Xinhao Li

North Carolina State University, Raleigh, NC

Phone: +1-919-607-1990

Email: xli74@ncsu.edu

Website: https://xinhaoli74.github.io

Ph.D. candidate in Chemistry (expecting to graduate in ***Nov*. 2020**). Expertise in applying **machine learning** techniques and **cheminformatics** to solve chemistry problems, *e.g.,* QSAR modeling. Well-versed in programming languages **Python** and **R** and machine learning toolkits such as **PyTorch** and **Scikit-Learn**. Proficient skills in chemical data mining, curation, analysis, visualization, and modeling.

Experience

Graduate Research Assistant *Advisor*: **Denis Fourches** *Aug 2017 – Nov* 2020 (*expected*)

- Development of Novel Quantitative Structure-Property/Activity Relationship (**QSPR/QSAR**) Modeling Methodologies:
  - **MolPMoFiT**: an effective **transfer learning** method based on **self-supervised pre-training + task-specific fine-tuning** for QSAR modeling. MolPMoFiT pre-trained a universal molecular structure prediction model using one million unlabeled molecules from ChEMBL and then fine-tuned it for various QSPR/QSAR tasks. (*J Cheminform* 2020, 12, *27*)
  - **Hierarchical QSAR**: An effective ensemble/stacking modeling method that Integrating binary, multiclass, and regression models for predicting acute oral systemic toxicity. (*Chem. Res. Toxicol.* 2020*, 33, 353–366*)
- **SMILES Pair Encoding** (**SPE**): a data-driven substructure tokenization algorithm for deep learning.
  - SPE splits SMILES into human-readable and chemically explainable substrings and shows superior performances on both generative and predictive tasks compared to the atom-level tokenization (*ChemRxiv* 2020)

Computational Sciences Intern **GlaxoSmithKline, Collegeville, PA** *May 2020 – Aug 2020*

- Explored transfer learning approaches to QSAR modeling for lead optimization endpoints.
- Developed twp pre-trained models based on LSTM and Transformer for transfer learning.
- Benchmarking machine learning algorithms (SMILES-based and Graph-based deep learning models, lightGBM etc.) on two internal and eight public datasets.

Education

PhD in Chemistry	**North Carolina State University**	Raleigh, NC	2017 – 2020.11 (*expected*)
MS in Chemistry	**Beijing University of Chemical Technology**	Beijing, China	2013 –2016
BS in Chemistry	**Beijing University of Chemical Technology**	Beijing, China	2009 –2013

Skills

- **Programming Toolkits:** Python, R, Git, Linux
- **Cheminformatics Toolkits:** KNIME, RDKit, Schrodinger
- **Machine Learning Toolkits:** Pytorch, Keras, Scikit-Learn, Streamlit, Jupyter Notebook

Publications

1. **Xinhao Li**, Nicole Kleinstreuer and Denis Fourches. Hierarchical Quantitative Structure–Activity Relationship Modeling Approach for Integrating Binary, Multiclass, and Regression Models of Acute Oral Systemic Toxicity. *Chemical Research in Toxicology*. **2020**, *33*, 353-366.
2. **Xinhao Li** and Denis Fourches. Inductive Transfer Learning for Molecular Activity Prediction: Next-Gen QSAR Models with MolPMoFiT. *J Cheminform* **2020**, 12, *27*.
3. **Xinhao Li** and Denis Fourches. SMILES Pair Encoding: A Data-Driven Substructure Tokenization Algorithm for Deep Learning. *ChemRxiv* **2020**