XINYU DONG

7 New York Ave., Stony Brook, NY 11790.
+1 631 428 4324 ⋄ xinyu.dong.1@stonybrook.edu
Personal Page: https://xinyudong93.github.io/

EDUCATION

Stony Brook University, Stony Brook, NY — August 2018 - May 2024(anticipated).
PhD in Computer Science, GPA 3.94/4.0.
Advisor: Prof. Fusheng Wang.
Research work on building machine learning based predictive models on large scale healthcare data and provide interpretations (e.g. opioid overdose risk prediction)
Teaching Assistant for courses ISE305(Database Design and Practice), CSE526(Principals of Programming Languages) and CSE532(Theory of Database System), CSE534(Fundamentals of Computer Networks)

Stony Brook University, Stony Brook, NY — August 2015 - May 2017
Master in Computer Science, GPA 3.97/4.0

Tianjin University, Tianjin, China — August 2011 - July 2015
Bachelor in Electronic Commerce, Distinguished Thesis Honor

WORKING EXPERIENCE

Regeneron Pharmaceuticals — Tarrytown, New York
PhD Intern - Clinical Informatics, Manager: Nilanjana Banerjee — May 2022 - August 2022
· Participate in building BERT based models for clinical prediction.
· Mainly responsible for interpretation the prediction results from BHERT model.
· Program in python with Pytorch, scikit-learn, numpy and pandas.

Baidu USA — Sunnyvale, California
Research Intern, Manager: Tianyi Gao — June 2021 - August 2021
· Participate in building baidu's own computing resources analyze tool.
· Responsible for the analyze result validation and computation bottleneck detection.
· Program in python with scikit-learn, numpy and pandas packages.
· Use NVIDIA Nsight system, DLProf and their GUI tools to generate analyzed data.

SKILLS AND INTERESTS

Solid understanding on **machine learning** and **deep learning** algorithms.
Proficient in **Python** and **pandas**, **numpy** and **scikit-learn** packages
Experience in machine learning framework including **keras**, **tensorflow** and **Pytorch**
Worked with visulization tools including **Tableau** and **R-Shiny**.
Worked with **Vertica**, **PostgreSQL** and **MYSQL** databases.
Familiar with **Java**, **R**, **Linux** and **SQL**.

PUBLICATION

Dong X, Wong R, Lyu W, Abell-Hart K, Deng J, Liu Y, Hajagos JG, Rosenthal RN, Chen C, Wang F. An integrated LSTM-HeteroRGNN model for interpretable opioid overdose risk prediction. Artificial intelligence in medicine. 2023 Jan 1;135:102439.

Dong X, Rashidian S, Wang Y, Hajagos J, Zhao X, Rosenthal RN, Kong J, Saltz M, Saltz J, Wang F. Machine learning based opioid overdose prediction using electronic health records. InAMIA Annual Symposium Proceedings 2019 (Vol. 2019, p. 389). American Medical Informatics Association.

Dong X, Deng J, Hou W, Rashidian S, Rosenthal RN, Saltz M, Saltz JH, Wang F. Predicting opioid overdose risk of patients with opioid prescriptions using electronic health records based on temporal deep learning. Journal of biomedical informatics. 2021 Apr 1;116:103725.

Dong X, Deng J, Rashidian S, Abell-Hart K, Hou W, Rosenthal RN, Saltz M, Saltz JH, Wang F. Identifying risk of opioid use disorder for patients taking opioid medications with deep learning. Journal of the American Medical Informatics Association. 2021 Aug;28(8):1683-93.

Lyu W, **Dong X**, Wong R, Zheng S, Abell-Hart K, Wang F, Chen C. A Multimodal Transformer: Fusing Clinical Notes with Structured EHR Data for Interpretable In-Hospital Mortality Prediction. AMIA Annu Symp Proc.

Deng J, Hou W, **Dong X**, Hajagos J, Saltz M, Saltz J, Wang F. A Large-Scale Observational Study on the Temporal Trends and Risk Factors of Opioid Overdose: Real-World Evidence for Better Opioids. Drugs-Real World Outcomes. 2021 May 26:1-4.

Yao H, Rashidian S, **Dong X**, Duanmu H, Rosenthal RN, Wang F. Detection of suicidality among opioid users on reddit: Machine learningbased approach. Journal of medical internet research. 2020;22(11):e15293.

Rashidian S, **Dong X**, Jain SK, Wang F. EaserGeocoder: integrative geocoding with machine learning (demo paper). InProceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems 2018 Nov 6 (pp. 572-575).

Rashidian S, **Dong X**, Avadhani A, Poddar P, Wang F. Effective scalable and integrative geocoding for massive address datasets. InProceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems 2017 Nov 7 (pp. 1-10).

Rashidian S, Abell-Hart K, Hajagos J, Moffitt R, Lingam V, Garcia V, Tsai CW, Wang F, **Dong X**, Sun S, Deng J. Detecting Miscoded Diabetes Diagnosis Codes in Electronic Health Records for Quality Improvement: Temporal Deep Learning Approach. JMIR Medical Informatics.

Chen J, Zhang L, **Dong X**. Clustering Personalized 3D Printing Models with Multiple Modal CNN. InChinese Intelligent Systems Conference 2017 Oct 14 (pp. 703-712). Springer, Singapore.

Rashidian S, Hajagos J, Moffitt R, Wang F, **Dong X**, Abell-Hart K, Noel K, Gupta R, Tharakan M, Lingam V, Saltz J. Disease phenotyping using deep learning: A diabetes case study. arXiv preprint arXiv:1811.11818 (2018).

RESEARCH PROJECTS DESCRIPTION

Opioid Epidemic Prevention August 2018 - present.

Apply deep learning methods on large scale clinical database to predict risk for opioid misuse related diseases risks of patients

- Responsible for the whole pipeline from dataset extraction, model implementation and performance evaluation.
- Predicted diseases includes opioid use disorder and opioid overdose.
- Use **SQL** to extract data from **large scale Electronic Health Records database** of nearly 9 million patients information.
- Apply **LSTM**, **Hierachical GNN** and **transformer** based **BERT** models mainly to build the mode.
- Apply sequential model to simulate the temporal progression of disease development.
- Employ graph network model to exploit knowledge from complex relations among patients and clinical concepts.
- Use different **interpretation** methods to make the model more understandable, including **shapley value**, **perturbation**, **permutation**, etc.
- Programming with Python and related packages including numpy, pandas, scikit-learn.
- Model implementation with deep learning frameworks including **tensorflow**, **keras** and **pytorch**.
- Achieved a prediction performance with AUROC score of 0.89 and F1 score of 0.8.
- Use **tableau** as the tool to visualize the work.
- Currently working on integrating clinical notes data with natural language processing methods.

EaserGeocoder: Integrative Geocoding for Address Data August 2016 - August 2017

Develop a geographic information system to process structured text address to return latitude and longitude.

- Mainly responsible for development and machine learning application work.
- Integrate with multiple New York state geographic datasets.
- The system can work offline to prevent private and sensitive data from being uploaded online.
- Implement the system with both Python and Java.
- Apply decision tree boosting classifier to enhance the searching result.
- Details available on http://bmidb.cs.stonybrook.edu/easergeocoder/index

OTHER INFORMATION

Attend kaggle competition of **The Nature Conservancy Fisheries Monitoring**, got a silver medal(ranking top 5%), available on https://www.kaggle.com/c/the-nature-conservancy-fisheries-monitoring/overview .

Certificate for **Neural Networks for Machine Learning**
available on https://www.coursera.org/account/accomplishments/certificate/R5TWU3W2F3K7

Certificate for **Data Science**,
available on https://www.coursera.org/account/accomplishments/specialization/certificate/HJJNX4467Z86