Hudanyun Sheng

hudanyun.sheng@outlook.com | (+86) 13683132915 | https://danniesheng.github.io/

Data Scientist with strong research-to-industry transition experience across healthcare, AI/ML systems, and computer vision. Proven ability to build and deploy ML solutions from raw data processing to model optimization, with strong cross-functional communication in both English and Chinese, with professional experience in both U.S. and China.

EDUCATION

University of Florida - *M.S. in Electrical and Computer Engineering (GPA: 3.86/4), December 2019*

Master thesis: Switchgrass Genotype Classification using Hyperspectral Imagery

University of Florida - *M.S. in Industrial and Systems Engineering (GPA: 3.87/4), December 2017*

Tongji University - *B.S. in Physics (GPA: 4.45/5), June 2015*

Bachelor thesis: The Correction of the Intensity Unevenness of X-Ray KB Imaging

TECHNICAL SKILLS

- Programming & Development: Python (NumPy, Pandas, SciPy, Streamlit, Plotly), Flask, HTML/CSS/JavaScript, MATLAB
- Machine Learning & Deep Learning: PyTorch, TensorFlow, Scikit-learn, Keras, Hugging Face, OpenCV
- NLP & Generative AI: ChatGPT API, LangChain, RAG, Weights & Biases, NLP pipeline design
- Data Engineering & Platforms: SQL, PySpark, BigQuery, Kedro, Data Pipeline Development, EHR Data Analysis
- Cloud & DevOps: GCP, AWS, Azure, Docker, Kubernetes, Git

PROFESSIONAL EXPERIENCE

Johnson & Johnson | **Data Scientist (Contractor),** Beijing, China — April 2024-present

Center of Excellence

- Built machine learning models to identify early adopters and prescribers using structured HCO/HCP and commercial datasets, supporting go-to-market strategy for new drug launches.
- Designed modular data pipelines using Kedro to integrate sales, promotional, and target lists data, ensuring reliable model inputs.
- Developed a prototype RAG workflow using ChatGPT to detect off-label discrepancies between localized Japanese FAQs and global drug labels.
- Contributed to Databricks migration and Feature Store transformation, improving feature reuse and scalability.

Zenni Optical | **Data Scientist,** Beijing, China — June 2023-Jan 2024

AI/ML Department

- Designed and deployed an OCR-based eyeglass prescription extraction API using FastAPI on GCP, including Japanese market adaptation.
- Integrated a multi-task deep learning model for blurriness detection, enhancing document quality control and downstream parsing accuracy.
- Conducted usage analysis with BigQuery and drove iterative product improvements based on user behavior insights.
- Delivered stakeholder-facing dashboards using Streamlit for progress tracking and feature demos.

University of Texas Southwestern Medical Center | **Data Scientist,** Dallas TX USA — Sep 2021-May 2023

Quantitative Biomedical Research Center

- Implemented Mask R-CNN for nuclei segmentation and classification in H&E histology images (82.5% detection, 82.0% accuracy).
- Created custom loss functions to handle partial/missing labels and support multi-cancer histology fusion tasks.
- Developed a CyTOF image processing pipeline (10× speed-up) with GUI and Docker packaging for easy deployment.
- Built EHR de-identification model and NLP preprocessing workflows to enable CLASI score prediction.

Donald Danforth Plant Science Center | **Data Science Researcher,** St. Louis MO USA — Feb 2020-Sep 2021

Data Science Facility

- Developed automated image analysis pipelines for plant phenotyping using RGB, thermal, and hyperspectral data.

- Led instance-level leaf segmentation using Mask R-CNN and tracking algorithms to quantify growth over time.
- Contributed to PlantCV open-source development, including new modules, testing, documentation, and user tutorials.

University of Florida Academic Health Center | **Data Science Intern** May 2019-Aug 2019

Precision and Intelligent Systems in Medicine Partnership Lab, Gainesville FL USA

- Preprocessed time-series patient vital signs and built cohort datasets for early risk stratification modeling.
- Reproduced and compared interpolation-based time-series models to evaluate clustering quality.

PERSONAL PROJECTS

MedKnow Recommender | Medical Knowledge Recommendation & Summarization System

- Designed and implemented a domain-specific recommendation system that integrates structured FAQs, drug label excerpts, and PubMed abstracts
- Supported bilingual drug name recognition (Chinese/English), unified multi-source content display, and summary generation using OpenAI GPT API
- Built a modular backend (Whoosh-based retrieval, PubMed crawler, LLM summarizer) and deployed an interactive Streamlit interface for real-time use

PUBLICATION

- Sheng, H., Wang S., et al. "MTIA: An open-source python package for systematic multiplexed tissue image analysis" (in preparation)
- Sheng, H., Gutierrez, J., Schuhl, H., Murphy, K. M., Acosta-Gamboa, L., Gehan, M., & Fahlgren, N. (2023). Increasing the Throughput of Annotation Tasks Across Scales of Plant Phenotyping Experiments. Authorea Preprints.
- Rong, R., Sheng, H., Jin, K.W., Wu, F., Luo, D., Wen, Z., Tang, C., Yang, D.M., Jia, L., Amgad, M. and Cooper, L.A., 2023. A deep learning approach for histology-based nucleus segmentation and tumor microenvironment characterization. Modern Pathology, 36(8), p.100196.
- Panda, K., Mohanasundaram, B., Gutierrez, J., McLain, L., Castillo, S. E., Sheng, H., ... & Slotkin, R. K. (2023). The plant response to high CO2 levels is heritable and orchestrated by DNA methylation. New Phytologist, 238(6), 2427-2439.
- Yu, G., Zare, A., Sheng, H., Matamala, R., Reyes-Cabrera, J., Fritschi, F.B. and Juenger, T.E., 2020. Root identification in minirhizotron imagery with multiple instance learning. Machine Vision and Applications, 31, pp.1-13.

ACADEMIC RESEARCH EXPERIENCE

Machine Learning and Sensing Lab | **Graduate Research Assistant**, Gainesville FL USA Mar 2017-Dec 2019

- Developed machine learning models for root detection in minirhizotron imagery using multi-instance learning.
- Proposed and implemented a Siamese-network-based dimensionality reduction method to classify plant genotypes.
- Designed and maintained hyperspectral/thermal data processing pipelines to support remote sensing research.

Institute of Precision Optical Engineering | **Undergraduate Research Assistant**, Shanghai, China June 2014-June 2015

- Conducted simulations of X-Ray KB (Kirkpatrick-Baez Microscope) imaging through programming, addressing and rectifying irregularities in the imaging process
- Developed expertise in correcting the unevenness of X-Ray KB imaging for improved accuracy and precision.
- Designed and implemented a user-friendly Graphical User Interface using MATLAB, facilitating a seamless and intuitive experience for navigating the simulated imaging process.

PROFESSIONAL STRENGTHS

- **Analytical & Fast Learner:** Quickly grasp new technologies and apply them independently; demonstrated by rapid onboarding and execution across diverse ML and NLP projects.
- **Effective Communicator:** Capable of translating technical concepts for both technical and non-technical audiences; frequently deliver stakeholder-facing demos and documentation.
- **Cross-cultural Collaboration:** Bilingual in English and Chinese with experience working across US and China teams, enabling smooth coordination in international and cross-functional environments.

CERTIFICATES, HONORS, REWARDS AND MISCELLANEOUS

- 1st Place of the "Swarm Behavior on the Grid" track in the Siemens "Tech for Sustainability Campaign 2023" (2023)
- Google Data Analytics Certificate– a rigorous, hands-on program that covers the entire scope of the data analysis process
- Co-Chair of the Committee for Scientific Training and Mentoring at Donald Danforth Plant Science Center