XINLEI DENG

National Institutes of Health, NIA & NIEHS, Baltimore, MD & Durham, NC
Tel: (518)253-2901/Email: xinlei.deng@nih.gov/Homepage: https://xinleideng.github.io

SUMMARY OF QUALIFICATION

- Proficient in Python, SQL, AWS, GCP, R (R-shiny), and Git.
- Applied experience of building online tools using machine learning, biostatistics, and deep learning techniques.
- Extensive experience in creating interactive visualization of data via Tableau, R-shiny, DASH, and PowerBI
- Published 28 research papers in peer-reviewed journals including The Lancet, Radiology, etc.

EDUCATION

Ph.D. (Epidemiology), State University of New York at Albany, USA   05/2022
Relevant Coursework: *Statistical Inference, Machine Learning, Deep Learning*

M.D. (Preventive Medicine & Biostatistics), Sun Yat-sen University, Guangzhou, China   05/2019
Relevant Coursework: *Mathematical Modelling, Biostatistics, VB Programming, Big Data and MATLAB*

CIRTIFICATIONS

- **Artificial Intelligence, Machine Learning, and Deep Learning**, DeepLearning. AI/Coursera   2023
- **Fine Tune BERT for Text Classification with TensorFlow**, DeepLearning. AI/Coursera   2023
- **Database Design and PostgreSQL**, University of Michigan/Coursera   2023
- **Practical Data Science on the AWS Cloud,** AWS/Coursera   2023
- **Biostatistics,** Peking University   2020

SKILLS AND STRENGTHS

- **Software/Language**: Python, R, SAS, MySQL, PostgreSQL, HTML, CSS, ArcGIS
- **Machine Learning**: Deep learning, NLP, BERT, Transformer, ChatGPT API, DALLE2 API, GPT-4 API
- **Data Analysis & Visualization**: Statistical Inference (A/B testing), Tableau, PowerBI, R Shiny, DASH
- **Search Platform & Data Management:** Elasticsearch, AWS/GCP, Docker, Git

WORK EXPERIENCE

Postdoctoral Research Fellow   06/2022-Present
National Institutes of Health --- NIA&NIEHS   Baltimore, MD & Durham, NC

- **Fine tuning BERT (transformer) for interview text classification** with **TensorFlow/Pytorch** and **TensorFlow Hub (Similar to HuggingFace)** to predict and classify mental status.
- Conducting **medical imaging analysis** by using **deep learning** methods such as ConvNet, max pooling, and filters.
- Developing **web-based platforms** to predict aging-related health outcomes via R shiny and DASH.
- Deploying **ChatGPT/GPT-4 API** into the online platforms to help explain the potential biological mechanisms.
- Building **machine learning models** (**Scikit-Learn**) including Catboost, XGboost, Multi-layer Perceptron, Explainable boost machine, Random Forest, and Bayesian models using **Docker** and parallel computing via Linux environment.
- Develop web-based platforms to predict aging-related health outcomes via **R shiny and Python DASH**.
- Regularly using **Git and Github** to manage multiple projects and version control and collaborating with multiple Clinical Research Organizations (Labcorp and DLH Corporation).
- Conducting **time-to-event** analysis such as **Quantile-Based g-Computation, log binomial regression, and COX model in longitudinal studies.**
- Using Shapley value, Boruta selection methods, Bootstrapping AUC, and Selection above Random methods improved predictive performance by 10% percent.
- Manage weekly reports, present at international conferences, and give invited talks in NIH.

Software Environment:
Python, Google Colab (GCP), AWS, Biowulf, PostgreSQL, Docker, Git, R, R-shiny, DASH, Tableau

Research Assistant & Data Manager   07/2019 – 06/2022
The Research Foundation of The State University of New York   Albany, NY
- Used large **CT scan data** and built **Deep Learning models** (**ResNet50**) via **TensorFlow** within Feature Pyramid Network, **VGG-16 (CONV), and Xception** to predict COVID-19.
- Built **pipelines** to transform **medical history text data** into digital text data and made it ready for text mining.
- Leveraged **NLTK, spaCy, word2vec, and GENSIM** to carry out **text mining** such as **tokenization, lemmatization, stemming, parsing, and embeddings.**
- Conducted sentiment analysis (**text classification**) with deep learning using **BERT model via PyTorch.**
- Managed and leveraged 20-year **NY-wide medical claims and electronic hospitalization data** (10 TB), NY utility & service data (2 TB), and nation-wide meteorological data (2 TB), and COVID-19 data (5 TB) in **over 20 projects**.
- Used **MySQL and AWS** to manage database and **Git and Github** to manage multiple projects.
- Developed an **innovative** combined model, **the two-stage downscaling model** for exposure simulations.
- Conducted statistical analysis including **conditional logistic regression, Quantile-Based g-Computation, log binomial regression, COX model, GEE model, and Bayesian spatial-temporal models with INLA** inference.
- Managed the **collaborations** among New York University, Emory University, Sun Yat-sen University, University of Rochester, Environmental Protection Agency, NIH, New York State Energy Research and Development Authority, and New York Public Service Department.
- Collaborated with researchers from China, used the **largest prospective birth cohort** in China to predict congenital heart diseases, and developed an **online predictive tool** to help patients (patent pending).
- Delivered the significant scientific findings and results to policymakers and managed the funding reports to NIH.

Software Environment:
Python, Google Colab (GCP), MySQL, Git, R, R-shiny, SAS, ArcGIS

Research Intern   09/2018 – 06/2019
Sun Yat-sen University Cancer Center
- Used genomics data to build **machine learning models** to predict nasopharyngeal carcinoma.
- Trained **image classification model** (CNN) by using MRI images to identify nasopharyngeal lesions.
- Handling **high-dimensional datasets** from detailed image reports to predict the progress of nasopharyngeal carcinoma.
- Led **nationwide time series clustering** study on the epidemiological characteristics of 41 notifiable **infectious diseases** from 2004-2016.
- Developed an APS (age, platelets, 2D shear-wave elastography) score using COX model to predict hepatocellular carcinoma in chronic hepatitis B retrospective cohort study.
- Served as a consultant of epidemiology and biostatistics for 12 clinical and epidemiology studies.

Software Environment:
Python, Docker, Git, R, R-shiny, SAS

Doctoral Intern   07/2015 – 01/2019
Third Affiliated Hospital, Sun Yat-sen University   Guangzhou, China
- Identified nasopharyngeal, hepatic, cerebral lesions on CT, MRI, and ultrasound.
- Studied CT, MRI, and ultrasound from renowned medical imaging physicians in the hospital's medical imaging department.
- Used open-source imaging tools including 3D Slicer to explore the brain structure.
- Used PACS software for medical radiology to access to patient folder and readily accessible.
- Led a clinical trial Phase II among 40 participants to explore the effectiveness of a new drug.

Software Environment:
Python, 3D Slicer, Git, R, SAS

SOFTWARE PACKAGES, TOOLS, & WEBPAGES

Developed R Package
- **Deng X**, Zhang W, Lin S. Package "APML" An Approach for Machine-Learning Modelling. DOI: 10.13140/RG.2.2.27638.42563. Published online 2020.

Developed Online Tools/Webpages (R-shiny):
- Predictive Model for Enteral Nutrition Intolerance in ICU Patients with Sepsis: xdeng3.shinyapps.io/NIPM/
- Predictive Model for Congenital Heart Diseases Built by a Large Birth Cohort Study: xdeng3.shinyapps.io/CHD_model/
- Predictive Model for COVID-19 vs Community-Acquired Pneumonia (CAP): xdeng3.shinyapps.io/COVID-19/

ACADEMIC SERVICES & HONORS

- Serving as **reviewer** for Environmental International, International Journal of Cancer, Health and Place, Indoor Air, The Journal of Nutrition, Environmental Pollution, Journal of Exposure Science and Environmental Epidemiology, Environmental Health, Journal of Environmental Management, Frontiers in Public Health, JMIR Public Health and Surveillance, Head and Neck, Journal of BioMed Research International, Hygiene and Environmental Health Advances, and Plos One
- Served as **Guest Editor** for *Plos One*
- Served as abstract **reviewer** for 2022-2023 APHA conference
- Distinguished Doctoral Dissertation Award 2022-2023
- Registration Award at ISEE conference 2021, 2023
- Served as moderator at the American Public Health Association (APHA) conference for **Big Data Analytics for Public Health Research session** 2021
- **Applied Public Health Statistics** (APHS) Student Membership Award from APHA 2021
- Health Research, Inc. David Axelrod Award for Outstanding Presentation in UAlbany 2021
- Student and New Researcher Network (SNRN) Award at ISEE conference 2020
- Outstanding Thesis Recipient in Sun Yat-sen University 2019
- **Honorable Mention Award** in American **Mathematical Contest in Modeling** 2018
- **First prize** at the provincial level in China Contemporary **Mathematical Contest in Modeling** 2017

SELECTED PUBLICATIONS

- **Deng X**, Thurston G, Zhang W, et al. Application of data science methods to identify school and home risk factors for asthma and allergy-related symptoms among children in New York. Sci Total Environ. 2021;770. doi:10.1016/J.SCITOTENV.2020.144746.
- **Deng X**, Li H, Liao X, et al. Building a predictive model to identify clinical indicators for COVID-19 using machine learning method. Med Biol Eng Comput. 2022. doi:10.1007/s11517-022-02568-2.
- Qu Y*, **Deng X***, Lin S, et al. Using Innovative Machine Learning Methods to Screen and Identify Predictors of Congenital Heart Diseases. Frontiers in Cardiovascular Medicine. 2022;8. doi:10.3389/fcvm.2021.797002.
- Hu K, **Deng X**, et al. Development and validation of a predictive model for feeding intolerance in intensive care unit patients with sepsis. Saudi J Gastroenterol Off J Saudi Gastroenterol Assoc. Published online 2021.
- Luo W, **Deng X**, Xu X, et al. Development of a Prognostic Model for Predicting Multiple Sclerosis After Optic Neuritis: A Secondary Analysis of Data From the Optic Neuritis Treatment Trial. Journal of Neuro-Ophthalmology. 2021. doi: 10.1097/WNO.0000000000001424
- GBD 2021 Diabetes Collaborators. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. The Lancet. Published online June 1, 2023. doi: 10.1016/s0140-6736(23)01301-6.

*Co-first author