**Yu Deng**

CONTACT INFORMATION

Feinberg School of Medicine, 303 E Superior street, Chicago, IL.

[Email:yudeng2015@u.northwestern.edu](mailto:yudeng2015@u.northwestern.edu)

Mobil: 3124839916

RESEARCH INTEREST

I am a PhD student majoring in biomedical informatics. My focus is on applying statistical learning methods for clinical research. I have applied various machine learning techniques to find disease subtypes and novel risk factors. I have used deep learning/joint modeling for disease prediction using time series data (i.e. repeated measurements) as well as using Natural Language Processing (NLP) for computational phenotyping.

# EDUCATION Northwestern University, Chicago, IL, USA

* PhD Candidate, biomedical informatics, Feinberg School of Medicine, expected graduation June 2021
* Major GPA: 3.79/4.00
* Coursework: Programming for Big Data, Advanced Biostatistics, Statistical Theory and Method

# INTERNSHIP Data Science Intern, National Institute of Health (NIH), Bethesda, MD, USA, 2019

* Interned at a natural language processing research group at National Library of Medicine.
* Analyzed full text chemical corpus for downstream analysis such as name entity recognition.
* Developed customized CNN architectures utilizing longitudinal data for cardiovascular disease prediction.

SKILLS

SELECTED PROJECTS

SELECTED PUBLICATIO NS/CONFERE NCES

HONOURS/ DISTINCTIONS

# Programming Skills

* Python, R, SQL, Hadoop

# Machine Learning Algorithms

* **Deep Learning** (MLP, CNN, RNN), Classical & Penalized Regression Methods (LASSO, Ridge), SVM, Random Forest, KNN, Adaboosting, Cox regression; K-means, hierarchical clustering, non-Negative Tensor Factorization

# Statistical Analysis

* A/B testing, causal inference, missing data imputation

**Developed Deep Learning-survival Model for Cardiovascular Disease (CVD) Prediction, 2019**

* Developed CNN based models to forecast cardiovascular disease using time series data. Our model tackled the problems of patient irregular visit patterns, subject drop-out, and correlation within repeated measurements.
* Designed 3 convolution layers with convolution kernels convolved with the layer input over the temporal dimension.
* Explore different parameter settings include learning rate, number of hidden layers, epoch, activation function, early stop.
* Improved AUC by 3% compared to cox baseline model.

**Use of Clinical Phenotypes and Non-negative Tensor Factorization (NTF) for Heart Failure (HF) Prediction, 2017**

* Performed NTF on large scale, sparse medical record data; Generated latent clusters.
* Performed dimension reduction including NTF, MTF, PCA on medical record data to get important features.
* Used the output of different feature reduction techniques as the input features of logistic regression.

**Use of NLP to Improve Systemic Lupus Erythematosus Criteria Identification in Electronic Health Records, 2019**

* Developed SQL queries to extract large scale clinical notes from hospital data warehouse.
* Used Metamap to extract name entities, used l1/l2 penalized logistic regression used for feature selection and prediction.
* Evaluated model performance using sensitivity and specificity. NLP based algorithm improved renal sensitivity from 0.4 to 0.7.

***Comparison of the State-of-Art Neural Network Survival Models vs Cox PH on Cardiovascular Disease Prediction***

**Deng Y.**, Peng Y., Wei Y., Lu Z., Zhao L. (2021). Manuscript in preparation

***Natural Language Processing to Identify Lupus Nephritis Phenotype in Electronic Health Records***

**Deng Y.,** Pacheco J., Walunas T., Luo Y. (2021). Manuscript submitted to BMC biomedical informatics and decision making

***Identification of Systemic Lupus Erythematosus Subtypes Using Latent Class Analysis***

**Deng Y.,** Ghosh A., Luo Y., Kho A., Goldman R., Walunas T. (2021). Manuscript in preparation

***Association of Second-line Type 2 Diabetes Medication and Chronic Kidney Disease***

**Deng Y.,** Ghamsari F., Lu A., Yu J., Kho A. (2021). Manuscript in preparation

# *BERTSurv: BERT-Based Survival Models for Predicting Outcomes of Trauma Patients*

Zhao Y., Hong Q., **Deng Y**., Wang Y., Petzold. (2021). IEEE international conference on Data Mining (ICDM).

***Significance of Intra-operative Medication Data and Predictive Model Selection for Predicting Post-operative Atrial Fibrillation in Patients without Prior Atrial Fibrillation***

Yu J., Johnson E., **Deng Y.,** Kho A. (2021). Manuscript submitted to International Journal of Medical Informatics

# *Natural Language Processing for EHR-Based Computational Phenotyping*

Zeng, Z., **Deng Y**., Li X., Naumann T., & Luo Y. (2018). Natural Language Processing for EHR-Based Computational Phenotyping. IEEE/ACM transactions on computational biology and bioinformatics.

***Characterizing Design Patterns of EHR-Driven Phenotype Extraction Algorithms***

Zhong, Y., Rasmussen L., **Deng Y**., Hripcsak G., Chute C., Luo Y. (2018). IEEE on Bioinformatics and Biomedicine

**Honorable Mention in Student Poster Competition,** International Chinese Statistical Association (ICSA), 2018

**First Prize in Student Poster Competition,** Northwestern Biomedical Informatics Day, 2017

Driskill Graduate Program Full Scholarship, Northwestern University 2015 – 2017