# 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. Experienced in applying **machine learning** techniques and **cheminformatics** for solving chemistry problems, *e.g.,* molecular activity/propriety prediction. Well-versed in programming languages including **Python** and **R**. Proficient skills in chemical data mining, curation, analysis, visualization, and modeling.

## Research Interests

- Cheminformatics
- QSAR Modeling
- Computational Toxicology
- Machine Learning/Deep learning
- Transfer Learning
- Data Mining, Analysis, Visualization

## Research

**Graduate Research Assistant** Fourches Lab, NCSU
*Advisor*: **Denis Fourches** *Aug 2017 - Current*

- 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. (*ChemRxiv* 2019)
  - **Hierarchical QSAR**: An effective ensemble/stacking modeling method that Integrating binary/multi classification and regression models for predicting acute oral systemic toxicity. (*Chemical Research in Toxicology* 2020)
- Chemical Data Curation and Validation:
  - **Mol**ecular **D**ata**S**ets (**MolDS**): A toolkit (Python Package) for curating, standardizing and diagnosing Molecular Data Sets for benchmarking machine learning methods. (GitHub)
- Software Development:
  - **CryptoChem**: A novel cryptographic and data storage method based on cheminformatics, machine learning, and big chemical data. Two independent software were developed: MOLWRITE and MOLREAD. MOLWRITE encrypts the text/image data into chemical message and MOLREAD decrypts the encoded chemical message back to text/image.

## Education

| | | | |
|---|---|---|---|
| **PhD** in Chemistry | **North Carolina State University** | Raleigh, NC | 2017 – **2021.05** (*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
- **Cheminformatics Toolkits:** KNIME, RDKit, Schrödinger, ChemAxon
- **Machine Learning Toolkits:** Pytorch, Keras, Scikit-Learn, Streamlit

# Data Science Related Courses

- **deeplearning.ai:** Deep Learning Specialization; **Algorithmic Toolbox** (Coursera)

# 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. (2019): Inductive Transfer Learning for Molecular Activity Prediction: Next-Gen QSAR Models with MolPMoFiT. *ChemRxiv*.
3. **Xinhao Li** and Jiaxi Xu. (2017): Effects of the Microwave Power on the Microwave-assisted Esterification. *Current Microwave Chemistry*. 158-162.
4. **Xinhao Li** and Jiaxi Xu. (2017): Identification of Microwave Selective Heating Effort in an Intermolecular Reaction with Hammett Linear Relationship as a Molecular Level Probe. *Current Microwave Chemistry*. 339-346.
5. **Xinhao Li** and Jiaxi Xu. (2016): Determination on temperature gradient of different polar reactants in reaction mixture under microwave irradiation with molecular probe. *Tetrahedron*. *35*, 5515-5520.
6. Shanyan Mo, **Xinhao Li** and Jiaxi Xu. (2014): In Situ-Generated Iodonium Ylides as Safe Carbene Precursors for the Chemoselective Intramolecular Buchner Reaction. *J. Org. Chem*. 19, 9186-9195.

# Presentations

1. **NC State Chemistry Recruitment Weekend (March 2019, Raleigh, NC)**
   **Xinhao Li,** Denis Fourches. Hierarchical H-QSAR Modeling Method that Integrates Binary/Multi Classification and Regression Models for Predicting Acute Oral Systemic Toxicity. (**Poster**)
2. **American Chemical Society Conference (April 2019, Orlando, FL)**
   **Xinhao Li,** Denis Fourches. Hierarchical H-QSAR Modeling Method that Integrates Binary/Multi Classification and Regression Models for Predicting Acute Oral Systemic Toxicity. (**Poster**)
3. **"Innovations in Agriculture" Scientific Poster Session at BASF (May 2019, RTP, NC)**
   **Xinhao Li,** Denis Fourches. Hierarchical H-QSAR Modeling Method that Integrates Binary/Multi Classification and Regression Models for Predicting Acute Oral Systemic Toxicity. (**Poster**)
4. **NCSU/BASF Poster Session (Aug 2019, Raleigh, NC)**
   **Xinhao Li,** Denis Fourches**.** Transfer Learning for Molecular Property/Activity Prediction. (**Poster**)
5. **Triangle Machine Learning Day (Sep 2019, Durham, NC)**
   **Xinhao Li,** Denis Fourches**.** Transfer Learning for Molecular Property/Activity Prediction. (**Oral Presentation** & **Poster**)
6. **AI Powered Drug Discovery and Manufacturing (Feb 2020, MIT, Cambridge, MA)**
   **Xinhao Li,** Denis Fourches**.** Inductive Transfer Learning for Molecular Activity Prediction. (**Poster**)

# Awards

1. **CINF Scholarship for Scientific Excellence awarded by ACS CINF (2019 Spring)**