AIMLAC CDT Newsletter

Our first-year cohort at the induction event in Cardiff, September 2019

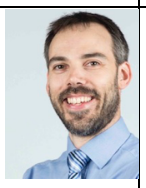
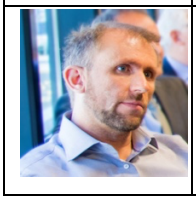
Welcome to our first AIMLAC CDT newsletter, here we introduce you to the students in our first cohort and update you on what they have been doing since we last met with you at the CDT induction event.

First of all, during this time of great uncertainty, we want to extend our best wishes for health and happiness to you, your family, friends and colleagues. We know this is a concerning time for all.

Concerning the global pandemic, we are following university, research council and government advice. All students and team members are currently working remotely. This includes teaching, training and supervision activities, to ensure students are as little affected as possible. Similarly, the CDT management team is continuing CDT support without interruption.

Meet the Management Board

http://cdt-aimlac.org/cdt-people.html

	Prof Gert Aarts - CDT Director Swansea		Prof Biagio Lucini - CDT Technical Director Swansea		Dr Shangming Zhou Swansea
	Rhian Melita Morris - CDT Centre Manager Swansea		Roz Toft - CDT Research Support Officer Swansea		Prof Reyer Zwiggelaar Aberystwyth
	Prof Stephen Fairhurst Cardiff		Dr Henning Flaecher Bristol		Prof Jonathan Roberts Bangor
	Owain Huw Super Computing Wales		Prof Malcolm Bremer Bristol		Prof Roger Whitaker Cardiff

Student Representative committee

Thomas Spriggs – Student Representative for CDT, Chair of the CDT Student Board, Student Representative for Swansea University CDT Governance group

Tonicha Crook – Student Representative for CDT, Co-Chair of the CDT Student Board

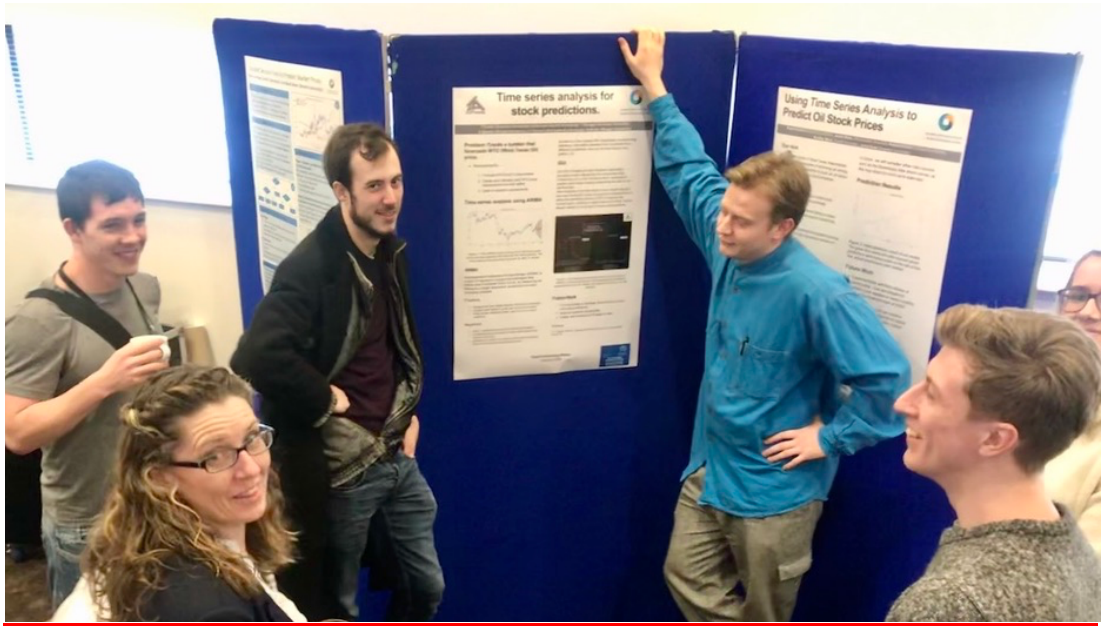
Sophie Sadler – Co-Chair of the CDT Student Board

“A New Decade of Supercomputing” January 2020

Supercomputing Wales hosted its inaugural conference from 23rd – 24th January 2020 at the Millennium Centre, Cardiff. The conference, which welcomed almost 90 delegates and a number of high-profile speakers, considered how High Performance Computing will develop over the next decade and the integral role of supercomputers in our future.

The CDT students attended the conference, and presented their coding challenge work with the SCW Research Software Engineers group

A number of presentations from the conference can be viewed on their website. https://www.supercomputing.wales/wp-content/uploads/2020/02/Conference-slides-pack.pdf



Students giving presentations on their coding challenge findings

Student volunteering

DataAid is an exciting new student led initiative that will provide data science consultancy to a group of carefully selected charities in collaboration with the DI and AIMLAC CDTs. This consultancy shall be carried out by the students and could include data analysis, visualisation and advice on how to improve existing data sets, generate new data or utilise existing open source data. This is a great opportunity for students to work together across multiple Universities and disciplines, share skills and collaborate on real world projects. Keep an eye out for DataAid at the upcoming CDT events.

Meet the Students

The first cohort includes students from the AIMLAC and the DI CDT's

Vanessa Cassidy - Bangor

Research area: My research is focussed on the different methods and potentialities of using machine learning algorithms to semi-automate the design and layout of a narrative data visualisation.
Research theme: T3: Novel mathematical, physical, and computer science approaches (data, hardware, software, algorithms)
PhD Title - Smart storytelling for scientific data visualisation
Hobbies/Interests: I enjoy playing piano and guitar and various outdoor activities, which I try and do as often as possible with my young family.

Tonicha Crook - Swansea

Research area: It has often been conjectured that the imperfections of game theory's ability to accurately describe human strategic behaviour is due to its neglect of the limited information processing abilities of the agents. Correspondingly, there has been a call to study bounded rationality, and to explore what rational behaviour under such information processing constraints means. In this project, I will study the consequences of demanding that agents employ computable strategies. A specific focus will be on multiplayer games.

Research theme: T3: Novel mathematical, physical, and computer science approaches (data, hardware, software, algorithms)
PhD Title: Consequences of demanding that agents employ computable strategies
Hobbies/interests: I enjoy reading and play games with friends, as well as cross-stitch and drawing.

Jamie Duell - Swansea

Research Area: Machine Learning in Medical Science
Research theme: T2: Biological, health and clinical sciences (medical imaging, electronic health records, bioinformatics)
PhD Title: Enhancing the Safe and Effective Use of Medicines in Hospitalized Patients: An AI Led Population Health Study
Hobbies/interests - I have strong interest in the state of the art with regards to AI and robotics with societal effect. Outside of academia, I previously spent my time competing at gaming events situated in the U.K. for various gaming organisations, this taking most of my free time previously.

Lily Major - Aberystwyth

Research area: A highly interdisciplinary project, involving computer science, stringology, bioinformatics, and algorithmics. Focusing around alphabet ordering methods (including bio-inspired randomized search heuristics such as evolutionary algorithms), Lyndon factorization, the Burrows-Wheeler Transform (BWT), and suffix arrays, using publicly available metagenomic data. This will further lead to an investigation of how factorization affects partitioning and parallelization of algorithms used for search and alignment in bioinformatics. Other investigations will involve compression technology similar to bzip2 and how alphabet orderings affect the BWT used in such algorithms.
Research theme: T2: Biological, health and clinical sciences (medical imaging, electronic health records, bioinformatics)
PhD Title: Big Data algorithmics for efficient search and analysis of large collections of genomes.
Hobbies/Interests: Retro-computing, electronics.

Michael Norman - Cardiff

Research theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD Title: Deep Learning for Real-Time Gravitational Wave Detection
Hobbies/Interests: Creative Writing, D&D, Pub Philosophy

Sophie Sadler - Swansea

Research area: The core idea of my research is to use data visualizations to provide explanations for the outputs of machine learning algorithms. Currently, I am working on a project involving social network analysis, with an aim to provide greater insight into various community finding algorithms. The approach we are using is to calculate a range of features of the structure of different networks and then use random forests to discern which of these features are most important when explaining how different community finding algorithms cluster the nodes of the graph. Though the

initial research I have done so far is theoretical, we are considering various follow-up projects applying this work to real-world social networks.
Research theme: T3: Novel mathematical, physical, and computer science approaches (data, hardware, software, algorithms)
PhD Title: "Visual Analytics for Explainable Graph-Based Machine Learning".
Hobbies/Interests: hiking & wild camping, creative writing, watching horror movies, and riding my horse.

Hattie Stewart - Bristol

Research area: The Square Kilometre Array (SKA) will generate extremely large datasets, containing a very high density of features that must be disentangled to classify the contributing sources and to examine radio propagation effects in the magnetised medium between the sources and the detector. The aim of the project is to develop and implement AI methods for inferring appropriate information from the six-dimensional data, initially using trial data from the SKA data challenges, but then working with real-life data from precursor projects. Attention will be paid to imperfections in the data that may cause systematic errors. Supervised and unsupervised techniques will be explored to optimise the efficiency and yield of object classification, including but not limited to: Convolutional Neural Networks, Gaussian Processes, Genetic Algorithms and Clustering methods.
Research theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD Title: AI techniques for extracting source information from Square Kilometre Array (SKA) datasets Hobbies/Interests: Rock climbing, surfing, travelling, plants, music

Cory Thomas - Aberystwyth

Research area: AI in Mammography
Research theme: T2 / T3 Biological, health and clinical sciences (medical imaging, electronic health records, bioinformatics) / T3: Novel mathematical, physical, and computer science approaches (data, hardware, software, algorithms)
PhD Title: Modelling the Development of Breast Cancer Abnormalities
Hobbies: Rugby, general sport
Interests: Computer modelling of real life and astrophysical problems

Bradley Ward - Cardiff

Research theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD title: Investigating the epoch of galaxy formation using artificial intelligence
Hobbies/interests - My hobbies include Cricket and amateur astronomy and astrophotography. I am also part of the Baseball and Aerial Fitness societies at Cardiff.

Robbie Webbe - Bristol

Research Area: Active Galactic Nuclei (AGN) form some of the most luminous known objects in the universe, with radiation emission being powered by accretion of matter on to Supermassive Black Holes. The x-ray variability of these objects is an

established part of astronomy, but in the last year two objects have been found which show variability on scales, and within time periods, which is far outside expected behavior, coining the term Quasi-Periodic Eruptions. By finding more exotically varying objects we will be better able to understand the behaviour of AGN in the most extreme cases. We already have large quantities of observations of objects, from missions like RXTE, Chandra and XMM, and with new observatories collecting more information, and more detailed information, by the day I hope to be able to use Machine Learning to find the needles within the haystack.
Research Theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD Title: Detecting exotic variability in Active Galactic Nuclei through Machine Learning
Hobbies/Interests: I am a keen runner, having completed races form 10K to marathon distance, and also enjoy playing team sports including football and hockey. Away from the sports pitch I enjoy problem solving and can often be found solving a crossword or sudoku.

Ben Winter - Bangor

Research area: My research area involves finding unique and optimal ways to enhance neural networks, through the evolution of their topology, network structure and weights. My goal is to create an optimal neural network for solving classification problems such as breast cancer detection and black hole detection, whilst also creating a grammar as to not only make the network more reproducible, but also to make neural networks and evolutionary algorithms less black box.
Research theme: T3: Novel mathematical, physical, and computer science approaches (data, hardware, software, algorithms)
PhD Title: Grammatical Neuroevolution
Hobbies/Interests: I enjoy watching Football and Formula 1, playing guitar and video games and socialising with friends.

Chris Wright - Bristol

Research area: The LUX-ZEPLIN (LZ) experiment aims to try to detect thus far elusive dark matter by observing the light produced when a dark matter particle recoils from an atomic nucleus using photomultiplier tubes (PMTs). My work involves the development and application of a convolutional neural network to pulse analysis and waveform reconstruction of the output of those PMTs. Preliminary results look promising that this can achieve results that are superior to the current generation software in dealing with pileup and the discrimination of pulses that occur very close together and the analysis of said pulses, and has potential applications for feature recognition in time series data more generally. This will hopefully contribute in some small way to allowing LZ to make the most of the data it obtains, and to maybe, if we're lucky, finally find a window into the roughly 95% of the "stuff" in the universe that we can't yet see!
Research theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD Title: Multi-channel waveform reconstruction for dark matter searches with LUX-ZEPLIN

Hobbies/Interests: Drama/Theatre, Tennis, and science fiction books, films and games. Also want to write a novel one day.

Data Intensive CDT

Elenora Parrag - Cardiff

Research area: Population studies of supernovae can track their evolution, and that of their progenitor stars, in different environments throughout the history of the universe. Tracking these populations over time can provide answers about cosmology and the evolution of the universe. However, population studies require comparisons of supernovae at the same epoch, as each stage of the supernova is determined by different physics. Datasets are largely incomplete in covering every stage of a supernova, as observation time and resources are limited. The aim of my PhD will be to develop a machine learning tool to relate supernovae spectroscopy and photometry at different times, connecting late and early time physics and allowing a more complete dataset to be inferred which in turn can allow for improved population studies.
Research theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD Title: Rewinding supernovae with machine learning
Hobbies/Interests: Indoor climbing and bouldering, hiking

Tom Spriggs - Swansea

Research area: Currently I am studying what is known as Spectral Reconstruction, it effectively consists of creating a model from which we could extract real world observables about particle physics (masses of particles etc), and then performing some ‘physics’ on this model until it is comparable to the data from lattice QCD simulations. This allows me to refine my model and/or parameters to better recreate the data, and more accurately extract the observables (which can then be checked against data from the LHC at CERN). The overall aim is to then test the lattice QCD simulations in conditions where experiments can also probe, to better trust our predictions in areas where they cannot, extending the reach of particle physics without the need for larger and larger colliders.
Research theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD Title: Lattice QCD - Investigating the spectrum, spectral features, and potential of hadronic states of QCD in a thermal medium
Hobbies/Interests: Football, film photography, and surfing

Raul Stein - Bristol University

Research area: I am interested in looking at ways to accelerate machine learning models and analysis on specialised hardware for triggering applications in particle physics detectors. Currently, my research is centred on improving the performance of the triggering system for DUNE (Deep Underground Neutrino Experiment) by implementing ML algorithms on FPGAs (Field Programmable Arrays). This has the potential to greatly improve data selection quality for large physics experiments where low latency (scale of 100ns) is crucial if it is possible to optimise and

compress the neural network to fit the resource limits of FPGAs. More specifically, I am looking into performing this online analysis on the detector readout to detect signs of the key physics events such as atmospheric neutrino interactions, supernova burst neutrinos, or nucleon decay. In the future I expect to also investigate the potential of IPUs (Intelligence Processing Units) for this task and applications in the other particle detectors such as the CMS at CERN.
Research theme: T1: data from large science facilities (particle physics, astronomy, cosmology)
PhD Title: FPGA implementation of machine learning for low latency data processing in particle detectors.
Hobbies/Interests: Fitness & Distance running, Break Dance, Fintech, Travelling.

Conferences, papers and projects

Tonicha Crook

36th British Colloquium for Theoretical Computer Science, 6th-8th April 2020, Swansea
https://cs.swansea.ac.uk/bctcs2020/

Topic and brief outline: The degree of non-computability of Nash equilibria in multiplayer games. Is there an algorithm that takes a game in normal form as input, and outputs a Nash equilibrium? If the payoffs are integers, the answer is yes, and a lot of work has been done in its computational complexity. If the payoffs are permitted to be real numbers, the answer is no, for continuity reasons. It is worthwhile to investigate the precise degree of non-computability (the Weihrauch degree), since knowing the degree entails what other approaches are available (e.g., is there a randomized algorithm with positive success change?).

Experience of presenting online: It still made me nervous but as there were only a few faces on screen when you share your slides, it makes it seem like there is only a few watching instead of a whole room. Even though it was online I prepared for it like a normal conference so hopefully this will help me for future face-to-face conferences.

How many people: Maximum was around 90

Fellowships

Congratulations to Professor Reyer Zwiggelaar FLSW for being elected as Fellow of the Learned Society of Wales.

Reyer is a Management Board member and lecturer on the RRI module.

Papers submitted

Recently Lily Major submitted a paper to PPSN 2020 - 16th International Conference on Parallel Problem Solving from Nature. The conference "welcomes works on all types of iterative optimization heuristics" or "connections between search heuristics and machine learning or other artificial intelligence approaches" (https://ppsn2020.liacs.leidenuniv.nl/) and all papers are presented as posters. In our paper - Evaluation of a Permutation-Based Evolutionary Framework for Lyndon Factorizations - we introduce a mutation-based evolutionary framework for finding Lyndon factorizations with two aims: minimization of the number of Lyndon factors, and balancing the length of each Lyndon factor for a given input string. Balancing is achieved using both the difference between the maximum and minimum length Lyndon factors, as well as using the standard deviation of the lengths of the Lyndon factors as fitness functions. Further, we present a modification to Duval's Algorithm for computing the Lyndon factorization of a string: the new heuristic algorithm (Flexi-Duval) does not assume the underlying alphabet is ordered but rather induces an ordering during the linear scan of the string. We also propose a problem-specific mutation operator for Lyndon factorization which is used in our evolutionary algorithm. Our results show that our framework is competitive with Flexi-Duval for minimization and yields high quality and robust solutions for balancing where no problem-specific algorithm is available.

Conferences attended

Lily Major attended COST Action CA15140 ImAPPNIO in 2019 in Coimbra, Portugal (https://imappnio.dcs.aber.ac.uk/training-school-2019/training-school-2019) with funding from COST. This involved training by experts in nature inspired optimisation. Topics covered included: machine learning, permutation-based problems, and theory of bio-inspired computation, among others (https://users.aber.ac.uk/thj10/flyer.pdf). The training at the event has was very useful in conjunction with the training received from the CDT in learning about the area.

External partners

The CDT works with a range of external partners, some examples are below. If you are not on the list and would like to work with and support us, please contact us at aimlac-cdt@swansea.ac.uk

Industry placements

We value our relationship with our industrial partners, and the industry placements are core to our CDT student development.

Next year our second-year students will be preparing for their 6 month placement in industry.

How it works:
6 months placement in year 2/3 with a two-week intro early in the second year

First-year cohort	first one-to-one contact with external partners at the CDT induction event, September 2019.
Second-year cohort	via “speed dating” events at the annual meetings, to exchange interests, expectations and demands a two-week placement at the selected external partner, to familiarise both the partner and the researcher with the company, scope and expectations and doctoral researchers will start working on a realistic task
Second/third- year cohort	straddling the second and third year, the doctoral researcher will spend a substantial time (up to six months) with the partner, working on a substantial task, after which the doctoral researcher will finalise their research project and thesis

2020 Cohort

We have recruited our second cohort of 12 excellent students to start on October 1st, 2020.

Contact

If you are interested in hosting a placement, or if you would like to know more about our industry placement process, and have a project in mind for a student, please get in touch with us at cdt-aimlac@swansea.ac.uk

Rhian Melita Morris - CDT Centre Manager (Swansea) r.m.morris@swansea.ac.uk
Mrs Roz Toft - CDT Research Support Officer (Swansea) r.toft@swansea.ac.uk

http://cdt-aimlac.org/cdt-main.html @AimlaCcommunity