Multimodal AI for de novo CNS Drug Design via Molecular Generation

Lars Arntsen
Deep Imaging Analytics Lab
Rensselaer Polytechnic Institute
Biomedical Engineering Society 2025 Conference

Introduction:

Discovering and optimizing small molecules for central nervous system (CNS) applications presents unique challenges related to accessibility and safety, due to the interdependent nature of key properties such as pharmacokinetics, blood-brain barrier (BBB) permeability, and toxicity. *De novo* drug design is traditionally target-oriented, beginning with a known disease-associated binding site and optimizing compounds that show activity against it. This study explores an alternative approach in which a generative model is trained to produce Simplified Molecular Input Line Entry System (SMILES) strings optimized for CNS properties, operating within the CNS chemical space before incorporating any target-specific considerations. Four multimodal classifiers for BBB permeability, neuronal cytotoxicity, mammalian neurotoxicity, and neural activity were developed to design a feature-based curriculum for reinforcement learning (RL) aimed at generating CNS-optimized compounds. The goal is to create a dynamic tool adaptable to various CNS applications that also accelerates the drug development process.

Materials and Methods:

A SMILES-based molecular generator was developed using RL with a curriculum-based reward strategy. The model was initialized with MolGPT, a GPT-2-based generator pretrained on the ZINC-15 dataset and sourced from the Hugging Face repository. It was optimized using policy gradient RL to produce valid and novel SMILES strings with CNS-relevant properties.

The reward function began by emphasizing validity and novelty. Validity was assessed using RDKit, an open-source cheminformatics toolkit, that conducts a sanitization process which involves multiple substeps checking if the molecule follows common chemical conventions. Novelty was measured by computing Tanimoto similarity between generated molecules and compounds in DrugBank's "All drugs" dataset. Every 1,000 training steps, a new feature was added to the reward function until all key molecular property constraints were included. These features were selected based on SHapley Additive exPlanations (SHAP) applied to four binary LightGBM classifiers trained to predict BBB permeability, neuronal cytotoxicity, mammalian neurotoxicity, and neural activity. The classifiers were multimodal, using chemical descriptors (from RDKit and PaDEL) and transformer-derived features extracted from Chemformer embeddings.

For each classifier, SHAP was used to rank features by importance. The most influential ones were introduced earlier in the reward curriculum. For shared descriptors across classifiers, only one threshold value could be applied to the curriculum, so the threshold that aligned with the desired directional effect was selected. To monitor progress, a checkpoint was saved every 5,000 steps, and a sample of 100 SMILES strings was evaluated to track average reward values and validity.

Results, Conclusions, and Discussions:

The curriculum-based RL approach led to significant improvements in the validity and property alignment of CNS-optimized SMILES strings. Validity increased from 79% at 5,000 steps to a peak of 94% between 45,000–50,000 and 80,000 steps, before slightly declining to 88% at 85,000 steps. The average reward, reflecting the proportion of satisfied property constraints, rose steadily from 15.19% to 78.13%, indicating that the model effectively learned to generate molecules aligned with CNS-relevant profiles.

The SMILES string-based generator, guided by multimodal property classifiers and curriculum RL, demonstrated robust learning and convergence. Performance plateaued after 45,000 steps, suggesting efficient optimization of the CNS-relevant chemical space. The final checkpoint produced valid, property-aligned molecules, confirming the effectiveness of curriculum RL for property-driven molecular design.

This study highlights the value of gradually introducing property constraints to balance novelty, validity, and multi-objective optimization. The plateau in performance suggests that key features were learned early, while later fluctuations in validity likely reflect the trade-off between novelty and strict constraint enforcement. Future work will alter training length, expand the property set, and incorporate target-specific information to generate molecules suitable for experimental validation and potential clinical development.

Figures:

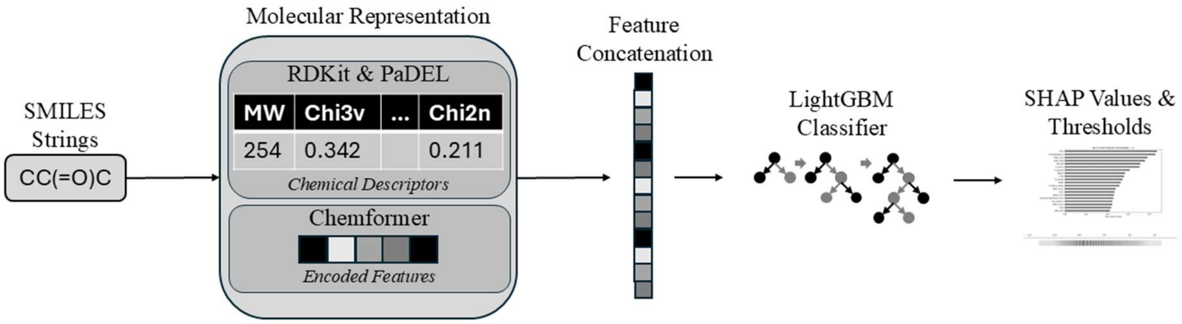

Figure 1. Curriculum Reinforcement Learning Preparation Pipeline. SMILES strings from a property-labeled dataset are featurized using three parallel encoding routes: RDKit and PaDEL generate chemical descriptors, while Chemformer produces transformer-derived embeddings. These features are concatenated and used to train a binary LightGBM classifier for each target property. SHAP analysis is applied to identify the most influential features and determine threshold values for use in the reward function of the reinforcement learning curriculum.

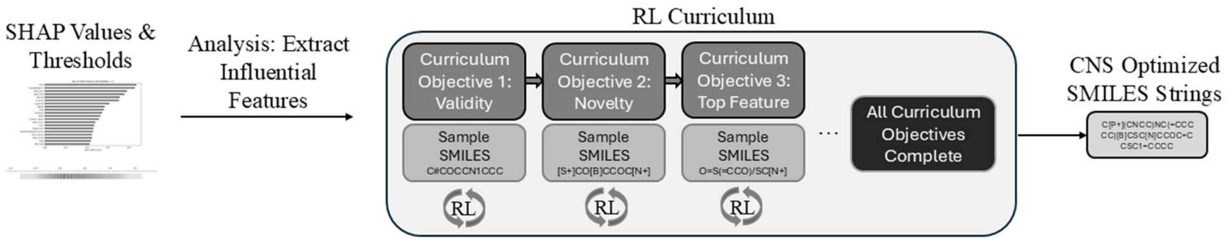
SHAP Values & Thresholds
Analysis: Extract Influential Features
RL Curriculum
Curriculum Objective 1: Validity
Curriculum Objective 2: Novelty
Curriculum Objective 3: Top Feature
Sample SMILES
C#COCCN1CCC
Sample SMILES
[S+]CO[B]CCOC[N+]
Sample SMILES
O=S(=CCO)/SC[N+]
RL
RL
RL
…
All Curriculum Objectives Complete
CNS Optimized SMILES Strings
C[P+](CNCC)NC(=CCC
CC)[B]CSC[N]CCOC=C
CSC1=CCCC

Figure 2. Curriculum Reinforcement Learning Design Overview. Key features and their threshold values were extracted from the property classifiers via SHAP analysis. For descriptors shared across classifiers, thresholds aligned with the desired directional effect were selected (e.g., the highest value if higher values correlated with improved outcomes). The curriculum began by optimizing for validity, followed by novelty, and then incrementally introduced one new SHAP-ranked feature every 1,000 training steps in descending order of importance. The final output is a SMILES string with a high likelihood of being valid, novel, and compatible with the CNS chemical space.

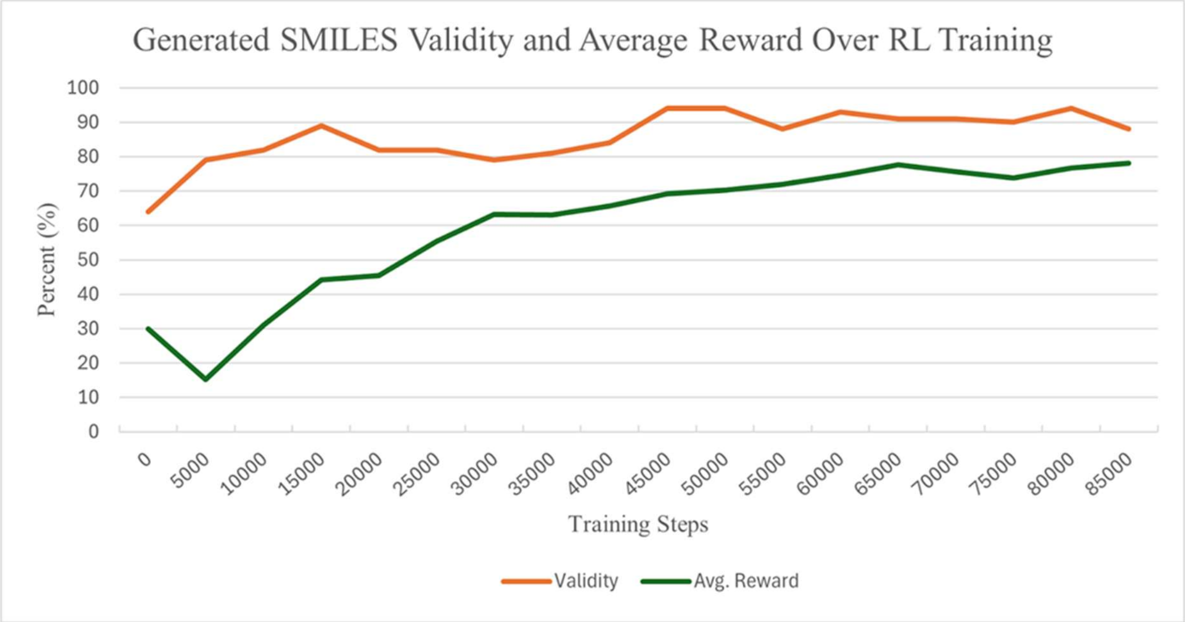

Figure 3. Reinforcement Learning Curriculum Training Overview. Over the course of 85,000 training steps, model checkpoints were saved every 5,000 steps. At each checkpoint, 100 SMILES strings were sampled. Their chemical validity and corresponding reward values were calculated and averaged to monitor training progress.