香港中文大學
The Chinese University of Hong Kong

Journal Club

Hybrid Computational Strategy for Predicting Complex Ligand–Metal Architectures

Zihan Li

Angew. Chem. Int. Ed. **2026**, *65*, e24655.

Content

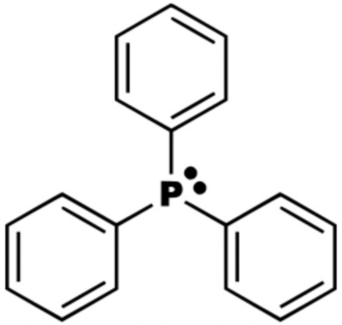
1. Introduction
2. Results
3. Summary

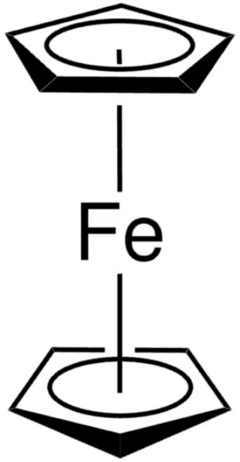
1. Introduction

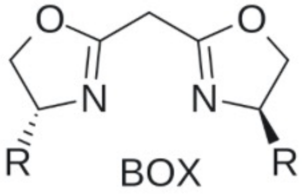
Introduction

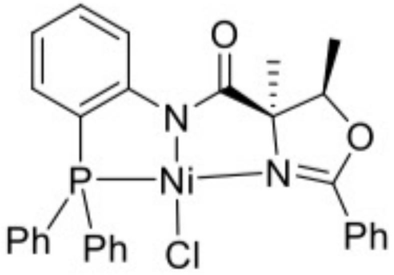
Ligand + Metal → Complex → Catalyze the reaction
(promote the reaction, change selectivity etc)

I^- CO









Different ligands have different ways to coordinate with different metals

↓

Current coordination model can't solve the problem:

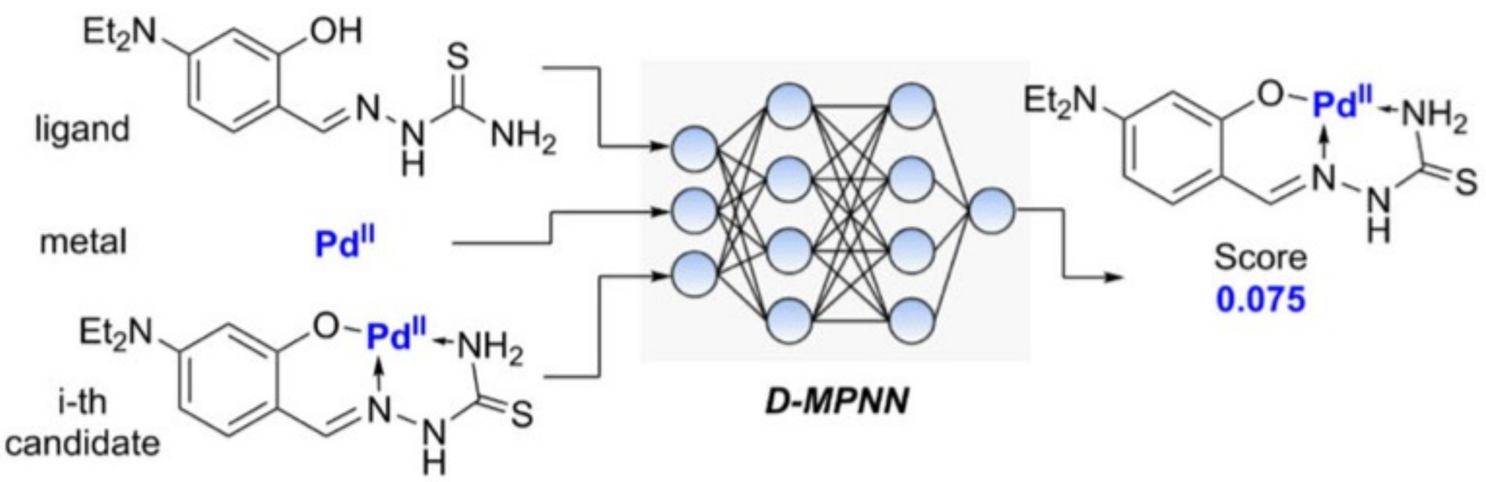
1. Don't consider the metal type
2. Can't be used for odd-hapticity ligands (eg. allyl) and high-denticity ligands

Workflow for this work

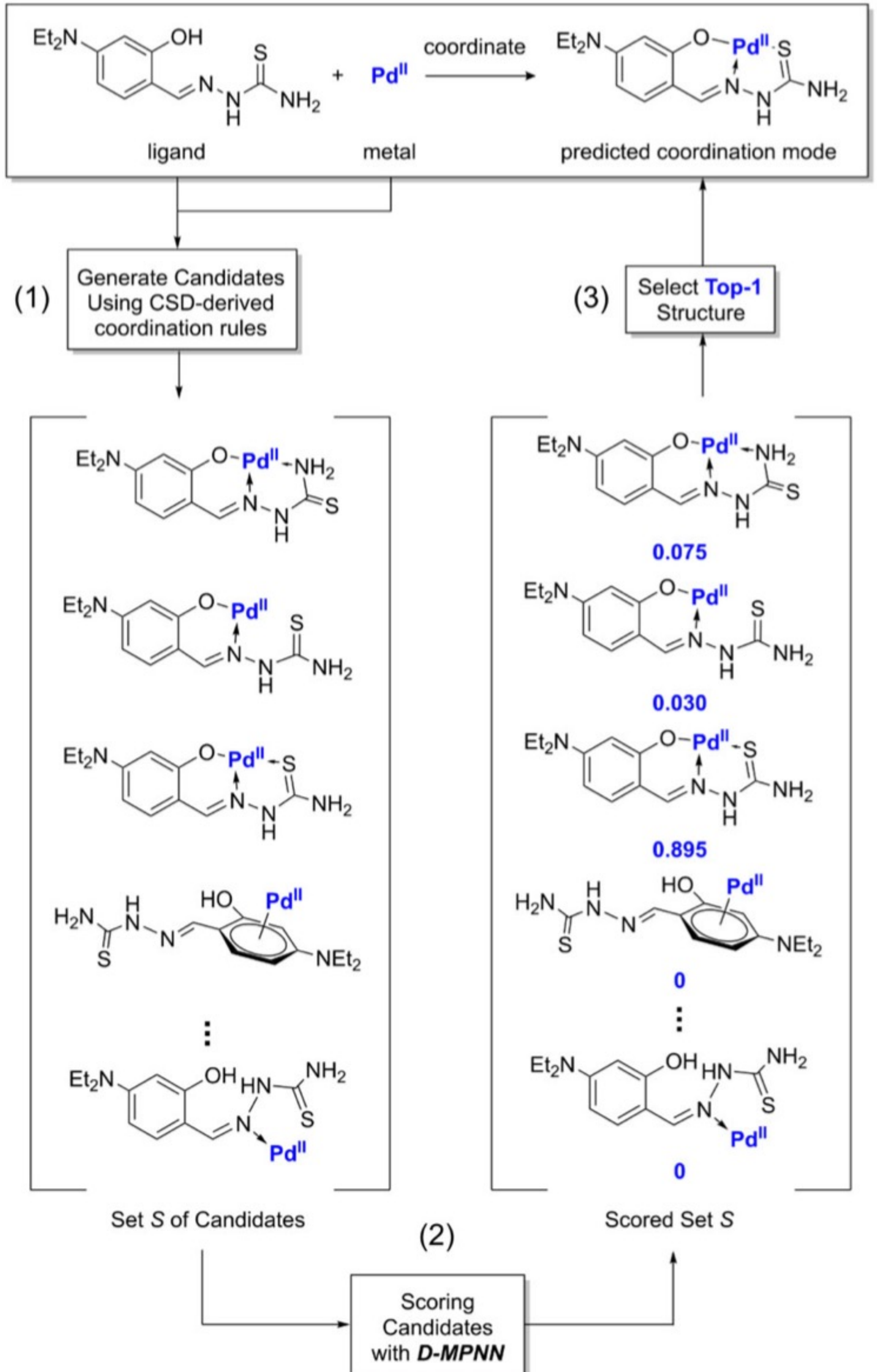
Rule + ML

From real CSD complexes

Scoring the candidates



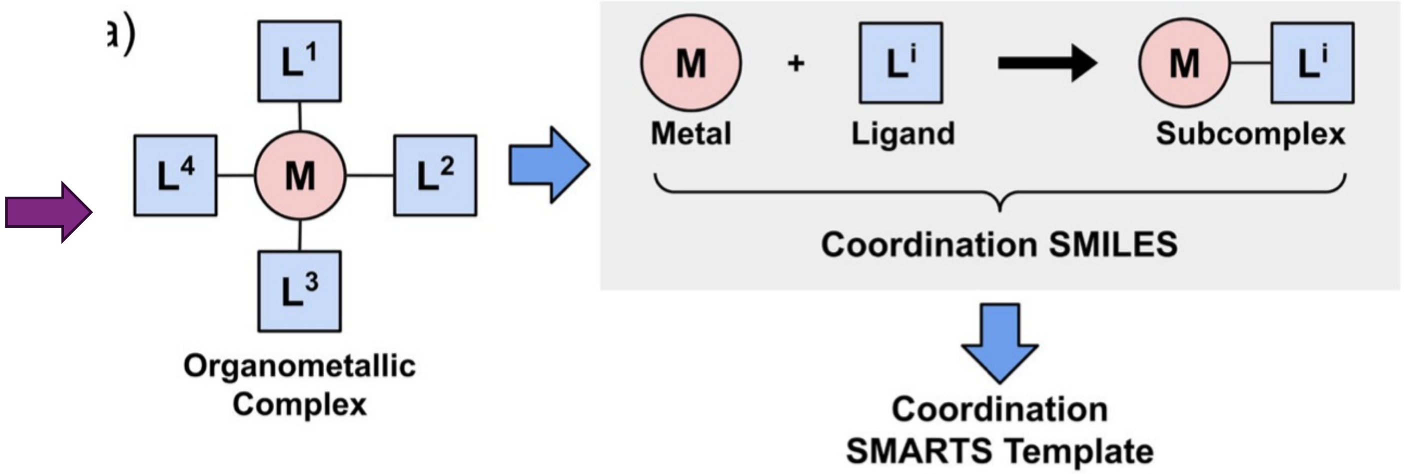
Directed Message Passing Neural Network

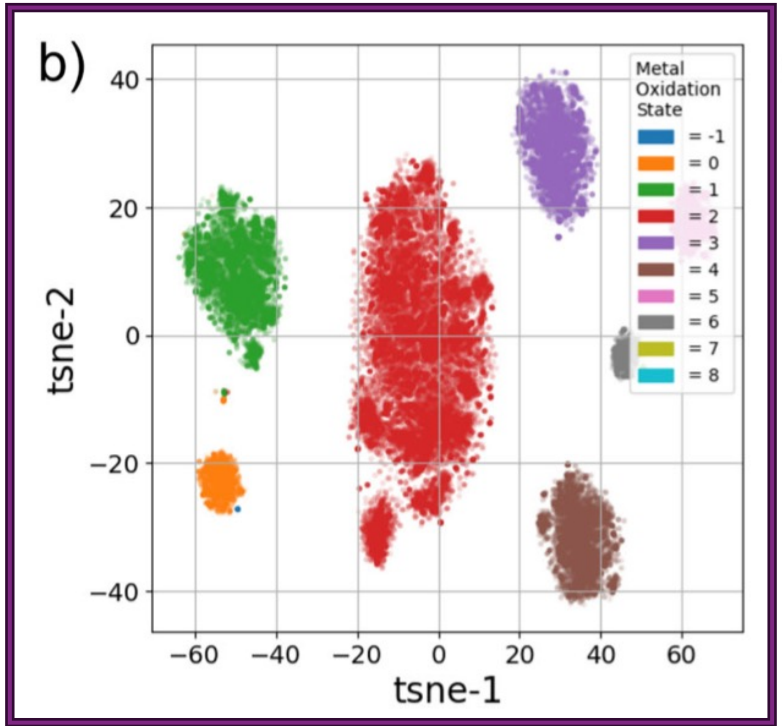


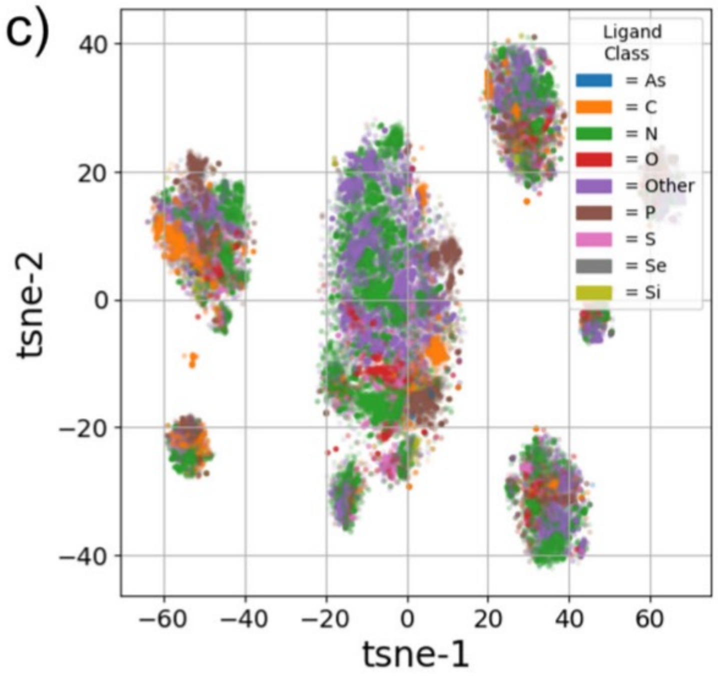
2. Results

Results—Data Preparation

DATABASE

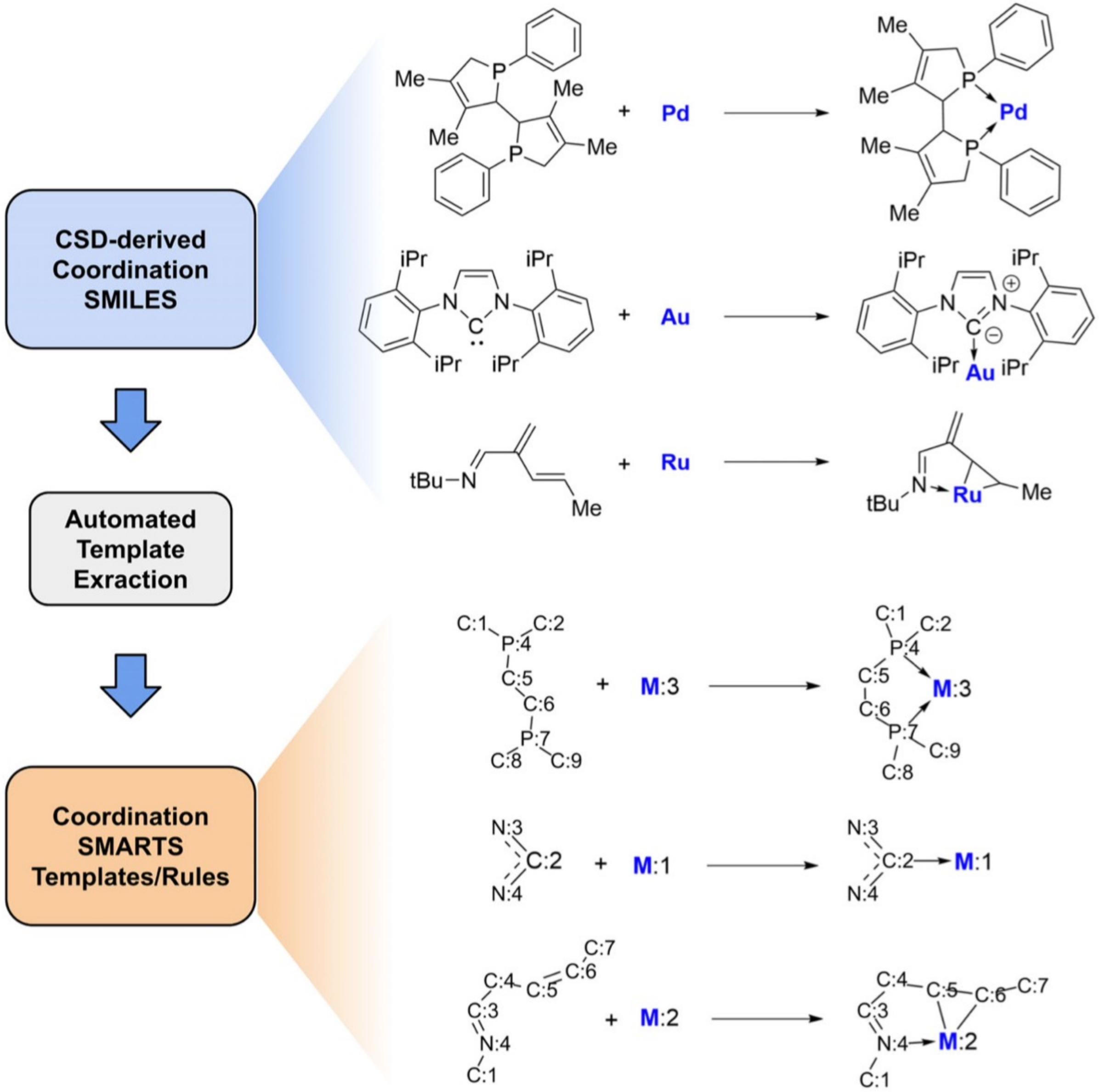






The authors minimized ligand overlap between the training/validation and holdout sets, particularly by using a Morgan fingerprint-based cluster split that **reduced ligand repetition** to about 0.80%, in order to prevent the model from merely memorizing training structures and to improve its generalizability.

Results—Extracting Coordination Rules



Atom-mapped coordination SMILES

Abstract coordination-templates (RDChiral)

Define the coordination core

Generic metal representation + combine fragments

RDKit RunReactants validation

$$\text{ligand} + \text{metal} \xrightarrow{\text{SMARTS template}} \text{predicted subcomplex}$$

Organize and add additional templates manually

Results—Generation of Coordination Candidates

RDKit RunReactants generation

$$1 \text{ metal} + 1 \text{ ligand} \rightarrow \text{ candidate subcomplexes}$$

Ni + NOS-tridentate ligand

Ni-N
Ni-O
Ni-S
Ni-(N,O)
Ni-(N,O,S)
…

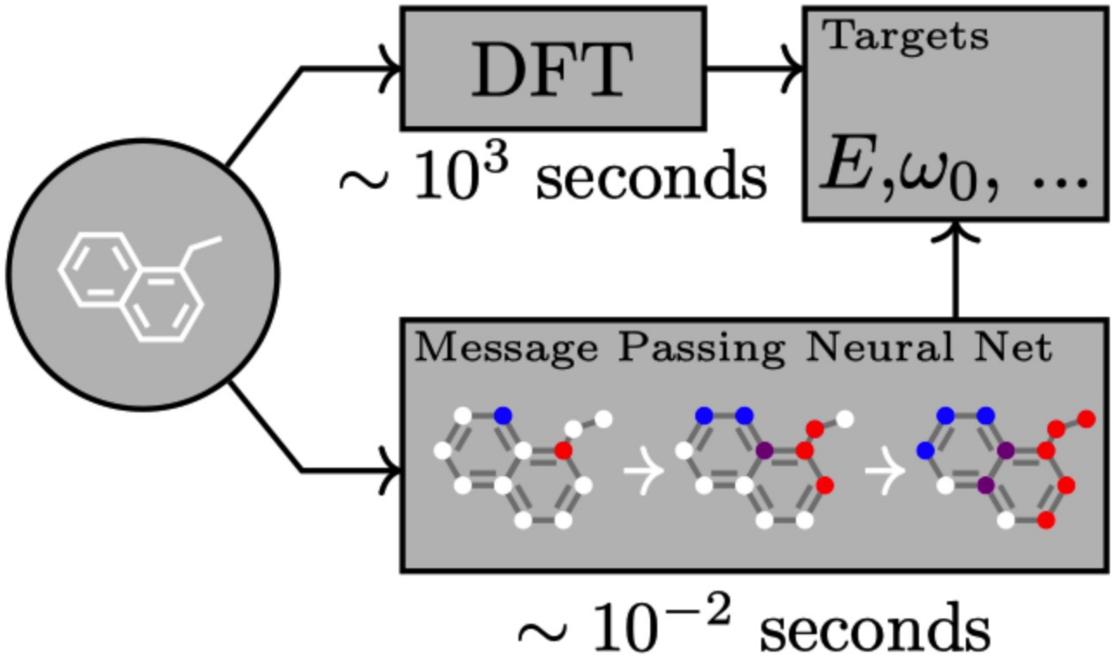
SMILES is a line notation used to represent a specific molecular structure as a text string, including atoms, bonds, branches, rings, and sometimes stereochemistry.

SMARTS is a rule-based language used to describe and match chemical substructure patterns, such as functional groups, donor-atom environments, or reaction cores within molecules.

Results—Scoring of Coordination Modes Using the D-MPNN Model

directed-message passing neural network, D-MPNN

Graph neural network that represents a molecule as atoms connected by bonds and passes information along directed **bonds**.
By repeatedly sharing information between neighboring bonds and atoms, it learns the **local and broader chemical environment** of a molecule.
The learned molecular representation can then be used to **score** or classify structures, such as predicting which metal–ligand coordination mode is most likely correct.



three SMILES inputs

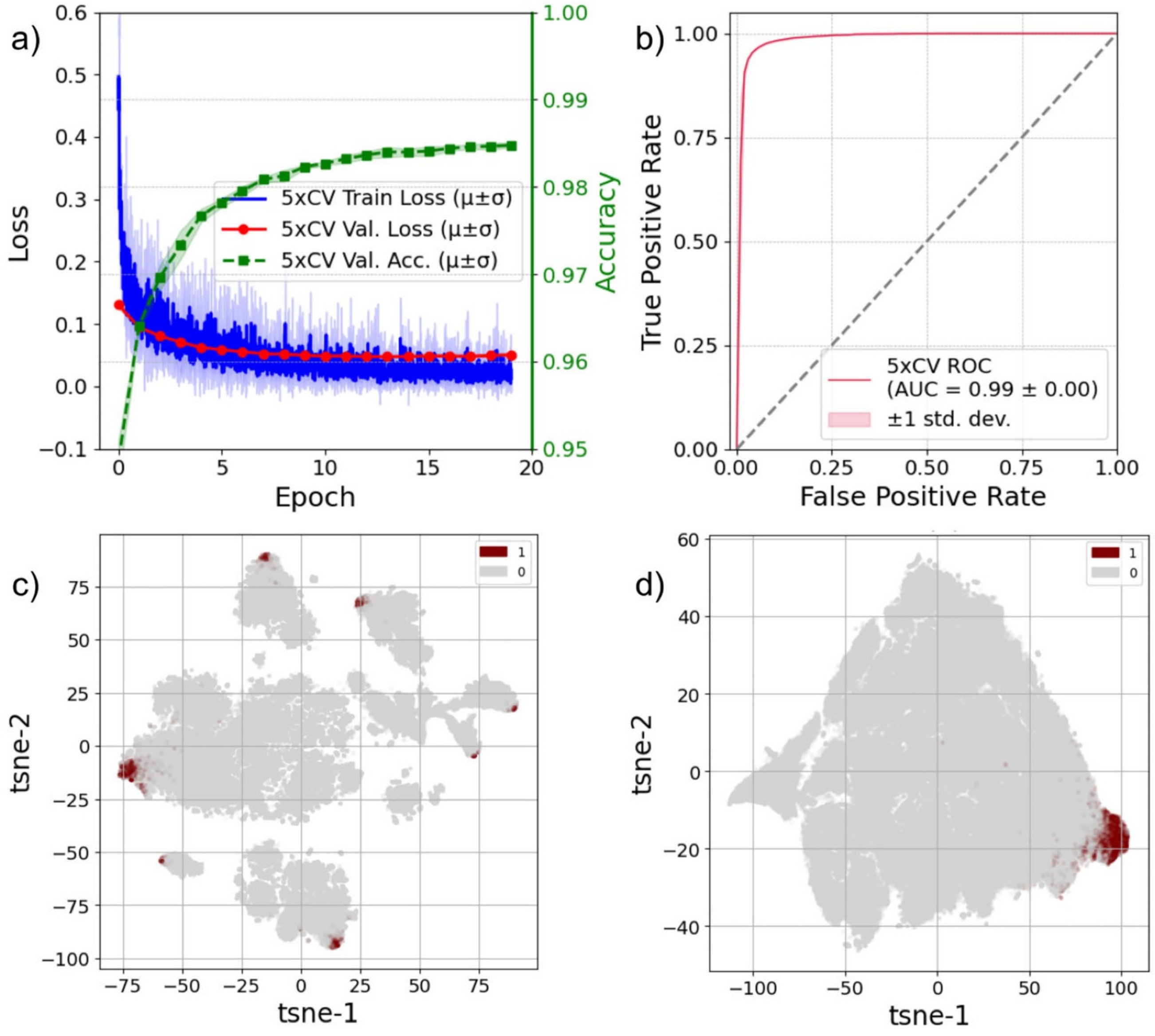
Ligand structure
Metal ion (oxidation state)
Candidate subcomplex

Chemprop →

atomic number, formal charge, valence, implicit hydrogens, unpaired electrons
bond type and bond direction

Score and classify →

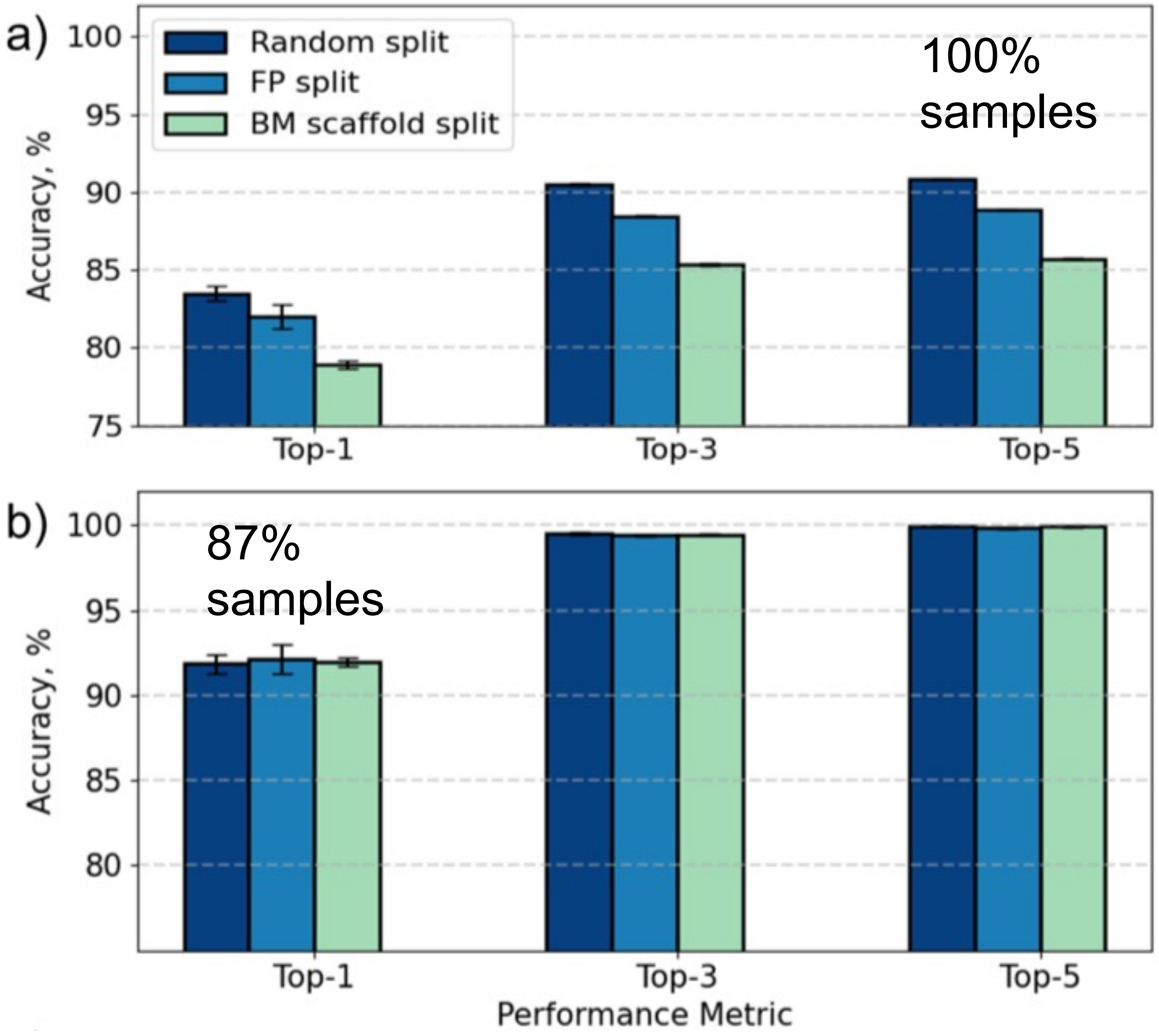
Results—Scoring of Coordination Modes Using the D-MPNN Model

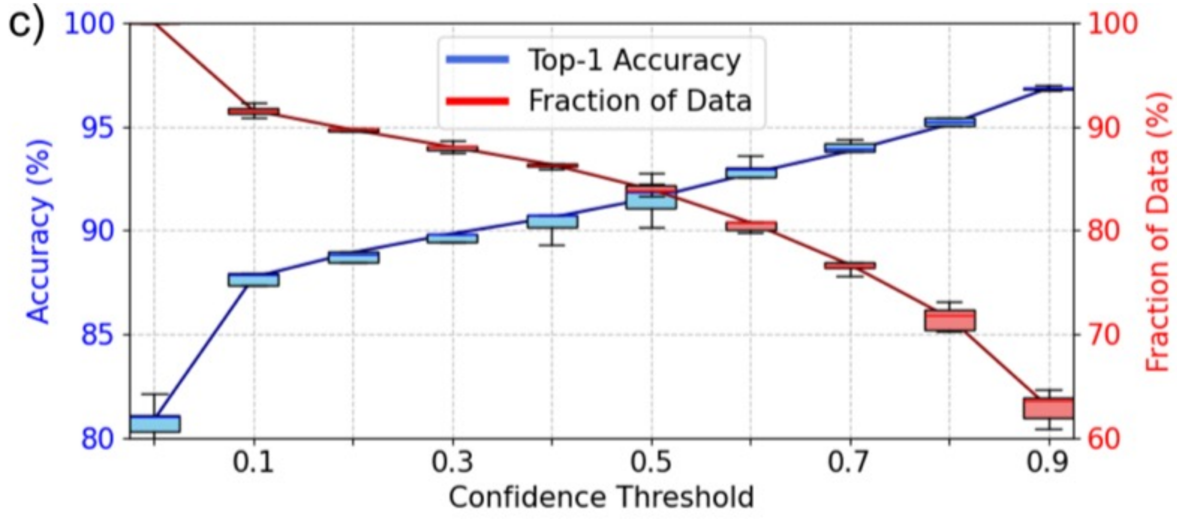


c) Before training

d) After training

Results—Top-K Prediction on the Holdout Set



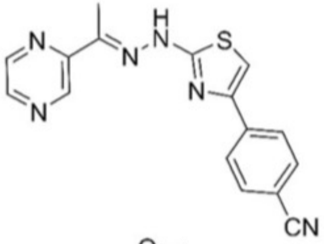
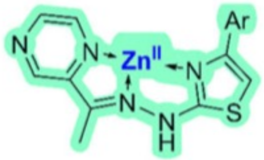
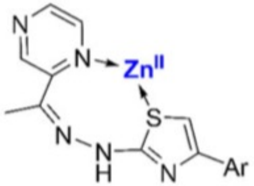
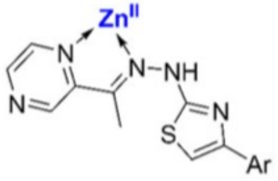
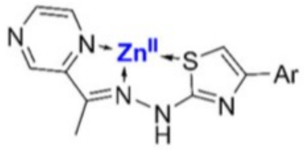
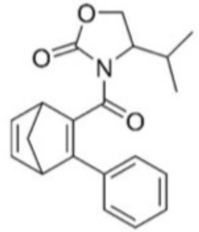
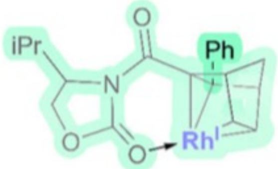
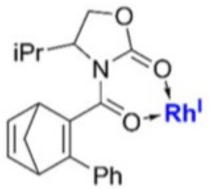
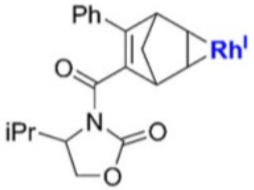
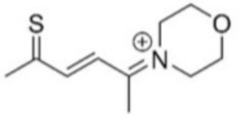
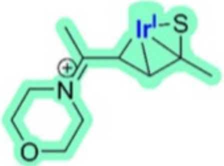
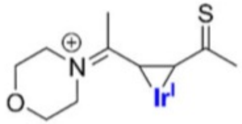
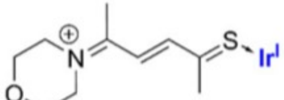
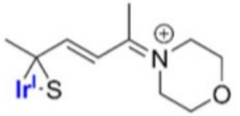
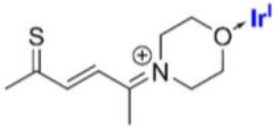
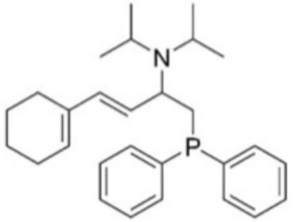
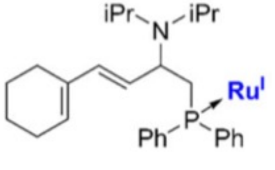
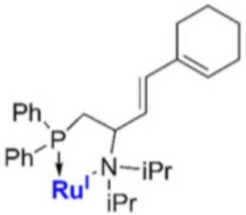
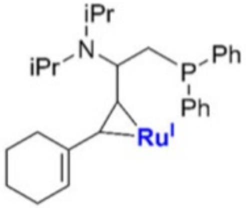


If no ML part (only use templates), the prediction average accuracy is 49.85%

Morgan fingerprint split (all structure)

Bemis–Murcko split (core-coordinate sphere)

Results—Application

#	Ligand	Metal	Top-1	Top-2	Top-3	Top-4	Top-5
1)		Zn^{II}	 0.99 ± 0.01	 0.00 ± 0.00	 0.00 ± 0.00	 0.00 ± 0.00	 0.00 ± 0.00
2)		Rh^{I}	 0.77 ± 0.13	 0.02 ± 0.03	 0.01 ± 0.01	 0.00 ± 0.00	 0.00 ± 0.00
3)		Ir^{I}	 0.77 ± 0.13	 0.01 ± 0.01	 0.01 ± 0.01	 0.00 ± 0.00	 0.00 ± 0.00
4)		Ru^{I}	 0.45 ± 0.22	 0.02 ± 0.01	 0.00 ± 0.00	 0.00 ± 0.00	 0.00 ± 0.00

3. Summary

This work developed a hybrid workflow that combines knowledge-based rules with machine learning to predict metal–ligand coordination modes.

TO Develop:

1. The model can only predict coordination patterns that are covered by existing templates.
2. Its scoring performance depends strongly on the coverage of the SMARTS template library.
3. Moderate performance for special ligand classes and ligands with weakly coordinating sites.
4. Its treatment of flexible ligands, sterically complex ligands and solvent effects needs improvement.
5. For multi-ligand complexes (INT), the current treatment remains rather approximate.
6. The model currently does not provide access to multi-metal complexes.

Thank You