

YI HAN

☎ 352-213-8893 ✉ joshuahane@gmail.com 🌐 ehanml.github.io 📍 Gainesville, FL USA

SUMMARY

- Solid understanding and skills in machine learning, deep learning, and (Bayesian) statistics, developed from 7+ years of statistical training and applied research experience in data science.
- Strong learning ability, technical leadership, and communication skills developed from leading 5 highly interdisciplinary projects that involve deep collaboration with experts and engineers in diverse domains.

EDUCATION

University of Florida , Agricultural and Biological Engineering	Gainesville, FL USA
Ph.D. Candidate, UF Informatics Institute (UFII) Fellow, Machine Learning	Expected May 2022
University of Florida , Statistics	Gainesville, FL USA
M.S., Statistics	Nov. 2019
University of Florida , Animal Sciences	Gainesville, FL USA
M.S., Quantitative Genetics (applied statistics)	Dec. 2016
China Agricultural University , Animal Sciences	Beijing, China
B.A., Animal Sciences	May 2014

SKILLS

Programming: Python, R, SQL, perl

Machine Learning: tensorflow, keras, pytorch, sklearn, hyperopt, optuna, XGboost, lightGBM, caret, pyspark.ml

Statistical modeling and testing: statsmodels, stan, JAGS, NILA, nlme, mgcv, lmerTest

Data wrangling and visualization: numpy, pandas, matplotlib, seaborn, tidyverse series (dplyr, magrittr, ggplot2, etc.)

PROFESSIONAL EXPERIENCE

Anomaly detection of estrus with wearable computing sensors <i>Python, HPC</i>	Oct. 2020 – Present
<ul style="list-style-type: none">• Lead data scientist in the FACT project (\$500,000), providing data-driven insights and machine learning solutions.• Time series modeling with highly imbalanced and poorly labeled data using anomaly detection methods (VAE, AAE, LSTM encoder-decoder, GAN), which improved 20% F_1 score from supervised learning approaches.	
Upgrade satellite precipitation images with machine learning <i>Python, Google Cloud</i>	Aug. 2019 – Oct. 2020
<ul style="list-style-type: none">• Derived a training objective (hurdle loss) to model zero-inflated data under the deep learning framework, which outperformed MSE trained models as judged by a comprehensive set of metrics (accuracy, RMSE, Pearson correlation, mutual information).• Semantic segmentation of precipitation images using FCNN (U-Net) and Google Earth Engine.• <u>Publication:</u> Han, Y, Fraisse, C, Bliznyuk, N. 2021 DeepHurdle: a novel approach for deep learning-based precipitation merging studies. <i>Agricultural and Forest Meteorology</i>. Under review.	
Uncertainty quantification (UQ) of time series forecasts <i>Python, R, HPC</i>	Aug. 2018 – Aug 2019
<ul style="list-style-type: none">• Proposed an ML-based probabilistic ensemble method that combines diverse ML models into a superior model for risk control.• Developed/implemented a suite of UQ techniques (Bayesian NN, probabilistic forecasting, ARIMA, GP, GAM, INLA).• Proposed a probabilistic RF algorithm inspired by randomized prior and Bayesian dropout inference, and implemented it by modifying the XGBoost engine. The results achieved state-of-the-art performances on most benchmarks (UCI) regarding NLL.• <u>Publication:</u> Han, Y, Bliznyuk, N. 2021 Machine learning-based probabilistic ensemble for urban water demand forecasting. <i>Environmental Modelling and Software</i>. Under review.	
Sptaio-temporal modeling of precipitation occurrence <i>R</i>	Aug. 2017 – Aug. 2018
<ul style="list-style-type: none">• Precipitation detection with cost-sensitive learning (SVM, class-dependent loss, MetaCost).• Statistical and machine learning modeling (GAM, GP, INLA, kNN, SVM, RF, XGboost, MLP) for precipitation classification.• Used XGBoost to track the information flow across time and space, and provided several data-driven insights to domain experts.• <u>Publication:</u> Han, Y, Fraisse, C, Bliznyuk, N. 2021 Improving remote-sensing predictions of precipitation events using weather stations data with machine learning. <i>Agricultural and Forest Meteorology</i>. Under review.	
Statistical modeling and testing with genetic and kinship data <i>R, perl, HPC</i>	Aug. 2015 – Mar. 2016
<ul style="list-style-type: none">• Run high dimensional mixed models with BLUP and likelihood ratio test, and identified 8+ important genomics regions.• Gene set enrichment analysis using Fisher's exact test and revealed 3+ significant functional terms.• <u>Publication:</u> Han, Y. and Peñagaricano, F., 2016. Unravelling the genomic architecture of bull fertility in Holstein cattle. <i>BMC genetics</i>, 17(1), pp.1-11. (66 citations)• <u>Publication:</u> Nicolini, P., Amorín, R., Han, Y. and Peñagaricano, F., 2018. Whole-genome scan reveals significant non-additive effects for sire conception rate in Holstein cattle. <i>BMC genetics</i>, 19(1), pp.1-8. (18 citations)	

LEADERSHIP

Teaching fellow (<i>Statistical Machine Learning, Bayesian Methods and Applications</i>)	Aug. 2017 – Present
<ul style="list-style-type: none">• Presented lectures and led discussion-based sections. Supervised students in final projects, graded exams, quizzes, and homework.	
Graduate assistant	Aug. 2017 – Present
<ul style="list-style-type: none">• Advised graduates about experimental design and machine learning. Mentored junior students and assisted in recruiting.	