Charles Harris, BSc

07792 541377 | charles.harris17@imperial.ac.uk | github.com/cch1999 | cch1999.github.io

Summary

- Current MSc student in Bioinformatics and Theoretical Systems Biology at Imperial College London
- Passionate about state of the art in machine learning being applied to solve problems in biology, particularly structural biology and drug design. Experience with GNNs for learning protein dynamics and protein ligand interactions.
- Strong leadership and communication skills developed from being Founder of student initiative to promote computational biology to students at Imperial College, rapidly grew to be one of the largest academic societies in the College
- Organiser and chair of then "Imperial AI in Drug Discovery" Conference
- Deep understanding of protein structural data, genomics and drug design
- Contributing to open source project (Graphein) to democratise access to protein graph data for geometric deep learning

Education

Imperial College London London, UK

MSc in Bioinformatics and Theoretical Systems Biology (Distinction in all work/exams so far) *Oct. 2020 – Sept. 2021*

- Component 1 (provisional grade: Distinction): Bioinformatics, Systems Biology, Mathematics, Modelling, Statistics, Computing
- Component 2: Group Computing Project (Provisional grade: Distinction), Data Analysis Project
- Component 3: Bioinformatics and Systems Biology Project

Imperial College London London, UK

BSc in Biochemistry (2:1) *Oct. 2017 – Jun. 2020*

- Final Year Modules: Bioinformatics (1st), Integrative Systems Biology (1st), Structural Biology and Drug Design (2:1 - 68.94%)

London School of Economics Summer School London, UK

Introductory Microeconomics *Aug. 2019 – Aug. 2019*

Technical Skills

Python: Tensorflow/Keras, PyTorch, PyTorch Geometric, Deep Graph Library/DGL-LifeSci, NetworkX, Scikit-learn, RDKit, Graphein, Numpy, Pandas, Data science (Linear regression, ROC curves), Matplotlib/Seaborn
Statistics/ML: Probability, Likelihood statistics, Bayesian statistics
Maths/Systems biology: Linear algebra, ODEs, Master equations, Gillispie simulations, Dynamical systems (linear stability analysis, bistability, phase portraits), Quantitative cell imaging, proficient in MATLAB
Bioinformatics: Genome assembly and annotation, RNA-seq data analysis, Sequence alignment, Protein function and structure prediction, NGS data analysis in R and Artemis
Structural Biology: Deep understanding of protein structure. Highly proficient in PyMol.
Biochemistry and molecular biology: Strong background of biology and biochemistry from BSc
Computing: Linux, SQL, GPU machines, familiar with use of computing clusters

Research - see GitHub - literature available upon request

MSc Final Project Supervisor: Prof Michael Bronstein

"Geometric Deep Learning for structure-based drug design" *June. 2021 – June. 2021*

- Designing a generative graph model (either a graph VAE or normalising flows) that can perform structurally aware generation of 3D models in a protein pocket

MSc Data Analysis Project - Awaiting grade Supervisor: Prof Michael Bronstein

"Learning interpretable protein dynamics using Geometric Deep Learning" *April. 2021 – June. 2021*

- Created a novel message-passing layer for protein graphs that learns interpretable protein dynamics

MSc Computing Project - Distinction Supervisor: Prof Michael Sternberg

"Prediction of the effect of missense variants involved in ligand binding via docking" *Jan. 2021 – April. 2021*

- Developing a chemoinformatic pipeline to automatically identify the biological ligand associated with an unannotated protein sequence, predict a 3D structure (likely using Phyre2) and predict whether mutations effect ligand binding by using docking
- Have become familiar with common chemoinformatic tools (e.g. RDKit)

Final Year Research Project - 2:1 (69.53%) Supervisor: Prof Michael Sternberg

"Deriving structural explanations for protein thermal stability"- Repository *Feb. 2020 – Jun. 2020*

- Final year research project looking at which structural effects (e.g. salt bridges and hydrogen bonds) contribute most to protein thermal stability

2nd Year Tutored Dissertation - 1st Supervisor: Prof Michael Sternberg

"In silico protein function prediction from structural data" *Jan. 2019 – Apr. 2019*

- Focused on the latest advancements of machine learning in protein function prediction

Personal Projects

Research into meaningful representation of 3D proteins structures for machine learning models *On going*

- Currently interested in ways of representing protein structures to be used in geometric deep learning models
- Graphein: Open-source contributor to package aiming to democratise access to protein graph data, expect to be published in the next year
- DMCNN: Implemented algorithm to predict protein function using protein distance matrices and convolutional neural networks. Achieved 90%accuracy

References

Prof Michael Bronstein - Imperial College London, Twitter m.bronstein@imperial.ac.uk

Chair of Machine Learning and Pattern Recognition at Imperial College

- MSc project co-supervisor

Prof Michael Sternberg - Imperial College London m.sternberg@imperial.ac.uk

Director - Centre for Integrative Systems Biology and Bioinformatics

- Final Year Research Project Supervisor, 2nd Year Dissertation Supervisor and MSc Course Director

Prof Bruno Correia - EPFL bruno.correia@epfl.ch

Head - Laboratory of Protein Design and Immunoengineering

- MSc project co-supervisor

Prof Stephen Curry - Imperial College London s.curry@imperial.ac.uk

Professor of Structural Biology

- Undergraduate personal tutor and X-ray crystallography lecturer

Dr Derek Huntley - Imperial College London d.huntley@imperial.ac.uk

Senior Teaching Fellow - Department of Life Sciences

- Computational Biology Society Advisor, 3rd Year Bioinformatics Module Convenor and MSc Co Director

Volunteering

Chair and Founder

Imperial College Computational Biology Society *Oct. 2019 – Present*

- Created and chaired Imperial's first **AI in Drug Discovery Conference** with top scientists, business leaders and entrepreneurs, over 1,400 tickets sold
- Moderated two panel discussions as part of conference (first one with 4 CEOs/Founders of AI in Drug Discovery companies and second with Prof Sir Tom Blundell, Prof Michael Bronstein and Dr Andreas Bender)
- Organised regular webinar schedule with top companies and academics in computational biology (including DeepMind (Dr Petar Veličković) BenevolentAI (Dr Nathan Brown), Nvidia, LabGenius, Imperial and UCL)
- Organised six workshops to teach biologist computational and coding skills
- Quickly grew the Society to over 340 members and developed a strong social media presence
- Showed strong leadership, organisational and team building skills by creating an effective team and culture from scratch. Managed 20 volunteers and numerous projects in the first term
- Demonstrated resilience and negotiating skills after five month long discussions with the Student Union to defend the creation the Society

Ambassador
Helen Arkell Dyslexia Charity *Oct. 2019 – Present*

Secretary
Imperial College Amnesty International Society *Oct. 2019 – Aug. 2020*

Kit Secretary
Imperial College Hockey Club *Aug. 2018 – Jul. 2019*

Stage Manager
Leighton Park School *Jun. 2015 – Jul. 2017*

Other Skills

Languages: Mandarin (A at GCSE Level)
Interests: Solo glider pilot, hockey, running, drone photography