Doan Bui

(765) 301-1005 | doanbui1607@gmail.com | github.com/doanbui1607

Background Education

The University of Chicago, Chicago, Illinois   **Expected June 2024**

Master of Science in Biomedical Science, Data Science Concentration   GPA: 3.73

Relevant coursework: Genome Analysis, Machine Learning, Deep Learning, Healthcare Data Science

DePauw University, Greencastle, Indiana   **Class of 2023**

Bachelor of Art in Biochemistry major, Data Science and Applied Statistics minor

Graduated Cum Laude

Academic Distinctions: DePauw University Dean's List, Merit Scholarship

Relevant courses: Data Structure, Bioinformatics, Applied Statistics, Calculus, Statistical Model Analysis

---

Technical Skills

Skills: R, Python, SQL, Linux, Tableau, Microsoft Office.

Libraries/Frameworks: Numpy, Pandas, Matplotlib, Seaborn, Scikit-learn, PyTorch, TensorFlow, statsmodels, Jupyter Notebook, tidyverse, ggplot, Rshiny, Bioconducter, DESeq2, maser, STAR, rMATS, Seurat, RMarkdown

Data Science pipeline: data cleaning, wrangling, exploration, visualization, modelling, interpretation, hypothesis testing, statistics

---

Research Experience

Biological Sciences Division, The University of Chicago

Diabetes Type II and Alzheimer's Disease – Single-cell Analysis   **January 2024 – current**

- Perform quality control, data visualization, and publication-quality graphics.
- Update the research team on new data analysis methods and relevant publications.
- Conduct DE analysis using Seurat and R packages on different T2D cell types.
- Collaborate closely with postdoctoral researchers and professors to provide comprehensive research insights.

Applying Machine Learning Model in Alzheimer Disease Prevention

- Conducted data mining, data cleaning and wrangling to ensure data quality and usability.
- Developed and optimized supervised machine learning models including tree-based and logistic regression on 400+ MRI images.
- Achieved over 80% AD prediction accuracy using feature engineering and hyperparameter tuning.

Gastric Cancer Prediction Analysis

- Leveraged NIH All of Us research program data of over 400,000 patients with malignant tumor in stomach.
- Investigated correlations between gastric cancer and social factors, lifestyle habits, and biological factors.

Stroke Prediction Analysis   **September – December 2023**

- Processed, cleaned, and explored NHANES data of over 5000 patients with both biological and social factors.
- Optimized models and identified 3 main predictors for stroke prediction from multiple regressions analysis.
- Generated statistical figures, interpreted findings, and revealed complex social impacts on stroke chances.

The RNA Institute, University at Albany   **June – August 2023**

ZIKA-virus infection on human progenitor cells

- Conducted FastQC test for over 130,000 RNA-Seq samples to ensure data quality and usability.
- Performed gene alignment, DE and alternative splicing analysis using STAR, DESeq2 and rMATS on Linux.
- Visualized and demonstrated analysis results on RStudio using R packages (maser, ggplot, dplyr)
- Identified 2 different genes with complex associations with ZIKA virus DNA-binding activities.
- Presented weekly progress updates and new cutting-edge ideas, research, and technology.
- Collaborated closely with professors and researchers to delivered project results in annual poster symposium.

DePauw University, Gurnon Lab   **June – August 2022**

Research Assistant

- Regularly presented project updates, new tools, methodologies, and recent publications.
- Evaluated and characterized protein structures of over 20 VUS using AlphaFold2 and ConSurf Server.
- Reported structural prediction improvements for UCSF AlphaFold2 multimer.