

# Hudanyun Sheng

[hudanyun.sheng@gmail.com](mailto:hudanyun.sheng@gmail.com) | (352) 281-3829 | <https://www.linkedin.com/in/hudanyunsheng/>

A data science researcher with 3 years of experience in machine learning algorithm development and image analysis to generate insights into data; making statistical analysis and compiling reports. Familiar with EHR data.

## EDUCATION

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

Master thesis: Switchgrass Genotype Classification using Hyperspectral Imagery

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

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

## TECHNICAL SKILLS

**Programming Languages and Tools:** Python (Numpy, Pandas, Scikit-learn, Seaborn, matplotlib, PyTorch, Tensorflow, keras, OpenCV), MATLAB, Git & Github, Jupyter Notebook, R, LaTeX, MS Office.

**Machine Learning:** Regression, Decision Tree, Random Forest, KNN, Naïve Bayes, SVM, PCA, SVD, Clustering, Time-series modeling, Nearest Neighbor, Neural Networks, K means clustering, NLP, Validation techniques

**Statistical Modeling:** Hypothesis Testing, ANOVA, Data Cleaning, Data Mining, A/B Testing

**Foreign Language:** Mandarin Chinese (native speaker)

## PROFESSIONAL EXPERIENCE

**Donald Danforth Plant Science Center | bioinformatics Researcher**, St. Louis MO USA Feb 2020-Present  
Data Science Facility

- Contributed to PlantCV (an open-source image analysis software package targeted for plant phenotyping) by improving and developing tools for image analysis, image segmentation, classification, and feature detection in image
- Established processing protocol for automated analysis of RGB, thermal and hyperspectral images to avoid repeated work; prepared data for analysis including batch merging and visualization of statistical analysis results
- Organizing data, guaranteeing data integrity, improving the quality, completeness, and consistency of data
- Collaborating with the research team to present data; communicating statistical outcomes using visualization tools
- Segmented leaf instances from images using pre-trained mask-RCNN model; linked time-series of every segmented leaf instance to understand the life-cycle of the plant during the time period

**University of Florida Academic Health Center | Data Science Intern**, Gainesville FL USA May 2019-Aug 2019  
Precision and Intelligent Systems in Medicine Partnership Lab

- Defined the analysis cohort and preprocessed raw data, including data cleaning, outlier detection, batch merging, and data visualization
- Extracted time-series feature by patients' major vital signs taken within their first 24-hour hospital admission
- Applied different methods including interpolation networks to deal with the irregularity of raw time-series data.
- Compared the time-series clustering results by generating comparison tables including statistical analysis automatically

## ACADEMIC RESEARCH EXPERIENCE

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

- Developed machine learning algorithms for automated root detection from mini-rhizotron images to save labor and time
- Established processing protocol for automated preprocessing of hyperspectral and thermal imagery of plants
- Developed algorithms for automated plant detection from hyperspectral images by hyperspectral endmember detection and hyperspectral un-mixing
- Classified genotypes of the same plant species by applying Siamese network on spectra of plant pixels

**Center of Applied Optimization | Graduate Research Assistant**, Gainesville FL USA June 2016-Feb 2017

- Preprocessed the electroencephalogram test data
- Boosted the classification performance with a run-time reduction by performing dimensionality reduction method

## PUBLICATION

G. Yu, A. Zare, H. Sheng, R. Matamala, J. Reyes-Cabrera, F. Frischi and T. Juenger, "Root Identification in Minirhizotron Imagery with Multiple Instance Learning," *Machine Vision and Applications*.