a decade of experience in the synthesis and application of ionic liquids, he has co-authored 22 publications and holds one Indian patent.