

# Donglei (Derek) Cai

Berkeley, CA | 508.322.0130 | [dcai@berkeley.edu](mailto:dcai@berkeley.edu) | LinkedIn:  
<https://www.linkedin.com/in/derek-cai-b83079155/>

---

## Education

**University of California, Berkeley** Aug 2017 - Present

**Double Major:** B.A. Applied Mathematics with Concentration in Statistics      B.A. Computer Science

**Relevant Coursework:** The Structure/Interpretation of Computer Programs, Data Structures, Introduction to Machine Learning, Principles/Techniques of Data Science, Linear Algebra/Differential Equations, Numerical Analysis, Discrete Mathematics, Probability theory, Time series, Convex Optimization

## Experience

**DNA Compass - Data Consultant**

June 2020 - September 2020

- Analyze COVID-19 metadata from UK Biobank for medical research

**Student Association for Applied Statistics - Education Director**

January 2021 - Present

- Develop introductory data science/machine learning lessons and teach weekly lessons
- Organize club workshops and mentor students new to data science/machine learning

**Student Association for Applied Statistics - Data Consultant**

September 2021 - Present

- Work on time series forecasting for a telecommunication company

**Berkeley RISELab - Computer Vision Researcher**

June 2021 - Present

- Develop efficient computer vision models for real-time video super-resolution

## Data Science Competition

Kaggle Competition Master (highest rank: 763/168k+)

**BirdCLEF 2021 - Birdcall Identification** Rank: 7/816

- Perform audio multi-class classification task and develop an ensemble model that consists of ResNeSt50, ReXNet100 and ReXNet150
- Apply data augmentation(Add background noise, Add Gaussiann noise, SpecAugmentation, mixup, secondary labels)
- Use Bayesian statistics to adjust threshold probability for classification

**SIIM-FISABIO-RSNA COVID-19 Detection** Rank: 20/1305

- Perform image multi-class classification and object detection task and develop an ensemble model that consists of EfficientNet V2L, EfficientNetB3 and different YOLOv5
- Use external data from BIMCV and RICORD as well as test data to generate pseudo labels for training
- Develop a binary classification to help filter out false positive predictions for object detection models
- Use Bayesian statistics to tune object detection accuracy using multi-classification model output

**SETI Breakthrough Listen - E.T. Signal Search** Rank: 27/768

- Perform image binary classification task and develop an ensemble model that consists of different EfficientNet B series, EfficientNet V2L and ResNet18
- Use external data to pretrain models and perform use test data to generate pseudo labels for training