From drugs to supercapacitor materials: Improving knowledge transfer methods for data-scarce materials discovery

Witold Taisner[1,*], Anna Przybylowska[1], Dariusz Brzezinski[1], Agnieszka Hryniewicka[2], Julia Kolodziejczyk[2], Kinga Cieciuch-Tymoniuk[2], Marta Eliza Plonska-Brzezinska[2]

[1]Institute of Computing Science, Poznan University of Technology, Piotrowo 2, 60-965 Poznan, Poland, [2]Department of Organic Chemistry, Medical University of Bialystok, Mickiewicza 2A, 15-222 Bialystok, Poland, *Correspondence to: witold.taisner@cs.put.poznan.pl

Accelerating materials discovery requires sifting through vast numbers of candidate molecules, a task where AI-driven inverse design shows promise. However, current generative models—tailored primarily for drug design—struggle with the unique chemical goals and small datasets of materials science. In this work, we investigate how to adapt these models to data-scarce material domains. Using the discovery of nanoporous carbon materials as a test case, we benchmark three generative AI architectures across transfer learning strategies including pre-training, fine-tuning, and reinforcement learning. Our experiments span two molecular representations that differ in their inherent chemical properties, enabling us to assess how representation choices influence the balance between generating chemically valid structures and satisfying domain-specific design rules.

One of the central challenges in generative chemistry is evaluating candidates against multiple, competing criteria. To address this, we introduce VUCS, a composite measure that quantifies the ratio of valid, unique, and constraint-satisfying molecules—a practical metric for data-scarce domains where multiple constraints must be simultaneously satisfied. Our benchmarking experiments show that combining fine-tuning and reinforcement learning can increase VUCS from ~1% up to 8% — more than an order-of-magnitude improvement over baselines.

We also propose a novel sampling-based molecular set extension method that aggregates generated candidates across multiple experiments and reuses them to enhance transfer learning, increasing the percentage of unique, valid candidate substrates from at most 8% to over 47%. We validate this approach by synthesizing three generated substrates, which form nanoporous covalent triazine frameworks upon pyrolysis, achieving highly competitive specific capacitance for this family of materials. These results demonstrate that systematic knowledge transfer, domain-aware evaluation via VUCS, and sampling-based data augmentation offer a practical route for adapting drug-design generative models to data-scarce materials science.

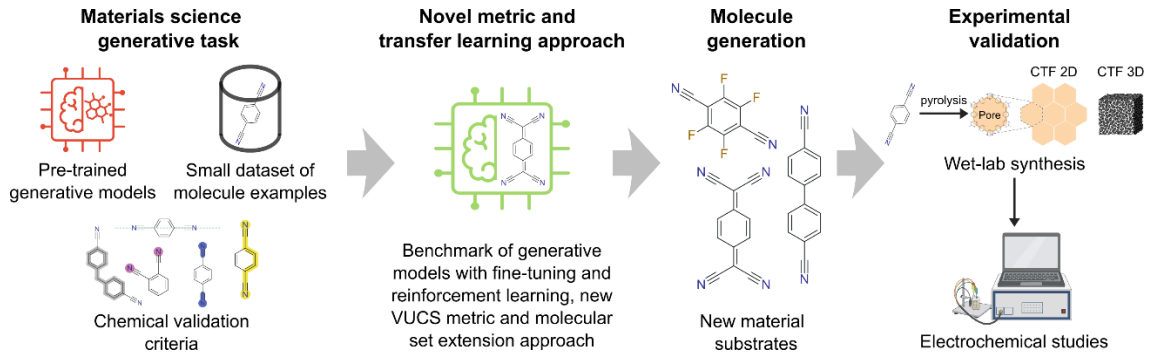

Figure 1: Workflow for adapting drug-design generative models to covalent triazine framework substrate discovery for supercapacitor electrodes.