Linhai Ma

Address: 9921 S.W 97th CT, Miami, FL, USA
Phone: +1-9198840535 Email: rinnkai.ba2019@gmail.com
Linkedin: https://www.linkedin.com/in/linhai-ma-6441861bb
Personal Page: https://sarielma.github.io/

EDUCATION

University of Miami, Department of Computer Science, Coral Gables, FL, USA. **08/2017-now**

- Doctor of Philosophy in Computer Science, specializing in Machine Learning and Medical Image Analysis (GPA:3.85/4.0)
- Related courses: Data Mining, Mathematical Statistics, Automated Reasoning, Statistical Learning, etc.

Institute of Software, Chinese Academy of Sciences, State Key Laboratory of Computer Science (ISCAS, SKLCS), Beijing, China **09/2014-07/2017**

- Master of Science in Computer Science, specializing in Software Testing and Concurrency (GPA: Top 10%)

Northeastern University, Software College, Shenyang, China **09/2010-06/2014**

- Bachelor of Science in Information Security (GPA: Top 5%)
- Related courses: C\C++, Java, Linux, Data Structure, Computer Architecture, Operating System, Computer Network, Advanced Mathematics, Linear Algebra, Probability and Statistics, Discrete Mathematics, etc.

PROGRAMMING LANGUAGES

- Python (including Pytorch, Pandas, Scikit-learn, Numpy, etc.), Java, C/C++, C#, Ruby, R, SQL, Prolog, etc.

INDUSTRY EXPERIENCE

Intern Software Engineer, Cadence Design Systems, Inc., San Jose, CA **05/2022-08/2022**

- Designed a machine learning system based on **Bayesian Optimization** to improve the integrated circuit simulation process, ideally reducing the circuit simulation time consumption by >90%.
- Built up the whole machine learning system, including data pipeline, prediction model, and training and evaluation module via **Python, Pytorch, Scikit-learn, Pandas, etc**., by which extensive experiments and testing were conducted.
- Conducted experiments on the machine learning system with different **kernel functions, data preprocessing, training batch size**, etc., to solve problems such as outliers, categorical inputs, early termination, etc.
- Visualized the experiments and results via **Python**, **Matplotlib, Seaborn**, etc., and had a presentation to show the potential benefit of this work to the company.

Intern Software Engineer, Cadence Design Systems, Inc., San Jose, CA **06/2021-08/2021**

- Designed a machine learning system based on **Random Forest** and **Gradient Boost Decision Tree,** aiming to predict the unfinished integrated circuit simulation results (measurements) based on the previous inputs (corners) and results (measurements), reaching a mean absolute error of 0.99 picoseconds (good result).
- Implemented the prototype of the machine learning system, including data preprocessing, fine-tuned prediction model, and evaluation module via **Python, Scikit-learn, Pandas, etc**., by which extensive experiments and testing were conducted.
- Developed a user interface via **PyQT5** and **matplotlib** to visualize the relationship between input variables and the prediction measures, which was used to select input features to improve the system.

Intern Software Engineer, Hillstone Networks, Beijing, China **06/2015-11/2015**

- Modified and Migrated an Intrusion Prevention System (IPS) server to a new server machine via **Ruby, Ruby on Rails, etc.**
- Developed a packet sniffer via **Rudy and C\C++** to periodically capture packets from the server.
- Developed an automatic **SQLite** manipulator via **Python, SQL, and C\C++** to automatically manipulate the SQLite database in the backend according to client requests.
- Developed HTTP URI filter in the firewall machine via **C\C++** to filter out abnormal requests**.**

Intern Software Engineer, **Huaxin Education Technology Co., Ltd.**, Shenyang, China **06/2013-08/2013**

- Developed a Vulnerability Scanner for web applications via **Java, JSP, Oracle, etc.,** to discover the vulnerability of the web pages against attacks, such as SQL Injection, Cross-Site Scripting vulnerabilities.
- Developed the user interface via **Java Swing** for a better user experience.

RESEARCH EXPERIENCE

Research Assistant, University of Miami, Department of Computer Science, FL **09/2019-now**

Project:

- Proposed methods to improve the following deep learning applications for better robustness:
  - **Unet**-based model (nnUnet) for Heart, Hippocampus and Prostate MRI images segmentation.
  - **Multi-task** Unet-based model for cephalometric landmark detection.
  - **YOLO** V5 for blood cell detection.
  - **Transformer**-based model (TransUnet) for abdominal organ segmentation.
- The experiments were conducted via **Python, Pytorch, Pandas, Scikit-learn, etc.,** on Tesla V100 GPU**.**
- The improved models have better accuracy under testing noises than the baseline models.

Project:

- Proposed a training method via increasing margin ideas to improve the neural networks for better robustness and gave the theoretical proof of the training method's convergence.
- Evaluated the proposed method with residual net classifiers on public image datasets, e.g., CIFAR10, TinyImageNet, etc., and real-world medical images, e.g., Covid-19 CT.
- The experiments were conducted via **Python, Pytorch, Pandas, Scikit-learn, etc.,** on Tesla V100 GPU**.**
- The proposed method significantly improved the model's robustness against adversarial noises with minimal accuracy degradation.

Project:

- Designed and implemented a CNN model for a variant-length 12-lead electrocardiogram classification task (from the China Physiological Signal Challenge 2018) via **Python, Pytorch, Pandas, Scikit-learn, etc.**, which achieved top-6 performance in the challenge.
- Designed and implemented a multiple-layer perceptron (MLP) model for electrocardiogram data classification (from PhysionNet's MIT-BIH dataset) via **Python, Pytorch, Pandas, Scikit-learn, etc.,** whose prediction accuracy is 92%.
- Proposed a regularization method to improve the deep-learning-based electrocardiogram automatic diagnosis for better robustness. The experiments show that the proposed method improved the model's robustness and outperformed other competitive methods in the paper.

Research Assistant, University of Miami, Department of Computer Science, FL and Northwestern University, Kellogg School of Management, IL **01/2018-09/2019**

- Parsed author and publication data (from 1991 to 2018) from Microsoft Graph database via **SQL and Python** to build up one author-to-author graph (millions by millions matrix) for each year.
- Designed a clustering machine learning algorithm based on the Monte Carlo method to cluster the graph into author groups. This algorithm is implemented via **Python and R**.
- Defined these author groups' evolution patterns (splitting, merging, etc.) in two adjacent years and collected these patterns via **Python** to show how each author group evolves over the years.
- Discovered the mathematical relationship between each group's evolution patterns and the group's success (e.g., the number of citations) via **Python and R** to get the conclusion, that is, what evolution pattern is more likely to contribute to a successful group.

Research Assistant, State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, Beijing, China **07/2015-06/2017**

- Designed and developed a C++ generator via **Python and C\C++** to automatically parse the C++ class (the concurrent data structure implemented with the pthread library) and to generate multi-thread test cases that invoke this C++ class.
- Proposed three adaptive algorithms to improve the C++ test case generation, which discovered up to 6% more potential concurrent errors and reduced the time cost by up to 10%.

COURSE PROJECT

Project*:* Sentiment analysis on review texts from Amazon

Analyzed the potential sentiment inside the review texts and predicted how buyers felt about the things they bought from their review texts.

- Encoded the review texts via **Vector Space Model (VSM)** with **Tf-idf** and **Latent Dirichlet Allocation (LDA)** into feature vectors via **Python, Java, MALLET, WEKA, and Windows Batch Script**.
- Used **SVM** and **Naïve Bayes** via **Scikit-learn** to classify the review texts into five categories, corresponding Star 1 to Star 5, to show their potential sentiment inside the review texts.
- The experiment shows that the performance of LDA is as good as VSM in this task.

Reviewer Experience:

- 1 paper in International Conference on Machine Learning 2022 (ICML2022) – AI top conference
- 3 papers in Neural Information Processing Systems 2022 (NeurIPS 2022) – AI top conference
- 1 paper in Scientific Programming – A peer-reviewed journal in software engineering
- 1 paper in IEEE Journal of Biomedical and Health Informatics– A peer-reviewed journal in biomedical informatics
- 4 papers in Expert Systems with Applications (ESWA) – A peer-reviewed journal in intelligent systems applications
- 1 paper in Artificial Intelligence (AI)– A famous peer-reviewed journal in AI
- 1 paper in Computers in Biology and Medicine (CIBM) – A peer-reviewed journal in biomedical informatics
- 2 papers in Computer Systems Science and Engineering – A peer-reviewed journal in computer systems science

---

PUBLICATIONS:

Linhai Ma, Liang Liang, Towards lifting the trade-off between accuracy and adversarial robustness of deep neural networks for medical image classification and segmentation." Accepted by SPIE Medical Imaging 2023.

Linhai Ma, Liang Liang, " Improving Adversarial Robustness of Deep Neural Networks via Adaptive Margin Evolution." Under review.

Linhai Ma, Liang Liang, "Adaptive Adversarial Training to Improve Adversarial Robustness of DNNs for Medical Image Segmentation and Detection." https://arxiv.org/abs/2206.01736. Under review.

Linhai Ma, Liang Liang. "Increasing-Margin Adversarial (IMA) Training to Improve Adversarial Robustness of Neural Networks." https://arxiv.org/abs/2005.09147. Under review.

Liang Liang, Linhai Ma, Linchen Qian, Jiasong Chen. " An Algorithm for Out-Of-Distribution Attack to Neural Network Encoder." https://arxiv.org/abs/2009.08016. Under review.

Linhai Ma, Liang Liang. "A Regularization Method to Improve Adversarial Robustness of Neural Networks for ECG Signal Classification." Computers in Biology and Medicine 144 (2022): 105345.

Linhai Ma, Liang Liang. " Enhance CNN Robustness Against Noises for Classification of 12-Lead ECG with Variable Length." 19th IEEE international conference on machine learning and applications (ICMLA 2020).

Linhai Ma, Liang Liang. "Improve robustness of DNN for ECG signal classification: a noise-to-signal ratio perspective." International Conference on Learning Representations (ICLR 2020) Workshop AI for Affordable Health.

Linhai Ma, Peng Wu, Tsong Yueh Chen. "Diversity driven adaptive test generation for concurrent data structures." Information and Software Technology 103 (2018): 162-173.