

Batu Yildirim

Email: by256@pm.me | Web: by256.github.io

I am a Research Engineer at InstaDeep in the BioAI team, where I conduct research and write software to design SARS-CoV-2 vaccines. Previously, I completed my PhD in the Molecular Engineering group at the University of Cambridge, where I researched machine learning methods and wrote software for characterization and generation of atomic and nanostructures. My thesis was titled *Machine Learning for Structural Characterisation and Generation: Applications to Small-Angle Scattering and Electron Microscopy*.

Education

University of Cambridge	PhD Physics and Machine Learning	Cambridge, UK (Oct 2018 - Oct 2022)
Queen Mary, University of London	MSc Data Science	London, UK (Sep 2017 - Sep 2018)
University of Manchester	MEng Materials Science	Manchester, UK (Sep 2013 - Jun 2017)

Employment Experience

Research Engineer II

London, UK (Mar 2023 - Present)

InstaDeep

- Designing and implementing both original and published neural network and machine learning methodologies for protein modelling and design in the BioAI team.
- Formulating experiments and employing hypothesis testing methods for analysis of data, experimental outcomes and comparative analysis of machine learning models.
- Development and continuous improvement of software for SARS-CoV-2 vaccine designs.
- Constructing data pipelines for large biological datasets, including the curation and refinement of data for experiments, analyses and model training/evaluation.
- Actively engaging with collaborators and stakeholders to consistently communicate and present my research methods, results and findings.

Research Scientist (Intern)

Harwell, UK (Sep 2019 - Sep 2020)

Science and Technologies Facilities Council

- Conducted research into the use of machine learning methods for materials characterization using electron microscopy and small-angle scattering data, resulting in two publications.

Data Scientist (Intern)

London, UK (Jun 2018 - Oct 2018)

StatusToday

- Created structured datasets from unstructured data and implemented ML models to classify user activity from automated system activity, leading to more accurate insights.

Competitions

Numer.ai

Remote (May 2021 - Present)

- ML quant competition, where I earned the rank of master (2nd) in 2022 and expert (17th) in 2023 ([profile](#) | [model](#)).
- Developed ML time-series models and performed statistical testing at multiple stages of my pipeline, notably during feature selection and model selection. Addressed non-stationarity and co-dependence between features and reduced my model's exposure to volatile features, resulting in stable performance over time.

Citadel Datathon

Dublin, IE (Jan 2019)

- 2nd place at the 2019 Citadel Dublin Data Open. \$5000.

Technical Skills

Python (advanced): experienced in building applications, writing scripts and data analysis; **C++** (foundational): solid grasp of syntax and capable of writing straightforward yet functional programs.

PyTorch (advanced): ability to build complex model architectures and train models; **NumPy** (advanced): numerical computing, encompassing array operations, matrix manipulations and advanced linear algebra techniques for scientific computing; **Pandas** (advanced): data manipulation, including data transformation, aggregation and merging for detailed analysis; **scikit-learn** (advanced): predictive and statistical modeling/analysis.

Git (proficient): regular use for version control, collaboration and maintaining projects; **Distributed computing** (basic): ability to implement and manage distributed computing processes, including parallel computations, data partitioning, and

process synchronization across multiple nodes and clusters; **Docker** (basic): ability to containerize applications and perform basic container management tasks such as building images and running containers.

Honours and Awards

- Fitzwilliam College, Senior Scholarship: awarded in recognition of significant research progress made during COVID-19. (Nov 2020)
- Fitzwilliam College, Senior Scholarship: awarded on the basis of excellent work. (Nov 2019)
- Rolls-Royce/Tin-Plate Workers Award: awarded for achieving first-class honours and finishing top of my class at the University of Manchester during my third year. (Aug 2017)

Open-Source Projects

Core Developer

- **rdfpy**: a Python module for fast computation of 2-D and 3-D radial distribution functions. (Jun 2020 - Jan 2023)
- **ImageDataExtractor**: a Python framework for electron microscopy image quantification. (Dec 2019 - Jan 2023)

Contributor

- **Ensemble-PyTorch**: a unified ensemble framework for PyTorch. I extended the API by implementing the capability to use arbitrary loss functions when training ensemble models (see commit [bb7b988](#)). (Aug 2021)

Publications

1. B. Yildirim, J. Douth, and J. M. Cole, "Multi-Task Scattering-Model Classification and Parameter Regression of Nanostructures from Small-Angle Scattering Data" *RSC Digit. Discov.*, 2024 [10.1039/D3DD00225J](#).
2. B. Yildirim, A. Washington, J. Douth, and J. M. Cole, "Calculating Small-Angle Scattering Intensity Functions from Electron Microscopy Images" *RSC Adv.*, vol. 12, pp. 16656–16662, 2022, [10.1039/D2RA00685E](#).
3. B. Yildirim, J. M. Cole, "Bayesian Particle Instance Segmentation for Electron Microscopy Image Quantification" *J. Chem. Inf. Model.*, vol. 61, no. 3, pp. 1136–1149, 2021, [10.1021/acs.jcim.0c01455](#).
4. B. Yildirim, C. J. Court, J. M. Cole, "3-D Inorganic Crystal Structure Generation and Property Prediction via Representation Learning" *J. Chem. Inf. Model.*, vol. 60, no. 10, pp. 4518–4535, 2020, [10.1021/acs.jcim.0c00464](#).
5. K. T. Mukaddem, E. J. Beard, B. Yildirim, J. M. Cole, "ImageDataExtractor: A Tool To Extract and Quantify Data from Microscopy Images" *J. Chem. Inf. Model.*, vol. 60, no. 5, pp. 2492–2509, 2020, [10.1021/acs.jcim.9b00734](#).