Shuang Ni

+1 (514)549-0276 | ni_shuang@yahoo.com

SUMMARY

PhD candidate specializing in machine learning, dimensionality reduction, and applications to biological problems. Skilled in uncovering insight from large-scale high-dimensional data. Seeking internship or part-time opportunities in machine learning research and engineering, data analysis, or computational biology research to leverage my expertise and drive impactful solutions.

EDUCATION

Université de Montréal, Montreal, Canada *09/2021 - 12/2026*

- PhD. in Computer Science (GPA: 4.30 /4.30) *(Anticipated)*
- Doctoral Researcher at Mila - Quebec AI Institute

Concordia University, Montreal, Canada *01/2019 - 05/2021*

- M.Sc. in Electrical and Computer Engineering (GPA: 4.30 /4.30)
- Transferred credits from the University of Manitoba (09/2017 - 12/2018):
  Statistical Aspect of Machine Learning, Parallel Processing, Advanced Signal Processing, Digital Image Processing

University of Electronic Science and Technology of China, Chengdu, China *09/2013 - 07/2017*

- Bachelor of Electronic Information Engineering (GPA: 3.47/4.00)

SKILLS

- **Programming:** Python, R, PyTorch, TensorFlow, Scikit-learn, C, C++.
- **Tools**: MATLAB, Mathematica, Linux, GitHub, HPC.

EXPERIENCE

Senior Data Science Intern | Analysis Group - Montreal, Canada

Synthetic Patient Journey Generation using Generative AI *05/2025 - 08/2025*

- Designed a disease progression model to simulate obesity and related cardiometabolic conditions concurrently over time.
- Adapted the Conditional Restricted Boltzmann Machine (CRBM) architecture to model longitudinal patient data under diverse synthetic interventions across multiple time steps.
- Extended CRBM model by incorporating additional patient history and integrating an autoencoder for fixed-length embeddings.
- Improved real-world pattern preservation, resolved systematic drift in synthetic variables, identified and diagnosed performance issues, and supported evaluation and validation of the improved model.

Doctoral Researcher| Mila - Quebec AI Institute - Montreal, Canada

Modeling Dynamic Mechanisms of Drug Treatment Effects with GeoSinkhorn Flow *05/2024 - present*

- Collaborated on the design and validation of the GeoSinkhorn Flow model, assessing its capability to model geodesic distances and dynamic transport across distributions.
- Conducted experiments on synthetic manifolds to demonstrate the model's ability to simulate geodesic distances on manifolds.
- Evaluated the model on scRNA-seq datasets, predicting drug combination effects and temporal treatment responses, demonstrating its effectiveness in dynamic modeling.

Representation Learning for Continuous Population Structure in Human Genomic Biobanks *05/2024 - present*

- Designed dimensionality reduction methods combining autoencoders and PHATE for multiple human genetic biobanks.
- Applied archetypal analysis to estimate admixture signals, revealing ancestral genetic contributions within populations.
- Benchmarked the learned representations against alternative methods using both quantitative metrics and qualitative visualization, demonstrating improved continuity and cluster separation aligned with population genetics.

Enhancing Semi-Supervised Visualization with Autoencoders and Random Forest Proximities *09/2023 - 09/2025*

- Designed and implemented various novel architectures combining geometry-regularized autoencoders with Random Forest-based PHATE to extend embeddings for out-of-sample data points while preserving the intrinsic manifold structure.
- Introduced a proximity-based landmark selection method, reducing training time by 40% without compromising extension quality.
- Evaluated performance on large-scale datasets, showing robust and meaningful embeddings across different parameter settings.
- Conducted extensive experiments on image and single-cell datasets, benchmarking against other visualization methods through quantitative and qualitative analysis. RF-AE consistently outperformed all competing models.

Machine Learning-Based Biomarker Discovery and Cellular Landscape Analysis in Multiple Sclerosis (MS) *09/2022 - 09/2023*

- Corrected batch effects and integrated scRNA-seq data from multiple patients of MS and other related diseases.

- Profiled and annotated cell types and subtypes by clustering scRNA-seq data and assigning identities based on known marker genes, using PHATE and Multiscale PHATE.
- Built and validated regression models using clinical and immunological data to identify key biomarkers and predict 1-year EDSS progression in MS patients, achieving $R^2 \geq 0.765$ on held-out validation data.

Graduate Research Assistant | Concordia University - Montreal, Canada

Muscle Oxygen Saturation Quantitative Measurement and Error Detection *09/2017 - 09/2020*

- Developed a methodology to non-invasively and quantitatively measure muscle oxygen saturation (SmO2) using 5-wavelength diffuse reflectance continuous-wave near-infrared spectroscopy (NIRS).
- Employed non-linear least squares fitting of the 5-wavelength measured attenuation spectrum to a Taylor expansion attenuation model, enabling the determination of SmO2 values for calf muscles in various workout statuses.
- Utilized SVM with an RBF kernel in Scikit-learn to classify labeled attenuation spectrum datasets and predict the accuracy of collected data.
- Designed an application for SmO2 calculation and fault diagnostics using MATLAB App Designer.

Deep Learning Mechanism of Revenue Maximization in Mobile Edge Computing *09/2019 - 04/2021*

- Designed and implemented a revenue maximization incentive mechanism using deep learning techniques.
- Developed a custom neural network model with TensorFlow to optimize revenue in the context of cooperative task offloading within a mobile edge computing system.
- Formulated a Lagrangian function as a loss function to effectively address the revenue maximization problem while considering three practical constraints.

PUBLICATIONS

- A. Aumon*, S. Ni*, M. Lizotte, G. Wolf, K. R. Moon and J. S. Rhodes, "Random Forest Autoencoders for Guided Representation Learning," *NeurIPS 2025* (Poster, accepted)
- S. Ni, A. Aumon, G. Wolf, K. R. Moon and J. S. Rhodes, "Enhancing Supervised Visualization Through Autoencoder and Random Forest Proximities for Out-of-Sample Extension," *2024 IEEE 34th International Workshop on Machine Learning for Signal Processing (MLSP)*, London, United Kingdom, 2024, pp. 1-6
- G. Li, J. Cai and S. Ni, "Truthful Deep Mechanism Design for Revenue-Maximization in Edge Computing with Budget Constraints," in *IEEE Transactions on Vehicular Technology*, vol. 71, no. 1, pp. 902-914, Jan. 2022

ACTIVITIES

Poster Presentation:

- "PHATE effectively captures continuous population structure in human genomic data", Probabilistic Methods in Genetics, Cold Spring Harbor, March 2025.
- "Enhancing Supervised Visualization Through Autoencoder and Random Forest Proximities for Out-of-Sample Extension", MLSP, September 2024
- "PHATE representation can effectively capture continuous population structure in human genomic data", Machine Learning in Computational Biology, September 2024

Tutorial:

- "PHATE representation can effectively capture continuous population structure in human genomic data", GRAM Workshop @ ICML 2024, https://colab.research.google.com/drive/1B-pJMtgQJ97XL4UhuscR5NibF2D3Izt2?usp=sharing

HONORS & AWARDS

- Bourse d'exemption UdeM, 2021
- Concordia Merit Scholarship, 2019
- International Graduate Student Entrance Scholarship of University of Manitoba, 2017
- Third-class Scholarship of UESTC, 2016
- Outstanding Volunteer of UESTC, 2014