

# CHUFAN GAO

219-239-8008 ◇ [gaoandy1445@gmail.com](mailto:gaoandy1445@gmail.com) ◇ Pittsburgh, PA, United States  
LinkedIn: [chufangao](#) ◇ Github: [chufangao](#) ◇ Google Scholar: [rBIZICgAAAAJ](#)

**Research Statement:** I am currently a staff researcher at Carnegie Mellon University with Professor Artur Dubrawski in the Auton Lab. My areas of focus include Weak Supervision, Unsupervised Feature Representations, and Machine Learning for Healthcare in general. I am also broadly interested in sequential data and Generative Adversarial Networks.

## EDUCATION

### Masters of Science in Robotics *Carnegie Mellon University*

August 2020 – August 2022  
*Pittsburgh, PA*

- School of Computer Science, Research Masters (GPA Fall 2021: 3.89)
- Relevant Courses - Math Fundamentals for Robotics, Computer Vision, Probabilistic Graphical Models, Machine Learning, Convex optimization, Statistical Techniques in Robotics Kinematics Dynamic Systems and Control

### Bachelor of Science *Purdue University*

August 2016 – May 2019  
*West Lafayette, IN*

- Majors: Computer Science, Mathematical Statistics **with Honors and Distinction** (GPA: 3.88)
- **CS/ML** courses (*\* indicates graduate level*) – Machine Learning\*, Algorithms\*, AI\*, Graphical Models, Data Structures and Algorithms
- **Math/Statistics** courses – Advanced Linear Algebra, Differential Equations, Real Analysis, Probability\*, Statistical Theory\*, Vector Calculus

## RESEARCH

### Research Analyst *Carnegie Mellon University Robotics Institute*

August 2019 – Present  
*Pittsburgh, PA*

- Conducted various research projects in partnership with University of Pittsburgh Medical Center including:
- Investigating methods to augment text classification with weak supervision, in submission to AAAI
- Augmenting weak supervision with active learning to improve labelling function parameters in an interpretable manner, in submission to ICLR
- A total of 2 accepted student abstracts in AAAI Student Track
- Publishing 1 medical abstract as well as 1 more pending in American Thoracic Society

### Robotics Institute Summer Scholars (RISS) *Carnegie Mellon University Robotics Institute*

June 2019 – August 2019  
*Pittsburgh, PA*

- [Robotics Institute Summer Scholar \(RISS\) program](#) (2-3% acceptance rate) - Investigated methods into detecting physiological state changes via deep unsupervised learning mentored by Professor [Artur Dubrawski](#)
- Methods include a custom Pytorch implementation of dilated CNNs for sequence embedding and autoencoders with attention
- Resulted in acceptance to NeurIPS ML4H workshop as well as a staff research position

### Visiting Undergraduate Researcher *DePaul University College of Computing and Digital Media*

June 2018 – August 2018  
*Chicago, IL*

- [Medix REU Program](#) (<10% acceptance rate) - Implemented a custom 3D Generative Adversarial Networks and 3D CNN to improve performance of Computer-Aided Detection systems under [Professor Jacob Furst](#) and [Professor Daniela Raicu](#)
- Resulted in oral presentation and publication of *Augmenting LIDC dataset using 3D generative adversarial networks to improve lung nodule detection* in SPIE Medical Imaging conference

### Undergraduate Researcher *Purdue University*

May 2017 – May 2018  
*West Lafayette, IN*

- Student Learning Research - Conducted statistical analysis on the effects of active learning classes on future student performance Mentored by [Professor Clarence Maybee](#). Presented *Out of the Box: Impact of Active Learning on Future Student Performance* via invited oral presentation in Purdue Journal of Undergraduate Research.

- Deep Health Metric Prediction - Acquired experimental data, automated preprocessing workflow, and wrote a program to determine health metrics such as Heart Rate, HRV, and PPG from videos using DeepFace mentored by [Professor Vaneet Aggarwal](#)

## ACADEMIC PROJECTS

---

**PCALG in Python:** Read respective papers on 4 Causal learning algorithms (PC, FCI, GES, LINGAM) and ported over R and C++ code from [pcalg](#) (R Causal Inference Package) to Python.

**Weak Text Classification:** Combined Data Programming (Snorkel) and glove word similarities to extract useful keywords for imdb reviews text classification. Resulted in a technical report: The word is mightier than the label: Learning without pointillistic labels using data programming”.

**Time Series Anomaly Detection:** Extended an existing AAAI paper’s graph neural network to account for uncertainty in estimates. Resulted in a technical report: Learning graph neural networks for multivariate time series anomaly detection”.

### Class Projects

- **Web-crawler** capable of extracting links and indexing over 100,000 websites given a seed website. Supported keyword searching as well.
- **Terminal shell** in C,C++ with working regex, history, typing, application execution capabilities
- **Online question answering platform** with custom HTML, JavaScript, CSS, JQuery, Bootstrap, Django Backend. [youtu.be/0frgr0mw\\_tg](https://youtu.be/0frgr0mw_tg)
- **Custom javascript for generation of arrows** in [causality101.net](https://causality101.net)

## TEACHING AND MENTORSHIP

---

### AI4ALL Mentor

*Carnegie Mellon University*

Summer 2021

*Pittsburgh, PA*

- Advised 5 high school students one-on-one over a 2-week time period to use transfer learning and DenseNet to achieve over 90% accuracy in plant disease classification.
- Created and tested project template code in Google Collaboratory, ensuring that memory usage was as low as possible, for the above task.

### Purdue Teaching Assistant

*Purdue University*

January 2017 – May 2017

*West Lafayette, IN*

- Coordinated CS177 Python lab activities and review sessions. Graded and reviewed labs, projects, and tests.

### Computer Science Peer2Peer Mentor

*Purdue University*

August 2017 – December 2017

*West Lafayette, IN*

- Mentored and guided incoming computer science freshmen. Set up weekly meetings, informal lunch discussions, and various other events to support mentees.

## HONORS AND AWARDS

---

### Scholarships and Academic Awards

Boeing Scholarship - disbursed based on academic merit in CS. 5/1900 CS students at Purdue.

- Purdue Presidential Scholarship - disbursed based on high academic achievement; leadership and service in school/community. 830/40,000 students at Purdue.
- Gordan L. Walker Scholarship - disbursed based on continuing academic achievement in mathematics.
- Purdue West Lafayette Deans List (all years), Honors College Member (all years). 1 out of all Math students at Purdue.

### DJI Drone Challenge

Summer 2019

- Led a team of 5 in a drone challenge following a path specified by aruco tags autonomously. Implemented functionality of viewing AR holographic images through the drone camera. Created android app to switch between drone modes.
- 1st place out of 8 teams and 40 competitors

### UBTech Humanoid Challenge

Summer 2019

- Led a team of 6 in programming and teleoperating a humanoid robot in ROSpy with a raspberry pi that could effectively grasp and move a small object.
- 1st place out of 6 teams and 40 competitors.

## TECHNICAL STRENGTHS

---

**Programming Languages** – Python, R, Java, C, C++, Javascript

**Frameworks / Tools** – Keras, Tensorflow, Pytorch, Sklearn

## ADDITIONAL PROJECTS AND SERVICE

---

### Committees

- Carnegie Mellon University Robotics Institute Summer Scholars (RISS) Admissions Committee (2020-Present): Reviewed applicants on quality of fit to RISS. Produced forms and documentation that was used to streamline application process.

### Reviewer Duties

- International Conference on Learning Representations (ICLR), 2022
- NeurIPS, 2019-2021
- ACM Conference on Health, Inference, and Learning (CHIL), 2020

### Food Pantry Volunteer

*ACE Campus Food Pantry*

January 2019 - May 2019

*West Lafayette, IN*

- Facilitated the operations (checkout and stocking) of ACE Campus Food Pantry in its efforts to contribute to greater food security for the Purdue Community

## PUBLICATIONS

---

- [1] J. H. Yoon, C. Gao, J. Kim, J. H. Kim, T. Lagattuta, S. Helman, M. Hravnak, M. R. Pinsky, and G. Clermont, "Prediction of hypovolemic instability in normal volunteer blood donors using machine learning," in *TP55. TP055 Mechanical Ventilation, ICU Management, and CV*, American Thoracic Society, 2022.
- [2] C. Gao, A. Dubrawski, M. Pinsky, G. Clermont, and J. Yoon, "Identification and explanation of severity of bleeding-induced hypovolemia using unsupervised deep learning," in *TP55. TP055 Mechanical Ventilation, ICU Management, and CV*, pp. A2841–A2841, American Thoracic Society, 2021.
- [3] M. Goswami, L. Chen, C. Gao, and A. Dubrawski, "Modeling involuntary dynamic behaviors to support intelligent tutoring (student abstract)," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 13799–13800, 2020.
- [4] S. Peng, L. Chen, C. Gao, and R. J. Tong, "Predicting students' attention level with interpretable facial and head dynamic features in an online tutoring system (student abstract)," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 13895–13896, 2020.
- [5] C. Gao, F. Falck, M. Goswami, A. Wertz, M. R. Pinsky, and A. Dubrawski, "Detecting patterns of physiological response to hemodynamic stress via unsupervised deep learning," in *NeurIPS 2019 ML4H Workshop*, 2019.
- [6] C. Gao, S. Clark, J. Furst, and D. Raicu, "Augmenting lidc dataset using 3d generative adversarial networks to improve lung nodule detection," in *Medical Imaging 2019: Computer-Aided Diagnosis*, vol. 10950, p. 109501K, International Society for Optics and Photonics, 2019.
- [7] C. Gao, "Out of the box: Impact of active learning on future student performance," *The Journal of Purdue Undergraduate Research*, vol. 8, no. 1, p. 38, 2018.

### Technical Reports (\* denotes Equal Contribution)

1. S. Ray\*, S. Lakdawala\*, M. Goswami\*, and C. Gao\*, "Learning graph neural networks for multivariate time series anomaly detection" arXiv preprint arXiv:2111.08082, 2022
2. C. Gao\* and M. Goswami\*, "The word is mightier than the label: Learning without pointillistic labels using data programming" arXiv preprint arXiv:2108.10921, 2021

### Preprints Submitted and Under Consideration

1. C. Gao\*, M. Goswami\*, J. Chen, A. Dubrawski, "Classifying Unstructured Clinical Notes via Automatic Weak Supervision" In Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL), 2022

### Preprints To Be Submitted

1. M. Goswami\*, C. Gao\*, B. Boecking, A. Dubrawski, "Active Learning for Weakly Supervised Model Refinement" In Uncertainty in Artificial Intelligence (UAI), 2022