XIAOBING YU

Department of Computer Science & Engineering
Department of Genetics
Department of Radiology
Washington University in St. Louis, St. Louis, MO, U.S.
513-306-3721 | xiaobing@wustl.edu
Github: https://github.com/XiaobingDean

EDUCATION

Washington University in St. Louis — St. Louis, MO
Master of Science in Computer Science — Jan. 2021 – Dec. 2022

- GPA: 3.94/4.00
- Visiting Graduate Scholar at Johns Hopkins University (Jun. 2021 – Aug. 2021)
- Selected Courses: Machine Learning, Algorithms for Computational Biology, Applications of Deep Neural Networks, Theory of Artificial Intelligence and Machine Learning, Introduction to Artificial Intelligence, Algorithms for Nonlinear Optimization, Algorithms for Biosequence Comparison.

Miami University — Oxford, OH
Bachelor of Science in Computer Science — Aug. 2017 – Dec. 2020

- Major GPA: 3.46/4.00
- Dean's List (Jan. 2020 – May 2020, Aug. 2019 – Dec. 2019)
- Selected Courses: Object Oriented Programming, Operating System, Database System, Data Structure, Advanced Algorithm, High Performance Computing, Discrete Mathematics, Linear Algebra, Statistics.

PUBLICATIONS

1. Diab NS[#], Barish S[#], Dong W[#], Zhao S[#], Allington G, **Yu X**, Kahle KT, Brueckner M[&], Jin SC[&]. 2021. Molecular Genetics and Complex Inheritance of Congenital Heart Disease. ***Genes***; 12(7):1020. # Co-first authors; & Co-corresponding authors. **Cover story**.
2. Qiu GW[#], **Yu X**[#], Sun B, Zhang L, Wang Y. 2020. Metastatic Cancer Image Classification Based on Deep Learning Method. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering,* ***ICCECE*** *2021*; 658-661. # Co-first authors.
3. **Yu X**, Chen H, Yan R, Stahr M. 2021. Design and Implementation Knowledge Graph for Curriculum System in University. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering,* ***ICCECE*** *2021*; 767-770.
4. Wang, Y.-C.; Wu, Y.; Choi, J.; Allington, G.; Zhao, S.; Khanfar, M.; Yang, K.; Fu, P.-Y.; Wrubel, M.; **Yu, X**.; Mekbib, K.Y.; Ocken, J.; Smith, H.; Shohfi, J.; Kahle, K.T.; Lu, Q.; Jin, S.C. Computational Genomics in the Era of Precision Medicine: Applications to Variant Analysis and Gene Therapy. ***J. Pers. Med***. 2022, 12, 175.
5. Pan X, **Yu X**, Mintz A, Wang J, Mokkarala M, Narra V, Marcus D, Bierhals A, and Sotiras A. A Generative-Discriminative Deep Learning Approach to Classify Radiology Reports based on the Presence of Follow Up Recommendations ***SPIE Medical Image Robert F. Wagner all-conference best paper finalist award***
6. **Pan X**[#], **Yu X**[#], Mintz A, Wang J, Mokkarala M, Narra V, Marcus D, Bierhals A, and Sotiras Large-scale evaluation of machine learning models in identifying follow-up recommendations in radiology reports (*Submitted to* ***JAMIA***) # Co-first authors

PRESENTATION AND POSTER

1. Yu X. 2021. Design and Implementation Knowledge Graph for Curriculum System in University. *2021 IEEE International Conference on Consumer Electronics and Computer Engineering, ICCECE 2021*; 767-770.
2. Yu X. 2022. Mask R-CNN for glomeruli instance segmentation in kidney transplant biopsies. Master Project Oral Defense.

RESEARCH EXPERIENCE

Washington University in St. Louis (Department of Radiology) — St. Louis, MO
Research Assistant to Professor Aristeidis Sotiras — Nov. 2021 – Present

Text Classification on incidental findings on clinical report

- Text classification based on medical incidental findings on radiology report. We will design an algorithm and compare with other popular text classification algorithm.
- We focused on two parts, findings, and impression.
- We have included CNN, RNN, SVM, Fasttext, Paragraph Vector, Hierarchical Attention Networks (HAN), BERT and other popular text classification models on our radiology report, used F1 score as our primary metrics to compare their performance with our own model.
- Submitted to JAMIA.

- We had another 4M+ data from radiology department, and we focused on the same area as the previous research.
- We also included CNN, RNN, SVM, Fasttext, Paragraph Vector, Hierarchical Attention Networks (HAN), BERT with different version, Graph Convolutional Network (GCN) and other popular text classification models on our data, used F1 score and Balanced Accuracy to further evaluate how DL models can perform on more radiology report.

Washington University in St. Louis (Department of Radiology) St. Louis, MO
Research Assistant to Professor Aristeidis Sotiras July. 2022 – Present
Automatic impression generation based on finding for radiology report

- We aimed to develop a new model to generate impression from finding and background part since we believed impression section can be generated by those DL models from text generation research.
- We have investigated several existed methods such as BART and other sequence-to-sequence based model.
- We keep investigating more existed SOTA algorithms and models for radiology report summarization and further refine our own idea.

Johns Hopkins University (Department of Computer Science) Baltimore, MD
Research Assistant to Professor Benjamin Langmead and Dr. Daniel Baker Jun. 2021 – Sep. 2021
Designing new algorithms and machine learning methods for Phip-Seq

- Designed new algorithms and machine learning methods for designing the "library" - the set of peptides that the Phip-Seq assay should use. A major computational question was how to design a library that consisted of a given small number of peptides (constrained by practical costs) that, nonetheless, covered the interesting portions of the target set. A specific goal of this project was to improve the network architecture and optimize the hyperparameters for this specific task.
- Was responsible for designing machine learning and deep learning architecture to fit the spatial information from the sequence. Used CNN 1D as options for model selection since CNN has abilities to process the spatial information which were essential features for the target sequence.
- Discovered a reasonable approach to encode the sequence which contained as much information as possible.
- Learned Bepipred-2.0 and generated results from this machine learning approach.

Washington University in St. Louis (Department of Computer Science) St. Louis, MO
Research Assistant to Professor S. Joshua Swamidass Jan. 2022 – Aug. 2022
Deep learning networks for improving glomeruli instance segmentation in kidney transplant biopsies

- I aim to solve the problem that pixel-wise semantic segmentation currently works well, but instance segmentation of individual glomeruli is currently performed with conventional image processing (blob detection), which occasionally leads to erroneous grouping of close-set clusters of glomeruli.
- Our goal is to develop a deep learning approach for processing kidney biopsy semantic segmentation images to improve glomeruli instance segmentation.
- I have achieved some improvements by using Mask R-CNN with different optimization, and I am trying to give full analysis on the results to identify the strengths and weakness of this approach.
- I have passed the oral defense and written a report about my result and conclusion.
- From our previous research, we found the common object detection model and Non-Maximum Suppression (NMS) is not suitable for our target, which is calculating the global glomerulosclerosis.
- We are attempting to improve these two parts by using more relative model such as CircleNet, and further develop NMS by replacing different IOU method.

Miami University (Department of Computer Science) Oxford, OH
Research Assistant to Professor Michael Stahr Jul. 2020 – Dec. 2020
OpenVaccine: COVID-19 mRNA vaccine degradation prediction (Kaggle)

- Used GRU/LSTM layer to process these data and merge the result using GNN v1 and GNN v2 to obtain the result, since it contained all features from a regular sequence.
- Pre-processed and tokenized the sequence, secondary structure, and loop type based on its physical and biological features. Used all the information to train a model on degradations recorded by the OpenVaccine researchers.

Miami University (Department of Computer Science) Oxford, OH
Research Assistant to Professor Michael Stahr Jan. 2020 – Sep. 2020
Metastatic cancer image classification based on deep learning methods

- Proposed a novel method that combined the deep learning algorithm in image classification, the DenseNet169 framework, and the Rectified Adam optimization algorithm. The connectivity pattern of DenseNet was direct connections from any layer to all consecutive layers, which could effectively improve the information flow between different layers.
- Investigated all state-of-the-art models on Computer Vision and got familiar with how to use OpenCV.
- Learned how to do model selection and hyperparameter searching on different scenarios.

SELECTED AWARDS AND HONORS

- Dean's List 2019, 2020
- Second Prize on Kaggle Deep Learning Challenge (2%); Ranking: Top 28 2020

WORK EXPERIENCE

Tencent Shenzhen, China
NLP Internship Jun. 2020 – Jul. 2020

- Learned and practiced the architecture design of several NLP algorithms for solving sequence problems, which included CNN, RNN, LSTM and BERT.
- Evaluated the advantages and disadvantages among several text classification models by referencing recall, precision, and F-score, and wrote a report on performance comparison reports.

Teaching Assistant
Assisted with the following units:

- CSE 417T: Introduction to Machine Learning (WashU) Aug. 2021 – Dec. 2021; Jan. 2022 – May. 2022
- CSE 517T: Machine Learning (WashU) Jan. 2022 – May. 2022
- CSE 566S: High Performance Computer Systems (WashU) Jan. 2021 – May 2021
- CSE271: Object Oriented Programming (MU) Aug. 2019 – Dec. 2019, Aug. 2020 – Dec. 2020
- CSE274: Data Structure (MU) Aug. 2020 – Dec. 2020
- CSE385: Database System (MU) Aug. 2019 – Dec. 2019, Jan. 2020 – May 2020
- CSE381: Operating System (MU) Jan. 2020 – May 2020

Technology Consult (MU)

- Assisted with Senior Design Aug. 2020 – Nov. 2020

ADDITIONAL INFORMATION

Additional Professional and Extracurricular Experiences

- Professional Gamer (May 2016 – Mar. 2017)

Interests

- Coding (Kaggle competition)
- Basketball

Computer and Language Skills

- Software (Genetic related): GATK, Hail, Annovar, VEP, Plink, Plots Read, TADA-R.
- Software & Libraries: MySQL, MongoDB, Visual Studio, Google Cloud Platform, Linux System, NetBeans, Eclipse, Jupyter Notebook, TensorFlow, PyTorch, OpenCV, BioBERT, AWS, NumPy, Pandas, Keras, scikit-learn etc.
- Programming Skills: C++ (Advanced), Java (Advanced), Python (Advanced), SQL (Advanced), R (Advanced), Shell Script, C, C#, Html, PHP, JavaScript, Answer Set Prolog, Scheme, Perl, Matlab.
- Language: English, Chinese, Japanese.

REFERENCES

Prof. Aristeidis Sotiras aristeidis.sotiras@wustl.edu
Prof. S. Joshua Swamidass swamidass@wustl.edu 314-362-7424
Dr. Daniel Baker dbaker@pacificbiosciences.com 765-702-8477