

WENKANG WEI

Current Address: Central, South Carolina, 29630 || **Phone:** 864-324-4885 || **Email (Preferred):** wenganw@g.clemson.edu

Tech Blog: wenkangwei.github.io || **LinkedIn:** www.linkedin.com/in/wenkang-wei-588811167/ || **Academic Profile:** meritpages.com/wenkan

SUMMARY

I'm a second-year master student with 5 years of experience in programming and 3 years of experience in Machine Learning. I'm actively looking for a full-time job or summer internship of **Software Engineer**.

EDUCATION

Clemson University, SC

- Master of Computer Engineering, Minor of Computer Science Expected in May 2021 GPA 3.79/4.0
- Bachelor of Electrical Engineering May 2019 GPA 3.84/4.0

TECHNICAL SKILLS

- Programming:** Python (PyTorch, Tensorflow, Pandas, etc), PostgreSQL, C/C++, Latex, Markdown, Object-Oriented Programming(OOP), data structure, client-server system
- Software Toolkits:** Linux, Git/Github, Docker, Flask(server framework) , GCP (Google Colab), AWS (EC2, RDS), VSCode, PyCharm, GDB, Qt
- Distributed Computing:** Apache PySpark, Hadoop MapReduce, Databrick Distributed Cluster
- Data Processing:** Data Pipeline Construction, Data ETL (Extract, Transform, Load), Data Wrangling
- Machine Learning:** Recommendation System (Matrix Factorization, etc), Deep Learning (CNN, RNN), Classification, Regression, Clustering
- Model Evaluation and Improvement:** Cross-Validation, ROC, AUC, Feature Importance, Ensemble Learning, etc.

PROFESSIONAL EXPERIENCE

Machine Learning Research Assistant

Clemson University, Summer 2020-Current

- Proved convergence and convergence rate of Multiple Update Algorithm (MUA) in Non-Negative Matrix Factorization Problem
- Formulated Matrix Factorization Problem to Constraint **Optimization** Problem and simplify problem with Linear Algebra, Lagrange multiplier
- Utilized Lipschitz gradient, convex optimization to prove the convergence and convergence rate of MUA algorithm
- Implemented MUA and **ALS (alternative least square)** algorithm in Google Colab and Matlab to verify convergence results
- Wrote a paper in AAAI format using **Latex** (unpublished due to copyright)

Team Leader in Kaggle Competition: Cassava Leaf Disease Image Classification

Kaggle, Fall 2020- Current

- Lead 2-person team to build a multi-task image classification system to classify cassava leaf diseases via a noisy image dataset from real world
- Construct data pipeline in **PyTorch** to extract and load Cassava leaf disease image dataset (5.76GB compressed data)
- Leverage data reduction methods to analyze data distribution and applied Image augmentation (cutout, mixup, etc) to transform images
- Utilize **efficient-net** and **visual transformer** in multi-task classification task, tune models with label smoothing, early stopping, weight decay techniques, evaluate model using cross validation and improve model accuracy in public score by 3% using ensemble learning
- Achieve **0.905** accuracy in public score and rank **top 3% out of 3133 teams** in kaggle leaderboard currently

Leader in Human Activity Data Collection (20GB) and Wrangling

Clemson University, Fall 2020-Current

- Write tutorial documents with Markdown in Github and train **10 students** to collect wrist motion data from daily life to analyze eating behaviors
- Mentor and assist each student to collect, clean and label 2GB individual data in 2 weeks and transform time series data for data wrangling
- Construct robust data pipeline to solve buffer overflow problem to extract, load and transform **large-scale time series dataset (20GB)** in **1 min**
- Visualize and analyze imbalanced data with Pandas, Seaborn, Matplotlib and smooth data using moving average for data augmentation
- Build Convolution Neural Network to classify and segment eating period and achieve the **best weighed accuracy 96%** in cross validation

Coordinator in Hackathon Competition: Human-Detection based Smart Reminder

Clemson University, Spring 2018

- Ignited and designed a novel software solution, a reminder Raspberry-Pi robot that uses computer vision to detect and remind users of taking break after usage of computer to improve users' health
- Collaborated with **3 multi-disciplinary** teammates (CS, CpE, EE) to design and build client-server system and human detection system in **1 day**
- Applied **Flask** server in Raspberry Pi to send images to Laptop client and utilized OpenCV and pre-trained Fast-R CNN **Tensorflow** model in client for human detection in Python
- Assisted to integrate, test server-client systems and achieved the **Best Overall award** in Hackathon competition

SELECTED PROJECTS EXPERIENCE

Recommendation System based on MovieLens 25M dataset (PySpark, Hadoop, HDFS, SQL)

Fall 2020

- Constructed a recommendation system using **Collaborative Filtering** and **Matrix Factorization** methods to analyze the genres contributing to ratings and make recommendations to users
- Utilized **PySpark** and **SQL** to load, query and analyze movielens dataset (**25 million ratings**) in databrick distributed cluster platform
- Implemented and applied mapper, reducer functions in **Hadoop** File system to analyze different movie genres to ratings

Real-time Signal Visualization System (C++, Qt, GDB)

Spring 2018

- Designed a visualization software system based on Qt package, Arduino using C/C++ to visualize voltage signal in real time with self-motivation and initiative
- Designed GUI components and class modules for software interface in Qt and software framework to control data visualization behaviors in C++ using **data structure** (queue) and **Object-Oriented Programming (OOP)** techniques
- Integrated, tested and debugged GUI components with software framework using GDB toolkit and Qt IDE
- Wrote technical document for software system in Github with video demo. Link to demo: <https://github.com/wenkangwei/SerialPlot>

HONOR

Eta Kappa Nu (HKN) (Spring 2018 - present);

Golden Key Honor Society (Fall 2017- present);

President's List (Summer 2019);

Dean's List (Fall 2016 – May 2019);

Best overall hack award of CUhackit Competition (Spring 2018);