AIMLAC

UKRI CENTRE FOR DOCTORAL TRAINING IN **ARTIFICIAL INTELLIGENCE, MACHINE LEARNING** AND **ADVANCED COMPUTING**

Table of Content

Aberystwyth University

Project title: A machine learning approach to predict cardiovascular diseases from electrocardiogram (ECG) data using 3D visualisations

1st Supervisor: Dr Faisal I Rezwan
2nd Supervisor: TBD

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T2 - biological, health and clinical sciences T3 - novel mathematical, physical and computer science approaches

Project description: The electrocardiogram (ECG), a low-cost, rapid and simple test, is a widely used tool used by cardiologists and non-cardiologists for decades and provides a window into the physiological and structural condition of the heart. However, at present there is no tool available that can produce a 3D representation of the heart from ECG data, showing structural conditions of the heart and predicting heart conditions based off this new information. Moreover, artificial intelligence (AI) has been utilized to detect ECG signatures and patterns that are unrecognizable by the human eyes for predicting cardiovascular diseases. None of these have utilised the potential of 3D visualisation of heart.

Our group has developed a prototype software tool to visualise the electrical signature of heart. This tool allows for visualisation of the electrical signal propagation in three dimensions by projecting onto vectors, planes, and mesh framework using new 3D indices produced from the ECG data. These new indices allow for a summary description of the electrical signature in a numerical format for data analysis to produce initial 3D shapes of heart signals. Also, these 3D indices, containing minimum information, can be utilised to predict variety of cardiovascular diseases using appropriate machine learning models along with identifying the location where damage is caused by the heart condition. Therefore, this project aims to develop machine learning models to identify cardiovascular diseases using ECG data in conjunction with 3D visualisation to locate damaged areas of the heart. For this, we will utilise ECG data (~2 million ECG records from ~1.5 individual patients) from the Federal University of Minas Gerais (Brazil). Furthermore, this project will also explore the possibility whether this technique can further be utilised in portable ECG monitors for self-diagnosis of heart problems, which will lead to patient empowerment that is particularly important to reduce the stress that is regularly put on the NHS in recent times.

Objectives:

Developing robust machine learning models using 3D metrices from ECG data to predict cardiovascular diseases.

Integrating the machine learning model with the 3D visualisation tool to identify the location of abnormality in the heart.

Exploring the feasibility of implementing the framework using portable ECG monitors for self-diagnosis.

This project is a collaboration between Aberystwyth University, The Reading Guide Company, and Federal University of Minas Gerais (Brazil).

Project title: Optimisation Heuristics and Exploratory Landscape Analysis in Combinatorial Optimisation

1st supervisor: Dr Christine Zarges
2nd supervisor: Dr Thomas Jansen

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T2 - biological, health and clinical sciences T3 - novel mathematical, physical and computer science approaches

Project description: Over the last decades, general optimisation heuristics such as evolutionary algorithms have successfully been applied to a wide range of tasks in real-world applications [1]. At the same time, significant theoretical advances that shed light on their fundamental working principles have been made [2]. This project will further improve the applicability of such optimisation heuristics. It will build upon recent momentum and progress in both, theory and applications of such heuristics and further contribute to bridging the gap between these two branches of the research area [3]. It will make use of Exploratory Landscape Analysis [4], a class of methods used to determine properties of optimisation problems to solve the so-called Algorithm Selection problem, i.e., the problem of selecting the best suited optimisation heuristic for a given problem. Important goals of the project include the identification of correlations and gaps in correlation between features, landscapes, and algorithm performance, the discovery of further landscape features and the development of a recipe for designing an optimisation heuristic based on these landscape features. The focus is on combinatorial optimisation problems.

[1] Annual "Humies" Awards For Human-Competitive Results Produced By Genetic And Evolutionary Computation: https://www.human-competitive.org

[2] Doerr, B., and Neumann, F., editors (2020). Theory of Evolutionary Computation - Recent Developments in Discrete Optimization. Springer. http://doi.org/10.1007/978-3-030-29414-4

[3] COST ACTION CA15140: http://imappnio.dcs.aber.ac.uk

[4] Mersmann, O., Bischl, B., Trautmann, H., Preuss, M., Weihs, C. and Rudolph, G. (2011). Exploratory landscape analysis. In Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation (GECCO), pages 829–836, New York, NY, USA, ACM.

Project Title: Self-assessment in robotic imitation learning processes for robotic assembly tasks

1st supervisor: Dr Fei Chao
2nd supervisor: Dr Changjing Shang

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: Robotic assembly tasks require robots to own sufficient action planning abilities, while imitation learning is regarded as a crucial approach to transferring human action planning abilities to various types of robot. State-of-the-art video-guided robotic imitation methods remain to involve human experts for providing sparse rewards to indicate whether robots successfully complete their specified tasks. This project attempts to enable robots to autonomously evaluate whether their actions can complete tasks and produce real-time corrections in performing their actions. It addresses a fundamentally novel approach to imitation learning solution, involving video understanding, model-based deep reinforcement learning, deep learning-based action rewards evaluation, and other deep learning modules. With the assistance of such innovative imitation techniques, robots can effectively learn new and complex skills based on human demonstrations.

When a human subject learns a new action skill, the subject has to compare their actions with task specifications and make corrections promptly. This naturally raises the question of whether robots can use a similar pattern to learn novel skills from demonstrations, leading to three important abilities reasonably expected: (1) Understanding: Extract introduction images from demonstration videos via video interpretation, segmenting a certain observed behaviour into multiple stages; (2) Learning: Identify model robotic states and actions, while learning the staged tasks step-by-step; and (3) Assessing: Evaluate the action-state achieved at the end of each stage.

his PhD project is proposed to create the first integrated approach to learning robotic multi-stage tasks directly from input video demonstrations without human involvement. In particular, the project will cover the following work packages: a literature review on representation learning and reinforcement learning techniques; an innovative design of a computational mechanism to organically combine these two learning paradigms within a unified framework; a software specification and implementation of the algorithm to realize the framework; and an experimental comparison against state-of-the-art methods over both standard benchmark datasets and a certain application case study to be determined (with respect the interest of the candidate).

References

1. Q. Li, Z. Guo, F. Chao*, X. Chang, L. Yang, C.-M. Lin, C. Shang, Q. Shen. Robotic Trajectory Sequential Writing Problem via Learning Character's Structural and Sequential Information. IEEE Transactions on Cybernetics. Accepted, 2022

2. M. Zhang, S. Vikram, L. M. Smith, P Abbeel, M. J. Johnson, S. Levine. SOLAR: Deep Structured Representations for Model-Based Reinforcement Learning. ICML 2019: 7444-7453

3. L. M. Smith, N. Dhawan, M. Zhang, P. Abbeel, S. Levine. AVID: Learning Multi-Stage Tasks via Pixel-Level Translation of Human Videos. Robotics: Science and Systems 2020

Project Title: Analysing behaviour patterns to predict increasing risk of accidents in elderly

1st supervisor: Dr Patricia Shaw
2nd supervisor: TBD

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T2 - biological, health and clinical sciences T3 - novel mathematical, physical and computer science approaches

Project description: To support an aging population to live independently, Aberystwyth Smart Home Lab is simulating a home environment with various sensors to monitor activity around the home, as well as developing technologies to assist with independent living. A major risk for elderly is accidents occurring whilst alone in the home such as falls. Changes in behaviour can be an early indicator of problems that could lead to an increased risk of accidents occurring. If identified, interventions by family and social care support can help to address issues and thereby bring the risk of accidents back down again.

Using AI and machine learning techniques, this project will analyse data to recognise daily routines and detect increasing changes is behaviour to evaluate the potential risk of accidents occurring.

Project title: Understanding Deep Learned Features in Breast Cancer

1st supervisor: Professor Reyer Zwiggelaar
2nd supervisor: TBD

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T2 - biological, health and clinical sciences T3 - novel mathematical, physical and computer science approaches

Project description: We are interested in developing an enhanced understanding of deep learning approaches in computer aided diagnosis, especially with regard to mammography and histology (which is the main imaging modality for the early detection of breast cancer).

To facilitate this, we are expecting to look at the links between deep learned features and more traditional image processing information (the latter can be linked to morphological features of the mammographic abnormalities (which are used by radiologist for classification).

As a second part of the project, we will also investigate links between deep learned mammography and histology features. This would be to investigate how the morphology of breast cancer can be linked across different image resolutions (e.g. tissue level for mammography versus cell level for histology).

The work will be based on existing datasets (some publicly available). We have datasets, which have a large number of samples covering various mammographic abnormalities. In addition, we have a unique datasets, which contains mammography and histology images from the same women.

The work will be in close collaboration with clinical experts, who tend to be actively involved in discussions and evaluation.

Bangor University

Project Title: Visualisation of law using Artificial Intelligence

1st Supervisor: Dr Pete Butcher,
2nd supervisor: Professor Jonathan C. Roberts,
3rd supervisor: Dr Sarah Nason

Department/Institution: School of Computer Science and Electronic Engineering and School of Law Bangor University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: This project will ideate, design and programme computing solutions to apply visualisation techniques to legal advice services, underpinned by AI (Artificial Intelligence). Access to legal information and advice is a complicated and sometimes expensive process. An increasing number of legal services are being offered online, for free with the aim of making it easier for anyone to get access to the legal information they need (E.g., Resolver, Legal Utopia). Many of these systems, including those developed by us [1], require massive amounts of time and effort in the background, scouring legal code to produce simplified flow diagrams describing the connections between legal bodies and their processes. Artificial Intelligence (AI) techniques have the potential to help capture, organise, and analyse this material automatically. The use of artificial intelligence (AI) can help reduce the workload required by the humans in the loop behind the scenes. Data-visualisation has potential to elicit salient information to the public. For example, current legal visualisation research has tended to focus on litigation and private and commercial issues, whereas visualising public law (the law that governs relationships between individuals and public bodies) and the nature of law as pathways and visualising the reasoning or decision-making processes inherent to law, and the stories of those subject to law, remain under-explored aspects of legal design.

The work will design and implement data-visualization solutions that will help laypeople and policy makers alike understand the systems in which they are operating. The work will involve working as a team, and speaking with lawyers, ombudsmen, legal services, advisors etc. It will require the development of mathematically driven AI analysis solutions, and human-in-the-loop computing interfaces.

J. C. Roberts, P. Butcher, A. Sherlock, and S. Nason. "Explanatory Journeys: Visualising to Understand and Explain Administrative Justice Paths of Redress", IEEE Trans. on Vis. and Computer Graphics, vol. 28 1, 2022, pp. 518-528, DOI: 10.1109/TVCG.2021.3114818

Project Title: Artificial Intelligence in IoT Full stack (Internet of Things) for Environmental Science

1st Supervisor: Dr Noel Bristow
2nd supervisor: Dr Panagiotis Ritsos

Department/Institution: School of Computer Science and Electronic Engineering, Bangor University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project Description: Modern IoT (Internet of Things) data capture, storage and visualisation confront the developer and analysts with severe data challenges: massive volumes, rapidly changing, temporal, spatial, contaminated, wide, imbalanced (and so on). But the use of IoT environmental data has huge opportunities to help scientists understand the natural world, how the climate is changing, and influence policy makers [1-3]. This project will seek to develop novel and effective solutions where standard approaches are ineffective. The work will investigate theoretical and practical solutions that will transfer across from computing to application domains in computer science and environmental science and oceanography. One solution could focus on solutions to weakly supervise learning, within a full-stack framework. This could be used to extract actionable insights from the digital streams, reducing the volume of data stored, and enhancing visualisation. Consequently, the work requires expertise in coding but also full-stack development. We will work with sensors, and Bangor's network of LoRaWAN gateways linked to The Things Network, developing IoT Full Stack solutions that combine the use of Front End (e.g., HTML, CSS, JavaScript) with Back End development (e.g., Python, PHP) with Internet of Things (firmware systems, networking protocol, and sensors) underpinned with AI learning solutions. Experiments will be carried out to explore learnt analysis of techniques to highlight and store salient features and learnt behaviour and visualise the results to the end-user. We will work alongside various partners at Bangor University: electronics engineers to develop IoT sensors; and environmental scientists to enable the development of robust solutions for the challenges they face. This project, by developing an intelligent IoT pipeline, enhanced with AI, will provide the infrastructure to enable these challenges to be met.

[1] Zhang, S., Bristow, N., David, T. W., Elliott, F., O'Mahony, J., & Kettle, J. (2022). "Development of an organic photovoltaic energy harvesting system for wireless sensor networks; application to autonomous building information management systems and optimisation of OPV module sizes for future applications", Solar Energy Materials and Solar Cells, 236, 111550.

[2] A. M. F. Rigby, P. W. S. Butcher, P. D. Ritsos, and S. D. Patil, "LUCST: A novel toolkit for Land Use Land Cover change assessment in SWAT+ to support flood management decisions", Environmental Modelling & Software, no. 105469, Aug. 2022.

[3] Bristow, N., Rengaraj, S., Chadwick, D., Kettle, J., & Jones, D., "Development of a LoRaWAN IoT Node with Ion-Selective Electrode Soil Nitrate Sensors for Precision Agriculture", Sensors, (accepted November 2022).

Project Title: AI driven data-visualisation design and storytelling

1st Supervisor: Prof. Jonathan Roberts
2nd Supervisor: Dr Pete Butcher
3rd supervisor: Dr Panagiotis Ritsos,

Department/Institution: School of Computer Science and Electronic Engineering, Bangor University

Research theme: T3 - novel mathematical, physical and computer science approaches

Description: Visualising data is a creative process yet challenging. Showing the most important results, highlighting the most significant correlations, and displaying data in an effective way can be challenging for any researcher. Often researchers create visualisations and make presentations (telling stories with their data) to highlight the important lessons learnt. Our recent work has focused on design [1], layout techniques [2], and telling visualisation stories [3]. AI algorithms have the potential to extract salient features and remove effort of design and storytelling presentation. Potentially exemplars can be used to train the system. Teachers often explain and describe best practices. But while there are many examples, this is tacet knowledge and it is difficult to articulate what 'best' means. Local search solutions and weak supervision can help. Label placement algorithms use local searches (such as performed through a Tabu search), and we semi-automatically present stories of covid data, highlighting salient features [3]. But deep learning has the potential to go further, and suggest appropriate designs, guide researchers in design decision making, re-engineer design solutions, highlight salient features and provide appropriate visualisations, inspire design ideation processes, and so on. We will investigate weak supervision, semi-supervised learning, transudative learning and deep learning. Focus will be on both automatic design and the presentation of the information in automatic storytelling. Learning from current exemplar research, it will research algorithms, metrics, and methods to help researchers design visualisations and use them to tell effective stories with their data.

[1] Roberts, J. C., Headleand, C., & Ritsos, P. D. (2015). Sketching Designs Using the Five Design-Sheet Methodology. IEEE Transactions on visualization and computer graphics, 22(1), 419-428.

[2] Chen X, Zeng W, Al-Maneea HMA, Roberts JC, Chang R. (2021). Composition and Configuration Patterns in Multiple-View Visualizations. IEEE Transactions on visualization and computer graphics. 27(2):1514-1524.

[3] M. Chen, A. Abdul-Rahman, D. Archambault, J. Dykes, A. Slingsby, P. D. Ritsos, T. Torsney-Weir, C. Turkay, B. Bach, R. Borgo, A. Brett, H. Fang, R. Jianu, S. Khan, R. S. Laramee, P. H. Nguyen, R. Reeve, J. C. Roberts, F. Vidal, Q. Wang, J. Wood, and K. Xu, "RAMPVIS: Answering the Challenges of Building Visualisation Capabilities for Large-scale Emergency Responses," Epidemics, vol. 39, no. 100569, Jun. 2022.

University of Bristol

Project title: Searching for New Physics with the CMS experiment at the LHC

1st Supervisor: Sudan Paramesvaran
2nd Supervisor: Henning Flaecher

Department/Institution: School of Physics, University of Bristol

Research theme: T1: data from large science facilities

Project description: The CMS experiment at the Large Hadron Collider has already made ground-breaking discoveries since it started operating in 2010. However, the long awaited Run 3 of the LHC has now started, and is set to double the entire dataset currently collected by CMS. This will lead to an unprecedented amount of data which will need to be carefully scrutinised to make sure we leave no stone unturned in our search for new physics. In particular, the search for signatures which exhibit large amounts of missing transverse momentum - a characteristic of particles escaping undetected - are especially interesting as they cast a wide net for new physics, including dark matter candidates and more exotic models such as split-susy. In this project we will develop machine learning (ML) algorithms to significantly enhance the discovery prospects from our newly collected data; a variety of different ML techniques will be studied to sift through the hundreds of millions of collision events looking for signs of new physics. These algorithms can be applied at all stages of analysis from trigger and reconstruction to event selection and background estimation.

Project title: Identifying Higgs Bosons with Machine Learning at the LHC

1st Supervisor: Sudan Paramesvaran
2nd Supervisor: Jim Brooke

Department/Institution: School of Physics, University of Bristol

Research theme: T1: data from large science facilities

Project description: The CMS detector has been operating at the Large Hadron Collider for a decade, resulting in a wide range of ground-breaking measurements, including the discovery of the Higgs boson. The volume of data produced by CMS is enormous, at 40TB/s, far in excess of what can be recorded on disk. A "trigger" system is therefore required, which analyses and selects proton collisions as they are produced, to ensure crucial physics is stored for later analysis. An upgrade of the system is currently underway, which will increase the processing power of the system by over an order of magnitude. The system must cope with the huge increase in data expected from the high-luminosity LHC, which begins operation in 2029. This project will focus on developing Machine Learning algorithms for the future CMS trigger system, assessing their performance using simulations, as well as real data already recorded by CMS. These algorithms could have a significant impact on the physics programme of the high-luminosity LHC.

Project title: Fast Inference for Online Data Selection at DUNE

1st Supervisor: Jim Brooke
2nd Supervisor: Sudan Paramesvaran

Department/Institution: School of Physics, University of Bristol

Research theme: T1: data from large science facilities

Project description: The ProtoDUNE detector, currently operating at CERN, is a small-scale (800 tonne) prototype of the future 70,000 tonne Deep Underground Neutrino Experiment (DUNE). When complete, DUNE will measure neutrino parameters to unprecedented precision, casting light on the neutrino mass hierarchy and the matter/anti-matter asymmetry in the Universe. Both DUNE and protoDUNE use liquid Argon as a target, instrumented as a time project chamber (TPC), to measure particle interactions with exquisite precision. The data produced by the TPC is extremely well-suited to reconstruction via image recognition techniques using machine learning. The Bristol DUNE group is studying $\pi 0$ identification and reconstruction at protoDUNE, with a view to measuring the pion charge exchange cross-section ($\pi+ \rightarrow 1\ \pi 0$ + X). These studies are motivated by the importance of $\pi 0$ backgrounds to electron neutrino measurements at DUNE. Our analyses and solutions currently use traditional rules-based methods. This project will build on the established understanding of the problem within the Bristol group, to extend and complement current solutions with machine learning techniques. The problem lends itself well to studying a range of approaches and has numerous extensions in other areas of DUNE physics.

Project title: Coevolution of galaxies and supermassive black holes in the Euclid era

1st Supervisor: Sotiria Fotopoulou
2nd Supervisor: Malcolm Bremer

Department/Institution: Department of Physics, University of Bristol

Research theme: T1: data from large science facilities and T3 - novel mathematical, physical and computer science approaches

Project description: Black holes with masses more than a billion times the mass of the Sun lurk in the heart of most galaxies. During the so-called active phase of accretion, accumulated gas and dust around the black hole shine across the electromagnetic spectrum from the radio to the gamma rays. These sources are called Active Galactic Nuclei (AGN) or quasars, depending on their luminosity. We currently believe that there is likely a coevolution between the growth of these supermasive black holes and the stars in their host galaxy. However, the details of the coevolution remain unclear, and under vivid debate.

This project will leverage the upcoming optical and near infrared space telescope, Euclid, scheduled for launch in July 2023. The Data-Intensive AstroNomical Analysis (D.I.ANA.) group in Bristol is involved in the preparation of the data analysis of the mission (e.g. Fotopoulou & Paltani, 2018). We are particularly interested in the properties of the AGN population, and their evolution across cosmic time. We use traditional methods and a variety of machine learning approaches to identify and characterise the population of all Euclid sources (stars, galaxies, AGN), which are expected to be more than one billion.

The Euclid Consortium in an international team and comprises more than 2,000 scientists and engineers. With this PhD project you will become a member of the Euclid Consortium as the first data come in! The focus of this project will be to use Euclid and other telescopes to identify low luminosity AGN, and to compare these observations with leading models and simulations. We are also keen to exploit the data to identify peculiar populations of galaxies and look for unexpected discoveries.

References:

Fotopoulou & Paltani 2018 A&A, 619A, 14F, arXiv:1808.04977

Stevens et al., 2021 JOSS, 6, 65, arXiv:2109.05207

Selwood et al., 2022, arXiv:2210.04827

Project title: Accelerating event simulation with machine learning for the LHCb experiment

1st Supervisor: Kostas Petridis
2nd Supervisor: Jonas Rademacker

Department/Institution: School of Physics, University of Bristol

Research theme: T1: data from large science facilities

Project description: The proposed project aims to resolve one of the main issues that severely hinder future iterations of particle physics flavour experiments like the Large Hadron Collider beauty (LHCb) experiment.

The vast signal samples expected to be collected by the planned upgrade of the LHCb experiment require even larger amounts of simulation to match the statistical precision of the data. Current approaches to event simulation are computationally too slow and do not lend themselves to parallelisation. The main challenge stems from having to simulate billions of events containing tens of thousands or particles produced at the collision point that subsequently all interact with the various sub-detectors of LHCb. The proposed research aims at using hardware acceleration — GPUs, as well as novel architectures like Intelligence Processing Units (IPUs) [1] — to train Deep Generative Networks to reproduce fully simulated samples with high fidelity but at a fraction of the time and in a highly parallelisable fashion. The project will build on the method to simulate muons interacting with matter using Generative Adversarial Networks established by a previous CDT studentship on the SHiP experiment [2] using Generative Adversarial Networks, but extensively generalised in order to cope with tens of thousands of particles and multiple subdetector systems.

References: (If applicable, 5 max)

[1] https://link.springer.com/article/10.1007/s41781-021-00057-z

[2] https://iopscience.iop.org/article/10.1088/1748-0221/14/11/P11028

Project title: Machine Learning to find New Physics in Muon Decays

1st Supervisor: Joel Goldstein
2nd Supervisor: TBD

Department/Institution: School of Physics, University of Bristol

Research theme: T1: data from large science facilities

Project description: The Mu3e experiment at PSI will look for extremely rare muon decays; in particular it is designed to try to identify the lepton flavour-violating decay of a muon to three electrons at the level of one event in 1016. The experiment will use the latest advances in detector technology to identify electrons with high spatial and temporal resolution, and advanced pattern recognition algorithms will be implemented electronically to filter the data in real time.

In this project the student will apply the latest developments in machine learning to Mu3e event reconstruction and filtering, developing new techniques that could be faster, more flexible and/or more effective than conventional algorithms. This could lead not only to the optimisation of the physics reach for the three-electron channel, but also the capability to perform real-time detailed analysis to look for different signatures. The student will start by developing and optimising algorithms in simulation, and then will have the opportunity to commission and test them in early data from the running experiment.

Cardiff University

Project title: AI optimization of optoelectronic devices for gravitational-wave detectors

1st Supervisor: Dr Nicolás Abadía (early-career researcher)
2nd Supervisor: Prof Katherine Dooley
3rd Supervisor: Dr Federico Liberatore (early-career researcher)

Department/Institution: School of Physics and Astronomy, Cardiff University

Research theme: T1: data from large science facilities. Note: The objective of this Ph.D. is to develop the proof-of-concept of a device that will be used in a Michelson interferometer at Dooley's group with the potential to be employed in large science facilities to produce data.

T3 - novel mathematical, physical and computer science approaches. Note: This Ph.D. will combine applied mathematical approaches to physics modelling to optimize optoelectronic devices, including thermo-opto-electrical modelling, which was not done before.

Project description: This project will apply artificial intelligence algorithms to optimize optoelectronic devices to meet the demanding specifications of laser interferometers for detecting gravitational waves and other fundamental physics phenomena.

The typical method of designing such devices is by guessing an initial structure based on previous experience and hand-tuning its parameters to optimize them using models and simulators. It is time-consuming to explore all the design space to improve performance manually. Nevertheless, genetic algorithms, level-set methods, non-linear search algorithms, and other Artificial Intelligence techniques can do it.

If the design space can be optimized automatically, we can enlarge the device space significantly by adding complexity to the device structure. If we have a large enough design space, there will likely be solutions with substantially better performance than small manually-optimized sets.

A relevant example is the work in [1]. By only using dielectrics in manually designed 3D structures, including lenses and mirrors, it is impossible to confine light in a spot below the diffraction limit of light. Nevertheless, in [1], they use a tolerance-constrained topology optimization method to produce an unprecedented only-dielectric cavity that can confine light well below the diffraction limit.

In this Ph.D. studentship, you will employ and modify artificial intelligent algorithms to optimize a tuneable lens on a chip with extremely low loss (<1%) to have adaptative mode matching for optical cavities used in laser interferometers. This multidisciplinary project involves optoelectronic device modelling in Abadia's group, optimization algorithms in Liberatore's group, and quantum-enhanced laser interferometers in Dooley's group.

You will use in-house modelling tools and commercial simulators in Abadia's group [2,3] to model tunable lenses to meet the stringent specifications required for gravitational-wave detectors. You will collaborate with Dooley's group [4] at the School of Physics and Astronomy to integrate the lens design to a squeezed-light-enhanced table-top laser interferometer. To improve the lens figure of merit ~100x times, you will drive the modelling tools with an artificial intelligence algorithm to increase the design space and find the fittest solution. You will collaborate with Liberatore's group [5] in the School of Computer Science and Informatics to develop and integrate the artificial intelligence optimization strategy with the modelling tools.

You will have the opportunity to fabricate the optimized device at the Institute for Compound Semiconductors and test it in a complete instrument in Dooley's group.

For further information on this role, don't hesitate to get in touch with Dr Nicolás Abadía at abadian@cardiff.ac.uk. You can follow us on https://www.researchgate.net/profile/Nicolas-Abadia and https://linkedin.com/in/nicol%C3%A1s-abad%C3%Ada-9032a422a

References:

[1] Albrechtsen, M., Vosoughi Lahijani, B., Christiansen, R.E. et al. Nanometer-scalephoton confinement in topology-optimized dielectric cavities. Nat Commun 13, 6281(2022).

[2] Joe Mahoney, Mingchu Tang, Huiyun Liu, and Nicolás Abadía, "Measurement of thequantum-confined Stark effect in InAs/In(Ga)As quantum dots with p-doped quantum dot barriers," Opt. Express 30, 17730-17738 (2022). - https://doi.org/10.1364/OE.455491

[3] Md Ghulam Saber, David V. Plant, and Nicolás Abadía, "Broadband all-silicon hybrid plasmonic TM-pass polarizer using bend waveguides," AIP Advances 11, 045219 (2021). -https://doi.org/10.1063/5.0044490

[4] Sander M. Vermeulen, ..., Katherine L. Dooley and Hartmut Grote. "An Experiment for Observing Quantum Gravity Phenomena using Twin Table-Top 3D Interferometers" Class. Quant. Grav. 38, 085008 (2021).

[5] F. Liberatore et al. "A hierarchical compromise model for the joint optimization of recovery operations and distribution of emergency goods in Humanitarian Logistics," Computers & Operations Research 42, (2014) 3-13.

Project title: Predictive Emergency Service Operations Planning

1st Supervisor: Dr Federico Liberatore
2nd Supervisor: Prof Alun Preece

Department/Institution: Cardiff University COMSCM

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: Aims: The objective of this project is the exploration of the frontier between demand forecasting and operations planning to devise predictive, intelligent, and adaptive strategies in the context of emergency services (e.g., policing, fire services, medical services). More specifically, the research will follow and combine two main lines. The first concerns the furthering of spatiotemporal prediction models of the demand that adjust to the idiosyncrasies of the application framework by exploring the implication of context-specific variables (e.g., micro/macroeconomy, population distribution, environmental characteristics) while leveraging societal issues (i.e., the bias in the historical demand distribution due to racial or economic factors). The second regards the definition of optimisation models specifically tailored to the necessities of emergency services to devise predictive action plans. This research will be hosted in the Crime and Security Research Institute with connections to UK/international police force partners and other emergency services. As such, the models will be tested on real-world datasets.

Methods: This project draws from the disciplines of Machine Learning (ML) and Operational Research. In broad strokes, the project will be structured as follows. i) Analysis of the context and problem to solve. ii) Literature review for the identification of improvement opportunities with regards to the problem addressed. iii) Definition of mathematical models that solve the problem and expand on the literature. In this regard, the project is focused on (but not limited to) classical ML models, neural and deep models, and mathematical programming models. iv) Development and validation of the solution methodology.

Deliverables: The results of the project will be mostly published as journal papers or as conference papers when appropriate. Dissemination will be carried out at national/international top-tier conferences through talks or posters. Successful models will be implemented and developed into working software to be used by emergency services.

Keywords: Forecasting, Deep Learning, Machine Learning, Location Analysis, Vehicle Routing Optimisation, Districting Design, Operational Research.

References:
Camacho-Collados, M. and Liberatore, F., 2015. A decision support system for predictive police patrolling. Decision Support Systems, 75, pp.25-37.

Camacho-Collados, M., Liberatore, F. and Angulo, J.M., 2015. A multi-criteria police districting problem for the efficient and effective design of patrol sector. European journal of operational research, 246(2), pp.674-684.

Liberatore, F., Camacho-Collados, M. and Vitoriano, B., 2020. Police districting problem: Literature review and annotated bibliography. Optimal Districting and Territory Design, pp.9-29.

Project title: Kinetic Monte Carlo and Machine Learning for Magnetic Monopole Dynamics

1st Supervisor: Dr Felix Flicker (Cardiff)
2nd Supervisor: Prof. Biagio Lucini (Swansea)
3rd Supervisor: Prof. Sean Giblin (Cardiff)

Department/Institution: Cardiff University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: Why can't you pull the north pole off a magnet? If you try to cut a magnet in half you get two smaller magnets, each with a north and south pole. Repeat this process and you ultimately end up with a single atom with a magnetic moment, which again has both poles. Essentially, the answer is that while there are particles with electric charge (electric 'monopoles', such as electrons) there are no magnetically charged particles -- magnetic monopoles -- in our universe.

However, there is growing evidence that this is not the whole story [1]. In 1997 a mysterious new type of magnet was discovered, called a spin ice. Our leading theories suggest that the magnetic fields of individual ions in these crystals orient themselves so as to create local sources and sinks of magnetization, which bear a striking resemblance to magnetic monopoles. Recent numerical [2] and experimental work [3] indicates that magnetic monopoles produce unique signatures in the fluctuations of the magnetic field around spin ice crystals, signatures known as 'pink noise' [4].

This project is aimed at understanding how monopoles generate pink noise; how that understanding might be used to definitively prove the monopoles' existence; and how we might put this understanding to use in developing technological applications. Topics will include noise spectra and the detailed study of stochastic signals; modelling and understanding fractals (bond-percolation clusters); loop erased random walks; out-of-equilibrium dynamics; and glassy behaviour.

The work will require the development of advanced numerical methods. We will apply kinetic Monte Carlo (kMC) for modelling all-to-all interactions between large numbers of spins. The project will make heavy use of the HAWK supercomputing facility. Additionally, we will develop a machine learning approach to identifying magnetic monopoles in dynamical measurements, trained using the kMC numerics. The numerical methods are computationally intensive, and will require the development of new methods for running efficiently on the latest heterogeneous high-performance computing architectures (including GPUs and FPGAs). There is scope to undertake analytical modelling to support the numerics, and the project will involve working closely with experimental and numerical groups in Cardiff, Swansea, Oxford, and Cornell.

The project includes substantial external funding for personal development, including public speaking and media training. There is also funding available for 3 months' salary to undertake an industrial placement if desirable. Both are part of the supervisory team's commitment to Responsible Innovation.

References:

[1] See Dr Flicker's Royal Institution lecture for an overview: youtube.com/watch?v=S3xH97Su-KY
[2] F. K. K. Kirschner, F. Flicker et al., Physical Review B 97, 140402(R) (2018)
[3] R. Dusad, ... F. Flicker et al., Nature 571, 234 (2019)
[4] F. Flicker, Science, accepted for publication

Project title: Exploring molecular cloud formation with machine learning

1st Supervisor: Sarah Ragan
2nd Supervisor: Nicolas Peretto
3rd Supervisor: Paul Clark

Department/Institution: Cardiff University, School of Physics and Astronomy

Research theme: T1: data from large science facilities and T3 - novel mathematical, physical and computer science approaches

Project description: Stars form in dense molecular clouds throughout the interstellar medium (ISM). Although observational studies have established what the conditions necessary for star formation are, very few constraints exist on how, why, and where those conditions arise in the Galaxy. Our current poor understanding of the molecular cloud formation process is one of the major barriers to a predictive theory of star formation.

Observations provide only a snapshot in time of the ISM, and regions undergoing cloud and star formation can be difficult to pick out using traditional methods. Because of this we lack a reliable timeline and global census of the clouds the Galaxy has formed.

This project will use wide-field surveys of the Milky Way plane in a variety of line tracers (e.g. CO, HI, HCN) to map the distribution of clouds in the earliest phase of formation. Artificial intelligence techniques will be used to develop a holistic picture of the distribution, stability and internal kinematics of molecular clouds, and their relation to their local environment in the Galaxy. Through the comparison with synthetic observations produced by numerical simulations (provided by P. Clark), we will utilise machine learning techniques to develop an evolutionary sequence of molecular clouds, providing a potentially profound insight into current theories of molecular cloud formation and evolution.

The PhD is primarily as observational astrophysics project, requiring the development of ‘big data’ handling skills, programming and machine learning. You will collaborate with other observers and theorists in the Cardiff Hub for Astrophysics Research and Technology.

References:

Ragan et al. (2018), MNRAS, 479, 2361

Urquhart et al. (2021), MNRAS, 500, 3050

Project title: Combining machine learning and physical intuition to understand the properties of black hole mergers

1st supervisor: Stephen Fairhurst
2nd supervisor: Vivien Raymond

Department/Institution: Cardiff University

Research theme: T1: data from large science facilities

Project description: The LIGO and Virgo detectors have observed 90 gravitational wave signals from merging black holes and/or neutron stars. In spring 2023, they will begin their fourth observing run, with increased sensitivity and the expectation of observing several black hole mergers per week. This drives a need to develop fast and accurate methods for inferring the physical properties of the observed systems. The focus of this PhD project will be to develop machine-learning based methods to perform this rapid parameter estimation. We will incorporate key physical insights, such as how spin-induced orbital precession and higher gravitational wave multipoles impact the waveform, to guide the development of the machine learning algorithms.

The PhD project will also require active involvement in the analysis of data taken during the fourth LIGO-Virgo-KAGRA observing run.

References:

GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run, https://arxiv.org/abs/2111.03606

The population of merging compact binaries inferred using gravitational waves through GWTC-3, https://arxiv.org/abs/2111.03634

Measuring gravitational-wave higher-order multipoles https://doi.org/10.1103/PhysRevD.103.024042

Identifying when precession can be measured in gravitational waveforms https://doi.org/10.1103/PhysRevD.103.124023

Project title: Development of brain artery stiffness measures using machine learning and multi-scale computational models

1st Supervisor: Prof Kevin Murphy
2nd Supervisor: Dr Ian Driver
3rd Supervisor: Dr Leandro Beltrachini

Department/Institution: School of Physics and Astronomy, Cardiff University

Research theme: T2 - biological, health and clinical sciences and T3 - novel mathematical, physical and computer science approaches

Project description: The heart constantly pumps blood to the brain with great force. Fortunately, the brain's blood vessels are elastic, providing protection by soaking up energy; a process called buffering. Age-related arterial stiffening is a leading cause of neurological problems. The excessive cardiac pulsatile energy reaching the brain's capillary bed damages the blood brain barrier. Mounting evidence suggests a strong causal link between reduced arterial elasticity in the body and brain disorders such as dementia. This research project will help develop ways to measure how, over a lifetime, the force of the heartbeat damages the vessels, causes stiffening and reduces buffering, leading to neurological problems.

This PhD project is well suited to students who would like to use their knowledge of computational modelling to inform our understanding of human brain blood flow and how it can go wrong with age and disease. It will be embedded in a large £1.8m 5-year research programme supported by Wellcome entitled "The Beat Goes On: imaging brain artery stiffness and the damage caused by poor buffering of pulsatile blood flow".

Magnetic resonance imaging (MRI) is a versatile tool for measuring the properties of brain blood vessels at multiple scales: in the large feeding arteries and in the microvasculature. In the first half of the PhD, you will further develop our specialised magnetic resonance (MR) imaging techniques that measure large vessel stiffness, integrating machine learning methods to improve information quality. You will extend our recently patented DIMAC technique (Dynamic Inflow MAgnitude Contrast) by learning MR pulse sequence programming. Adding radial readouts (with help from Oxford collaborators) and compressed sensing techniques (with help from Cardiff collaborators) will speed up efficient data acquisition.

In the second part of the PhD, you will integrate previously collected eye artery measurements into a computational statistical model of the brain's smaller blood vessels. You will use this to simulate MR signals that we measure with a different type of scan (BOLD fMRI). Using this model of the human cortical vasculature, we will be able to optimise MR sequences to detect pulsatility in smaller blood vessels. You will have the chance to visit international collaborators in Utrecht to work on this aspect.

Throughout the PhD, you will have the opportunity to apply the techniques that you have developed to real brain imaging data, some which our lab is collecting and some from publicly available large datasets (e.g. the UK Biobank), to help us determine how brain arterial stiffness progresses with age.

You will be registered in the School of Physics and Astronomy and the project will take place in CUBRIC (Cardiff University Brain Research Imaging Centre). CUBRIC houses ~200 researchers and hosts state-of-the-art neuroimaging equipment including 4 human MRI systems and a computing cluster including a Nvidia DGX-1 server optimised for machine learning applications. This MR equipment combination is unique in Europe offering an excellent work environment with researchers across a diverse range of MR disciplines.

References:

Whittaker, J.R., Fasano, F., Venzi, M., Liebig, P., Gallichan, D., and Murphy, K. (2022). Measuring arterial pulsatility with Dynamic Inflow MAgnitude Contrast (DIMAC). Frontiers in Neuroscience, Brain Imaging Methods 15:795749

Báez-Yáñez, M.G., J.C.W. Siero, and N. Petridou, A statistical 3D model of the human cortical vasculature to compute the hemodynamic fingerprint of the BOLD fMRI signal. bioRxiv, 2020: p. 2020.10.05.326512

Okell, T.W., M. Garcia, M.A. Chappell, J.V. Byrne, and P. Jezzard, Visualizing artery-specific blood flow patterns above the circle of Willis with vessel-encoded arterial spin labeling. Magn Reson Med, 2019. 81(3): p. 1595-1

Swansea University

Project title: Data Lab Cymru

1st Supervisor: Dr Jon Kennedy
2nd Supervisor: Prof Sinead Brophy

Department/Institution: Swansea University Medical School

Research theme: T2: biological, health and clinical sciences

Project description: There are three projects choose from which are part of the ongoing activities of the Data Lab Cymru. The group works in population heath and has collaborations internationally and across the UK. The main areas of activities include:

MuM-PreDicT (https://mumpredict.org/)

This study examines the influence of multimorbidity in pregnancy. This is a nationwide collaboration across England, Wales, N. Ireland and Scotland and involves 7 universities. The work involves examining clustering of multimorbidity, the trajectories leading to the clusters, and the outcomes associated with the different clusters. The applicant would utilise data from the SAIL databank, CPRD with the possibility of accessing data in N. Ireland and Scotland.

Arthritis

Examining the longer-term implications of COVID in terms of health care utilisation and undiagnosed disease and long-term implications of a diagnosis during the pandemic.

Early years

Linking Census 2011 and Census 2021 data with health data (GP, hospital, A&E) and administrative data (education, social care) to examine the influence of the occupation of the parent on health and attainment of the child (e.g., shift working/NHS working, army/distance workers).

Using imputation techniques to use data from surveys to impute ‘missing data’ not collected in routine data, such as income or self-assessed mental wellbeing.

The applicant will work with the Data Lab Cymru team to utilise machine learning techniques to incorporate longitudinal data across multiple sources to answer questions set by public health Wales and Welsh Government.

Project title: Tests of the dark sector with gravitational waves

1st Supervisor: Ivonne Zavala
2nd Supervisor: Biagio Lucini

Department/Institution: Physics & Maths, Swansea University

Research theme: T1: data from large science facilities and T3: novel mathematical, physical, and computer science approaches

Project description: The recent discovery of gravitational waves from binary black holes and neutron stars has open up new possibilities to solve some of the greatest mysteries of the universe: the nature dark energy, dark matter, and gravity at the highest scales.

The project's research involves both theoretical and computational development aimed at developing tests of the dark sector, DM & DE, using gravitational waves, focusing on the science of the Einstein Telescope (ET). In particular, the project will develop tests of the dark sector in these theories - namely dark matter and dark energy - using gravitational waves. It will use and develop analytical techniques, as well as numerical methods, combined with suitable computational tools, to 1) test dark matter candidates, such as primordial black holes, ultralight scalars or vector fields, as well as possible dark matter particles accreting on compact objects; 2) test the nature of dark energy and possible modifications of GR at cosmological distances, using gravitational wave observations.

It is expected that the student will join the ET collaboration as a member of the Swansea Research Unit, as well as the Swansea-Liverpool node of the European Consortium for Astroparticle Theory (EuCAPT).

Project title: Using Machine Learning to understand Lattice QCD Data

1st Supervisor: Prof Chris Allton
2nd Supervisor: Prof Gert Aarts
3rd Supervisor: Prof Tim Burns

Department/Institution: Physics/Swansea

Research theme: T1: data from large science facilities

Project description: Lattice QCD provides a first-principles approach to relate the basic degrees of freedom in the theory of the strong interactions (QCD), namely quarks and gluons, to observables probed in elementary particle and heavy-ion collider experiments. This includes spectral quantities (hadrons, transport) as well as thermodynamic quantities (pressure, entropy, susceptibilities, order parameters). In recent years, machine learning (ML) has increased in popularity also in the lattice QCD community [1], resulting in the exploration of many ML methods to analyse large data sets, both for regression and classification tasks. In this project, we will explore these methods to data obtained in the context of QCD at nonzero temperature and the FASTSUM collaboration.

References: [1] Applications of Machine Learning to Lattice Quantum Field Theory, D. Boyda, G. Aarts et al, contribution to SNOWMASS 2022, https://arxiv.org/abs/2202.05838

Project title: Optimising Attack-Defence Trees (ADT) using Evolutionary Computing

1st Supervisor: Dr Hoang Nga Nguyen
2nd Supervisor: Prof Siraj Ahmed Shaikh

Department/Institution: Department of Computer Science, Swansea University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: Attack-Defence Trees (ADTs) are used to model threats and defences as structured hierarchies and, are increasingly being adopted by in the design and development process of modern cyber-physical systems (CPS). Enumerating the attack surface of modern CPS is a substantial challenge however given (i) the volume of interfaces and associated configurations that could be exploited, (ii) the complexity offered by system behaviours that could be subverted for advancing an attack goal, and (iii) the nature of component integration making it difficult to systematically evaluate security at a system-level. While ADTs are a useful modelling method, constructing ADTs is manually intensive, or rule-driven (where generating such rules remains an arduous task). Current methods to overcome this are either limited to higher abstraction [1], or narrow down to singular interfaces (and even then completeness could not be guaranteed given expert-driven knowledge) [2].

This research proposes concepts from the theory of economics to cybersecurity threat mitigation, aided by evolutionary computing (EC) methods to determine optimal defensive configurations. We propose algorithmic transformations of manually-constructed ADTs into cost-benefit models representing system-level threats with defensive control measures as benefits. To achieve a risk-proportionate and cost-effective defensive posture for a system, optimal control measures would be determined; fitness functions would be expressed to maximise returns and minimise opportunity costs (for comparative analysis of one control feature to other alternatives).

While a range of EC methods is available to explore, our preference would be for genetic algorithms (essentially metaheuristics inspired by natural selection) given their scale-efficient algorithmic implementations readily available and prior use for multi-criteria optimal security sensor placement problems [3]. Outputs from the optimisation step would feed into an algorithmic reconstruction of optimal ADTs.

For validation, we propose to use a real-world component for threat enumeration and system-level risk assessment to evaluate the proposed methodology for completeness and effectiveness. Given prior work by the supervisory team [4], we propose to use an automotive Over-The-Air (OTA) Update system for evaluation where the output of the research would be benchmarked against an ISO21434 risk assessment (where ISO21434 is the recently released global standard for threat analysis and risk assessment for automotive systems [5]). Both supervisors are CACSP (Certified Automotive Cyber-Security Professionals) certified by SGS-TÜV Saar, and certified risk assessors as such for ISO21434.

An EC-driven approach holds the potential to significantly overcome such a challenge, given the obscure nature of security threats (which could evade human intuition when it comes to threat modelling), the volume and diversity of attack vectors, and the efficiency gains to be had from an expert-in-the-loop intelligent approach.

The doctoral candidate would benefit (i) from this research as it sits on the interface of AI and cybersecurity, (ii) from the global academic and technical community in this area through training and events organised by the Automotive Security Research Group (ASRG) (https://asrg.io/), which both supervisors are active members of, and (iii) a range of modelling and testing tools that are supported by the newly set up Systems Security Group (SSG) at the Department of Computer Science at Swansea University.

This project will play an integral role in the growth of such research in the department, and SSG's external industrial collaborations.

References: (If applicable, 5 max)

[1] M Salfer, C Eckert: Attack Graph-Based Assessment of Exploitability Risks in Automotive On-Board Networks. ARES 2018: 21:1-21:10

[2] M Cheah, SA Shaikh, O Haas, AR Ruddle: Towards a systematic security evaluation of the automotive Bluetooth interface. Vehicular Communications 9: 8-18 (2017)

[3] Chen, H., Clark, J.A., Tapiador, J.E., Shaikh, S.A., Chivers, H., Nobles, P. (2009). A Multi-objective Optimisation Approach to IDS Sensor Placement. In: Computational Intelligence in Security for Information Systems. Adv in Int and Soft Comp, vol 63. Springer.

[4] S Mahmood, HN Nguyen, S A Shaikh: Systematic threat assessment and security testing of automotive over-the-air (OTA) updates, Vehicular Communications, Vol 35, Elsevier.

[5] ISO/SAE 21434:2021 (EN): Road Vehicles — Cybersecurity Engineering. ISO/SAE.

Additional Projects

Aberystwyth University

Project title: Deep Learning for the Environment

1st supervisor: Professor Bernard Tiddeman
2nd supervisor: TBD

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: Deep Learning has made great improvements to the performance of algorithms against a variety of machine learning challenges. While deep learning can contribute to solving many of the great challenges facing humanity, it has also become part of the problem. The increase in computing power required to train the latest AI models was shown to double every 3.4 months between 2012-2018 (https://openai.com/blog/ai-and-compute/) which amounts to a 300,000 times increase in that period (compared with only an 8 times increase that would result from a 2-year doubling period c.f. Moore's law). Training of large deep learning algorithms has been estimated to produce hundreds of thousands of tons of CO2 (https://arxiv.org/pdf/1906.02243.pdf), and while renewables are a viable option for powering data centres, an exponential increase in demand of this size would soon swamp any additional capacity provided. The algorithms used to train deep AI models often amount to little more than versions of gradient decent, and much of the improvement in performance seems to derive from larger networks, rather than smarter algorithms.

This PhD project would investigate developing more efficient algorithms for training deep learning models. This could include drawing on the large amount of pre-deep learning literature to find algorithms that could be applied in the deep learning context. It could also involve investigating unsupervised, semi-supervised, weakly supervised and few-shot approaches, which could be used to build models that generalise better with less training data, or to build more general-purpose models that could be adapted to a variety of tasks with minimal retraining. The project could also investigate using such approaches to tackle AI problems linked to the environment, such as plant phenomics (for developing climate change resistant crops) or remote sensing (for measuring and predicting the impacts of climate change).

Project title: When Fuzzy Rule Interpolation Meets Fuzzy Attribute Aggregation: Fuzzy Aggregated Model Interpolation and its Application for Assessment of Rural Healthcare Inequality

1st supervisor: Professor Qiang Shen
2nd supervisor: Dr Changjing Shang

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T2 - biological, health and clinical sciences T3 - novel mathematical, physical and computer science approaches

Project description: For complex systems modelling, especially for those that are novel in a less experienced problem domain, two extreme challenges often co-occur: On the one hand, there may not be sufficient historical data for learning the model; and on the other hand, there may exist a large number of potentially influential but intertangled attributes that have to be considered in the model. Currently, there lack data-driven modelling mechanisms to deal with these challenges, which typically escalate when the resulting model requires high interpretability. This project sets an ambitious aim to tackle such challenges, with a focused application on rural healthcare inequality evaluation.

Rural healthcare inequality becomes a novel problem at the post-pandemic era, owing to the uneven distribution of healthcare resources and services regarding different geographical regions. Given the novelty of this problem, there is limited a prior data whilst the available data tends to be imprecisely described and noisy. Hence, the knowledge base to be built is approximate and incomplete, making a suitable case for the adoption of fuzzy rule interpolation. Moreover, there may be many variables that affect the relevant decision-making processes (say, for governmental policy formation or evaluation on healthcare inequality). Thus, to be effective and efficient, aggregation of the hidden information embedded within the available data (however restricted) becomes a prerequisite. This naturally points to the need of a mathematical means to support the integration of certain domain variables, which may themselves be correlated in the first place. A natural appeal is therefore, to combine these two popular computational intelligence methods, in an effort to develop an integrated approach towards satisfactorily resolving the above-identified challenging problem.

This PhD project is proposed to create the first integrated approach to fuzzy rule interpolation involving aggregated domain attributes. In particular, it will cover the following work packages: a literature review on fuzzy rule interpolation and fuzzy attribute aggregation; an innovative design of a computational mechanism to organically combine these two techniques within a unified framework; a software specification and implementation of the algorithm to realise the framework, with respect to an identified rural healthcare inequality assessment case study; and an experimental comparison against state-of-the-art methods over both standard benchmark datasets and the particular case study.

References:

1. F. Li, C. Shang, Y. Li, J. Yang and Q. Shen. Approximate reasoning with fuzzy rule interpolation: Background and recent advances. Artificial Intelligence Review, 54:4543-4590, 2021.

2. P. Su, C. Shang, T. Chen and Q. Shen. Exploiting data reliability and fuzzy clustering for journal ranking. IEEE Transactions on Fuzzy Systems, 25(5):1306-1319, 2017.

3. H. Sabokbar, H. Mohammadi, S. Tahmasbi, Y. Rafii and A. Hosseini. Measuring spatial accessibility and equity to healthcare services using fuzzy inference system. Applied Geography, 136:102584, 2021.

Project title: Can Colloids Learn?

1st supervisor: Dr Adil Mughal
2nd supervisor: TBD

Department/Institution: Department of Mathematics, Aberystwyth University

Research theme: T3 - novel mathematical, physical and computer science approaches

Description: Self-assembly is the autonomous organization of simple building blocks into complex structures with desirable properties. While self-assembly can take place at all length scales (from atoms and molecules up to macroscopic objects) colloidal particles, which lie in the meso-scale, have long been recognised as an important class of materials in this field. One of the most attractive features of colloids is their tendency to organise into crystalline structures. By tuning the surface chemistry of the individual particles, a range of ordered structures can be fabricated. One way to modify the surface chemistry of colloidal particles is to decorate their surface with localised patches enabling particles to stick to each other.

However the promise of many of these technologies remain out-of-reach due to the limitations of current self-assembly techniques. One problem is that in order to obtain a target equilibrium structure (i.e. a particular cluster geometry) by a process of self-assembly, the nature of the interactions between the constituent particles has to be determined first. In recent years inverse statistical-mechanical methods have been devised that find optimised interactions that lead to the desired final arrangement of particles. However, solving such inverse problems remain difficult and computationally expensive. In addition, the manufacture of particles with exactly the right arrangement of surface patches (dictated by inverse modelling) remains extremely challenging.

This project will investigate a radically new machine-learning based approach which circumvents these difficulties. You will conduct simulations of a small number of colloidal particles in solution which are trapped within a box of a given volume, their small size renders them subject to Brownian motion (meaning they jiggle and are knocked about randomly). Provided the strength of interaction between the sticky patches is not too great (compared to the jiggling due to Brownian noise) then the particles will explore various ensemble configurations (forming and dissolving various temporary structures at a rate that depends on the ambient temperature). Suppose that each time the cluster is observed to form a “desirable” arrangement - i.e. a structure corresponding to a target structure or close to a target structure – the cluster is “rewarded”. Could such a system of reward and punishment be used to “evolve” target structures which are applications in plasmonics, drug delivery, and sensor technologies?

The idea of training a cluster of colloidal particles by gradually reprogramming the interactions between them has never been explored. Not only does this lead to a novel method for the targeted assembly of nanostructures, it also opens a window onto a new scientific paradigm that combines reinforcement learning, thermodynamics and self-assembly.

Project title: Machine learning approaches to the engineering of nanoparticle assemblies and next generation optical materials.

1st supervisor: Dr Adil Mughal
2nd supervisor: Dr Chris Finlayson

Department/Institution: Department of Mathematics, Aberystwyth University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: This project will develop machine learning based simulations offering theoretical insights into the bulk-scale assembly of nanoparticles into large-scale periodic structures with targeted optical properties.

Our recent research into the shear-assembly of nanoparticles in visco-elastic media (Q. Zhao et al, Nat. Comms. 7, p11661, 2016) facilitates new large-scale photonics platforms and applications, and fundamental studies of the physics of ordered materials, photonic crystals, and structural colour.

The present aim is to provide a full theoretical understanding of shear-induced self-assembly based on numerical simulations of packing geometries and inter-particle interactions. Working within the context of experimental advances, these studies will develop a deep-learning framework that predicts how key optical properties (i.e. the "band-gap" structure) of the assembly may be engineered.

The starting point will be to develop an understanding of shear-induced ordering, using molecular dynamics (MD) simulations in assemblies of spherical particles under oscillatory shear. From this, we will develop an understanding, which includes soft (i.e. viscoelastic) interactions of particles and anisotropic particle (e.g. ellipsoids, rods, nano-colloidal cubes) packing problems. Electromagnetic models of resultant structures using finite-difference-time-domain (FDTD) methods will compute the light scattering interactions and thus address questions in optical band-structure engineering.

Finally, a transformative "machine learning" approach is developed, to predict the shapes/sizes of particle that will give rise to the desired optical properties, giving a "dial-in" direction to experimental manifestations of soft matter photonics.

Project title: Learning with Quantum Hidden Markov Models

1st supervisor: Dr Jukka Kiukas
2nd supervisor: Dr Rolf Gohm

Department/Institution: Department of Mathematics, Aberystwyth University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: The general aim is to develop novel mathematical theory for quantum Hidden Markov Models (HMM), building on local expertise on quantum Markov processes [1-3]. While the classical HMM is one of the popular "graphical models" in machine learning [4], its quantum analogue has been studied considerably less. The results of the project will contribute to "quantum machine learning" [5], in the precise sense of identification, monitoring and control of quantum systems. This will, in turn, have impact on emerging data science based on quantum information processing.

Quantum state identification, or tomography, is crucial for applications of quantum information theory. By definition, the state can be identified by measuring an "informationally complete" set of physical quantities [7,8], and this becomes a learning task when one uses measurements with sequential structure, implemented by coupling the "hidden" object system to a sequence of identical copies of an ancillary probe and measuring there — this is a quantum version of a HMM. Another learning task with this model is the inference of the parameters of the model [2,5].

A number of very interesting mathematical problems emerge from this. The Markovian character of the HMM relates them to quantum Markov processes [1-3,6]. Depending on the coupling and measurements one obtains an increasing filtration of operator systems [9] determining which states can be learned in principle. Recent work [1] indicates conditions on the coupling under which one can learn the state if all measurements were accessible, but very little work is done on the practically relevant problem of finding smaller sets of implementable measurements that would suffice [10]. In particular, matrix product state tomography [11] and analogous techniques for classical HMM [12], have not yet been systematically utilised for our goal.

The basic methodology involves probability theory and matrix algebra. Depending on the preferences of the student, this can be extended to abstract operator algebraic methods. Numerical investigations involving simulated learning data can also be included.

The student will also be expected to attend international conferences to network with the relevant research community.

References:

[1] R. Gohm, F. Haag, B. Kümmerer, Commun. Math. Phys. 352, 59–94 (2017).
[2] M. Guta, J. Kiukas, Commun. Math. Phys. (2015) 335: 1397.
[3] J. Gough, M.R. James, Commun. Math. Phys. 287, 1109 (2009).
[4] S. Marshland, Machine Learning: An Algorithmic Perspective, 2nd Ed, CRC Press, Boca Raton, 2015.
[5] A. Monras, A. Winter, J. Math. Phys. 57, 015219 (2016)

Project title: Cancer Localisation and Staging using Breast Ultrasound Imaging

1st supervisor: Professor Reyer Zwiggelaar
2nd supervisor: TBD

Department/Institution: Department of Computer Science, Aberystwyth University

Research theme: T2 - biological, health and clinical sciences T3 - novel mathematical, physical and computer science approaches

Project description: We are interested in localization and staging/classification in computer aided diagnosis, especially with regard to breast ultrasound (which is the main imaging modality for the early detection and staging of breast cancer).

We will bring together and further develop state of the art deep learning methods, but at the same time will integrate more traditional morphological information. In addition, we will incorporate patient specific information.

The work will be based on existing, publicly available, datasets. In recent work, we have brought together a bench-mark dataset, which will form the foundation for the evaluation of the research. The developed approaches to the use of multi-modal data are expected to be transferable to a wider range of application areas (e.g. we also have a lung cancer dataset, which could be used in the evaluation).

The work will be in close collaboration with clinical experts, who tend to be actively involved in discussions and evaluation.

Cardiff University

Project title: Using AI to characterise the turbulent interstellar medium

1st Supervisor: Paul C Clark
2nd Supervisor: Ant P Whitworth

Department/Institution: Physics and Astronomy, Cardiff University

Research theme: T1: data from large science facilities and T3 - novel mathematical, physical and computer science approaches

Project description: Supersonic turbulent motions are thought to play a major role in star formation process. Theories have suggested that ’turbulence’ is responsible for shaping the interstellar medium (ISM), regulating star formation, and even setting the masses of stars as they form in young clusters. However, the properties of the turbulent motions in the ISM are difficult to determine: we can only use the doppler shifting of molecular emission lines to probe their velocities — giving us 1D, not 3D information — and these lines are often optically thick, such that their shapes are difficult to interpret.

Numerical simulations of the turbulent ISM can help here. By using simulations that track the formation and destruction of CO and other species, we can then make synthetic observations of the simulations and use these to study the properties of the turbulence. Since we know the ‘true’ underlying properties of the turbulence in our 3D simulations, we can then test how the standard techniques used on observational data perform on synthetic observations. We can also make use of a machine learning approach to develop new techniques for studying the properties of ISM turbulence in our simulations and observational datasets.

In this PhD, the student will perform magneto-hydrodynamics simulations of the turbulent ISM, using our world-leading ISM chemical/thermodynamical model (Clark et al. 2019). They will then perform radiative transfer postprocessing on the ISM simulations to create "mock" observations of commonly probed molecular lines (e.g. CO, HCO+, HCN, N2H+), which will then form the basis of an extensive machine learning study. The goal will be to see whether modern AI techniques can distinguish between different environmental conditions in the turbulent velocity fields. The student will then apply this technique to real molecular line data from nearby star-forming regions, to provide a new perspective on their turbulent evolution.

This PhD will introduce you to a wide range of ISM physics — such as ISM chemistry, molecular line radiative transfer, (magneto-)fluid dynamics — and also advanced statistics and machine learning techniques.

Project title: A Cold and Dusty Universe: Understanding the cosmic dust and cold gas in nearby galaxies.

1st Supervisor: Matthew Smith
2nd Supervisor: Steve Eales

Department/Institution: School of Physics and Astronomy, Cardiff University

Research theme: T1: data from large science facilities and T3 - novel mathematical, physical and computer science approaches

Project description: Over half the light ever emitted by stars in the Universe having been absorbed by cosmic dust, and the situation is worse (>90%) when looking at regions where star-formation is occurring. Luckily, the dust re-emits the energy absorbed in the far-infrared/sub-millimetre, and so by observing in these wavelengths we can understand these cold dense regions where stars are born.

Our group in Cardiff is leading several international teams to obtain some of the first ground-based sub-millimetre maps of local group and nearby galaxies. Ground based observations are crucial due to the much better resolution and long wavelength coverage. The student will become an active member of several international teams, getting the opportunity to work with researches across the globe. These teams include the HASHTAG, DOWSING teams using the James Clark Maxwell Telescope (JCMT) in Hawaii, the IMEGIN team using IRAM in Spain, and the MUSCAT team using the Large Millimeter Telescope (LMT) in Mexico.

The project will investigate the interplay between the cosmic dust and the other components of the interstellar medium (e.g., atomic gas, molecular gas, and metallicity). For example, we know dust provides a way to measure the 'dark gas' in galaxies and is a way to measure the physical conditions in the interstellar medium. However, very little quantitative analysis has been done extra-galactically due to previous limits on resolution. We also know relatively little about the dust itself, and our recent work discovered that the dust's properties vary significantly across a galaxy. In this project the student will help develop new analysis tools to maximise the information from our observations (for example high-resolution SED-fitters, and hierarchical Bayesian fitting), and applying these techniques to our new high-resolution datasets.

We will then study the interstellar medium and star formation on the scale of individual giant molecular clouds. This includes investigating what is causing changes in the cosmic dust, the amount of dark gas, and what is heating the dust. We will then look at what regulates the star-formation process in galaxies. Whether it's dominated by local properties (e.g., local density or radiation field), or larger-scale properties (e.g, galaxy morphology, disk dynamics). How global galaxy relations, like the correlation between surface-density of star-formation and gas (Schmidt-Kennicutt law), are built from the small scale giant molecular clouds will be investigated.

During the PhD you will learn key skills in data analysis, machine learning, big data analysis techniques, as well as presentation skills. It is expected that you will have the opportunity to learn hands-on by observing at an international telescope, and participate at international conferences. You will also have access to range of training events both within the department and organised by the University.

References:

The HASHTAG Project: The First Submillimeter Images of the Andromeda Galaxy from the Ground, https://arxiv.org/abs/2110.00011

The Dust in M31, Whitworth, A.P, et al. 2019, https://arxiv.org/abs/1908.03458
The star formation law at GMC scales in M33, the Triangulum Galaxy, Williams, T et al. 2018, https://arxiv.org/abs/1806.01293

The Herschel Exploitation of Local Galaxy Andromeda (HELGA). II. Dust and Gas in Andromeda, Smith, M.W.L, et al. 2012, https://arxiv.org/abs/1204.0785

Project title: Solving the “structure – optical response” relation by combining the resonant-state expansion with machine learning

1st Supervisor: Egor Muljarov
2nd Supervisor: Wolfgang Langbein
3rd Supervisor: Francesco Masia

Department/Institution: Cardiff University, School of Physics and Astronomy, School of Biosciences

Research theme: T2 - biological, health and clinical sciences and T3 - novel mathematical, physical and computer science approaches

Project description: Electromagnetic resonances are widely exploited in modern technologies, and significant advance in their simulation and underlying understanding has an extensive field of applications with a potential impact on society. For example, nanoplasmonics, dealing with optical phenomena in the nanoscale vicinity of metallic surfaces, has yielded practical applications in physics, engineering, biomedicine, and even for environmental monitoring and homeland security. Plasmonic resonances can probe nanoscale regions in the vicinity of a metallic nanoparticle (NP), enabling nano-spectroscopy for reading the chemical structure of the object with nanometer resolution, biological labelling, imaging and biomedical sensing.

A host of methods in computational electrodynamics are presently employed for calculating resonances in optical systems and modelling light scattering, and their computational codes have been optimised over the past decades to increase their accuracy and speed. However, the forward design of complex resonators, or solving the inverse problem using these methods is still severely restricted by available computational power. The resonant-state expansion (RSE), developed [1] by Egor Muljarov and Wolfgang Langbein, provides an alternative method for accurate and efficient calculation of resonances in arbitrary open optical systems, which can be orders of magnitude more computationally efficient [2]. Its continuous development over the last decade, with over 25 original works published during this period, has opened a new field attracting other research groups in the UK and worldwide.

The RSE development has improved the accuracy and speed in solving the forward problem, from resonator structure to optical response. However, solving the inverse problem, from optical response to structure, is still beyond reach, as it traditionally requires an iterative minimisation using many forward solutions. This project wants to address this issue, by combining the high speed of the RSE to provide a forward data set sampling the structural parameter space sufficiently to train machine learning (ML)/artificial intelligence models for solving the inverse problem of determining the structure from the optical properties.

Aligned with ongoing experimental activity and development of the Nanosizer technology [3] in the supervisors groups, the project will use as examples the extinction and absorption cross-section spectra of gold and silver NPs of various structural faceting classes, such as decahedra, tetrahedra, bipyramids, etc [4].

The student will use and further develop the RSE-based scattering-matrix method [5] to calculate the forward problem, determining the scattering cross-sections of NPs of different sizes and shapes, comparing results with COMSOL calculations and measured spectra. As an introduction to ML, a model will be trained for the forward problem, and the accuracy will be optimised by varying the type and structure of the ML algorithm. To solve the inverse problem, the student will use the RSE results of the forward problem as training set for ML of the inverse problem. As an outcome of this research, a generic platform based on the RSE and ML will be developed and made publicly available, which can be used by academia and industry.

References:

[1] E. A. Muljarov, W. Langbein and R. Zimmermann, "Brillouin-Wigner perturbation theory in open electromagnetic systems", Europhys. Lett. 92, 50010 (2010);

UK patent GB2544601 (2020).

[2] M. B. Doost, W. Langbein, and E. A. Muljarov, “Resonant-state expansion applied to three-dimensional open optical systems” Phys. Rev. A 90, 013834 (2014).

[3] L. M. Payne, W. Albrecht, W. Langbein and P. Borri, “The optical nanosizer – quantitative size and shape analysis of individual nanoparticles by high-throughput widefield extinction microscopy”, Nanoscale 12, 16215 (2020);

US patent US10996159B2 (2021).

[4] Y. Wang, Z. Sztranyovszky, A. Zilli, W. Albrecht, S. Bals, P. Borri and W. Langbein, “Quantitatively linking morphology and optical response of individual silver nanohedra”, Nanoscale 14, 11028 (2022).

[5] S. V. Lobanov, W. Langbein, and E. A. Muljarov, “Resonant-state expansion of three-dimensional open optical systems: Light scattering”, Phys. Rev. A 98, 033820 (2018).

Project title: Simulations of massive galaxies and their circumgalactic medium

1st Supervisor: Freeke van de Voort
2nd Supervisor: Mattia Negrello

Department/Institution: School of Physics and Astronomy, Cardiff University

Research theme: T1: data from large science facilities and T3 - novel mathematical, physical and computer science approaches

Project description: Galaxies live in large, dark matter-dominated, quasi-spherical structures called haloes. The gaseous component of these haloes is also known as the circumgalactic medium (CGM) and comprises the reservoir of material from which the embedded galaxies can grow. Galaxies do not only passively accrete gas from their environment, they also drive large-scale outflows into the CGM. These outflows enrich the CGM, magnetize it, and drive shocks and turbulence. Therefore, if we wish to understand the evolution of galaxies, we need to study the CGM. This PhD project will focus on the environments of galaxies more massive than the Milky Way. Such massive haloes are expected to host volume-filling hot gas with temperatures close to the virial temperature, which increases with halo mass. Radiative cooling allows some fraction of the gas, especially dense or metal-rich gas, to cool down to much lower temperatures. This creates a multiphase medium where the hot, diffuse gas and the cooler, denser gas are in approximate pressure equilibrium. The massive galaxies themselves behave very differently to lower mass galaxies, because they no longer rapidly grow and have mostly switched off star formation, a process known as quenching. Observations of the CGM have shown a large reservoir of cool gas even around galaxies with no cold interstellar medium or ongoing star formation. This came as a surprise: why does this cool gas not accrete onto the central galaxy and fuel star formation? These observations have thus led to the suggestion that our theoretical understanding is incomplete.

This project will use cosmological, magnetohydrodynamical simulations, which start when the universe was very young and follow not just the formation of galaxies but also the large-scale structure of the universe. Traditionally, such simulations focus most of their computational effort on creating the best galaxies and incorrectly assume that the treatment of the CGM is sufficiently accurate. New numerical methods are being developed in order to increase the resolution of the gas around galaxies and study the massive galaxy haloes in more detail than has ever been possible before. The PhD student will take charge of developing and running galaxy simulations with enhanced CGM resolution and investigate the importance of various physical processes, such as magnetic fields and thermal conduction. A single simulation is expected to generate 15 TB of data and will thus require novel analysis techniques. After developing a successful analysis package, the student will have the opportunity to exploit the simulation data to study gas accretion, large-scale outflows, chemical enrichment, and other properties of the CGM. By comparing their simulations to existing observations and making predictions for future observations, the PhD student will be able to reveal the crucial role the CGM plays in the formation of massive galaxies and to put galaxy formation models to the test.

References:
[1] Cosmological simulations of the circumgalactic medium with 1 kpc resolution: enhanced HI column densities
[2] Freeke van de Voort, Volker Springel, Nir Mandelker, Frank C. van den Bosch, Rüdiger Pakmor

[3] Monthly Notices of the Royal Astronomical Society: Letters, Volume 482, Issue 1, p.L85-L89

Project title: Design of open science standards for specialised brain imaging data

1st Supervisor: Leandro Beltrachini
2nd Supervisor: Derek K Jones
3rd Supervisor: Kevin Murphy

Department/Institution: Schools of Physics and Astronomy and Psychology; Cardiff University Brain Research Imaging Centre (CUBRIC)

Research theme: T1: data from large science facilities and T2 - biological, health and clinical sciences

Project description: Characterising structures at the cellular scale is key to understanding tissue properties in health and disease. Performing such a characterisation in vivo and non-invasively is an extremely challenging enterprise. Central to this mission is magnetic resonance imaging (MRI), which has shown its value for depicting tissue microstructure in unprecedented detail. The versatility of MRI has attracted a variety of scientists, creating a diverse landscape of acquisition and processing methods. The large heterogeneity in notations and acquisition parameters employed in the literature makes the adoption of data sharing practices a priority with the aim of facilitating scientific reproducibility and maximising efforts and investments in the area. Despite the attempts to solve this issue, scholars in the field of microstructural MRI have not embraced such open practices, mainly due to the lack of a robust and general methodology that can be adapted to heterogeneous data description as needed in MRI.

In this project, the student will develop and utilise data archiving standards for specialised MRI data acquired to depict tissue microstructure. More specifically, the candidate will generate a comprehensive and generalised data archiving framework for openly storing and sharing microstructural MRI information under the Brain Imaging Data Structure (BIDS) umbrella. This framework will allow to record all the data and metadata needed to replicate the MRI experiments performed, simplifying the data sharing and processing stages. This includes the description of all the time varying fields tuned in the scanner and relevant for understanding the recordings. The standard will work with multimodal contrasts used in the research area, with special emphasis in diffusion MRI and spectroscopy. In a later stage, the resulting standard will be used to automatise data processing algorithms depending on available acquisition parameters, paving the way for the application of Machine Learning methods.

The PhD project will take place in the Cardiff University Brain Research Imaging Centre (CUBRIC), a pioneer in brain imaging research. CUBRIC houses >200 researchers across Schools and Colleges, making it a vibrant multidisciplinary research community. Moreover, the centre hosts state-of-the-art MRI equipment that the student will benefit from.

The successful candidate will have a unique training opportunity, involving access to a heterogeneous and continuously growing MRI data archive, tuition for specialised MRI operation and sequence design, presenting research work in international conferences, establishing links with industrial partners (e.g., SIEMENS), and being part of the larger collaboration in the area.

References:
J Gholam, F Szczepankiewicz, CMW Tax, L Mueller, E Kopanoglu, M Nilsson, S Aja-Fernandez, M Griffin, DK Jones, L Beltrachini, “aDWI-BIDS: an extension to the brain imaging data structure for advanced diffusion weighted imaging”, arXiv:2103.14485v2, 2021.

Project title: Characterisation of tissue microstructure from non-invasive MRI using Machine Learning

1st Supervisor: Leandro Beltrachini
2nd Supervisor: Derek K Jones
3rd Supervisor: Kevin Murphy

Department/Institution: Schools of Physics and Astronomy and Psychology; Cardiff University Brain Research Imaging Centre (CUBRIC)

Research theme: T2 - biological, health and clinical sciences and T3 - novel mathematical, physical and computer science approaches

Project description: The characterisation of biological tissue microstructure in vivo and non-invasively is of outmost interest in science. If successful, it could reveal unique insights into biological processes, including aging and cancer. In this regard, MRI plays a crucial role due to its superb flexibility to depict soft tissues. Researchers in the field have utilised this technology mostly based on assumptions about the shape of cellular components (spheroidal/cylindrical), which have been shown to introduce unwanted errors. The reason for making such approximations is the intractability that arbitrary tissues may impose on the mathematical models employed.

To tackle the problem, the supervisory team has been exploring the introduction of a potentially disrupting methodology borrowed from materials science. It consists of measuring statistical descriptors (SDs) of tissue microstructure using the MRI scanner, from which histology-like representations may be reconstructed. These SDs have the advantage of describing the statistical nature of tissue components without relying on any assumption on shapes and arrangements, making the technique potentially useful to depict tissue microarchitecture as never before. One of the major drawbacks of this technique resides in the instability and computational demands of the reconstruction step, which can last for days even in modern computers.

In this project, the student will provide a solution to the problem by introducing machine learning (ML) approaches in the process: first, to generate fast reconstructions of tissue microstructure based on MRI-based SDs; and second to perform quick simulations of MRI signals for any given microstructure, as those generated from SDs. Synthetic datasets representing biological tissues will be generated and used to train and test the algorithms, with special emphasis on prostate cancer. It is expected that the adoption of ML will bring the extra boost to successfully bring this technology to the medical imaging domain.

The PhD project will take place in the Cardiff University Brain Research Imaging Centre (CUBRIC), a pioneer in brain imaging research. CUBRIC houses >200 researchers across Schools and Colleges, making it a vibrant multidisciplinary research community. Moreover, the centre hosts state-of-the-art neuroimaging equipment that the student will benefit from, including the Connectom scanner with ultra-strong gradients.

The successful candidate will have a unique training opportunity, involving the possibility to attend courses in ML and MRI, tuition for specialised MRI operation, present research work in international conferences, establish links with industrial partners (e.g., SIEMENS), and be part of the larger collaboration in the area.

Project title: Exploring the Gravitational-Wave Sky with Machine Learning

1st Supervisor: Prof Patrick Sutton
2nd Supervisor: Dr Vivien Raymond
3rd Supervisor: Prof Stephen Fairhurst

Department/Institution: School of Physics and Astronomy, Cardiff University

Research theme: T1: from large science facilities data

Project description: Gravitational waves (GWs) are produced by some of the most violent events in the Universe, such as the mergers of black holes and the explosive deaths of massive stars. Rapid detection of such signals can trigger follow-up observations by other facilities including ground-based telescopes, satellites, and neutrino observatories, greatly increasing the scientific payoff of such discoveries.

For example, combined observations of the binary neutron star merger GW170817 revealed the origin of heavy elements in the Universe and provided a new way to measure the Hubble constant.

Detecting such GW events before the electromagnetic counterpart fades requires analysis of the gravitational-wave data on timescales of minutes or less. Recently, machine learning techniques based on Convolutional Neural Networks (CNNs) have been demonstrated to have the potential for sub-second-latency analysis of data from GW detector networks. The aim of this project is to develop and implement a CNN with the ability to detect generic transient signals, such as those expected from newly formed or perturbed black holes and neutron stars. The analysis will be run in real time during upcoming observing runs by the LIGO-Virgo-KAGRA detector network, to detect, characterise, and determine the location on the sky of GW signals. We will issue public alerts about detected events, with special emphasis on signals associated with gamma-ray bursts, core-collapse supernovae, and other relativistic astrophysical phenomena.

This project involves developing expertise in programming, signal processing, high-throughput computing, and machine learning. You will collaborate with other GW observers, GW theorists, and astronomers in the Cardiff University Gravity Exploration Institute and in the LIGO Scientific Collaboration. The project may optionally involve a long-term (up to four months) secondment to one of the LIGO observatories in the US.

References: https://arxiv.org/abs/2009.14611

Project title: Vector Symbolic Architecture for Edge AI

1st Supervisor: Alun Preece
2nd Supervisor: Yuhua Li

Department/Institution: School of Computer Science and Informatics, Cardiff University

Research theme: T3 - novel mathematical, physical and computer science approaches

Project description: The proliferation of the Internet of Things (IoT) has and will continue to result in a constant generation of massive volumes of data from distributed electronic devices. Cognitive processing (e.g., reasoning, problem solving) is about making sense of all the information gathered in order to operate effectively. The current techniques that make sense from data largely rely on machine learning (ML). However, conventional ML, and deep learning in particular, demand great computational power and consumer high energy so they are suitably implemented in the cloud and data centres. Due to constraints of security and privacy, response latency and communication cost, the limited energy capacity and computing resources of edge devices make the current computationally demanding machine learning and cognitive processing algorithms impractical for mass deployment. Many future applications will therefore require IoT devices that perform cognitive processing at the network edge. Therefore, it is crucial to develop computational, energy and communications efficient algorithms that enable data handling and intelligence embedding on edge devices such as the next generation of neuromorphic processors.

The proposed research harnesses advancements from an emerging computing framework termed Vector Symbolic Architectures (VSA) or Hyperdimensional Computing, to develop lightweight algorithms for efficiently performing cognitive processing at the network edge. VSA is a form of brain-inspired computing for representing and manipulating data as hypervectors in a high-dimensional vector space. Unlike classical computing, VSA operates on hypervectors using binding (e.g., XOR, permutation) and bundling operations (e.g., superposition) and obtains the result using vector comparison (e.g., 'clean-up' memory). Its distributed representation and manipulation of information inherently makes the computing robust to noise, is scalable, energy efficient and requires less time and data for training and inference. The proposed research aims to investigate a new paradigm for machine intelligence based on the inherent features of VSA, by developing algorithms addressing challenges at key stages of a VSA pipeline including semantic vector generation, vector binding and bundling, synchronisation and communication across multiple devices and distributed 'clean-up' memory management.

In harnessing the latest advances of VSA, the proposed research aims to explore the potentials of VSA to establish a systematic cognitive processing framework based on VSA that maps naturally onto the Observe, Orient, Decide & Act (OODA) cycle. It will develop a suite of novel learning methods that can naturally address the weakness of conventional machine learning (including deep learning and transformer networks) to deal with one/few shot learning, learning with very few examples, continual learning and lifelong learning. The developed new learning methods will exploit the distribution, composition and superposition properties of VSA to naturally accommodate data related issues including data contamination, missing values, data imbalance, data rarity and data drift. Recent work has shown how the VSA Sparse Distributed Memory (SDM) model closely approximates the attention mechanism used in transformer networks. VSA provides insights into why attention is so effective and suggests approaches to further improve the performance of these networks by combining them with the analogical and abductive reasoning capabilities of the vector representation.

References:

Bent, Graham, Chris Simpkin, Yuhua Li, and Alun Preece. "Hyperdimensional computing using time-to-spike neuromorphic circuits." In 2022 International Joint Conference on Neural Networks (IJCNN), pp. 1-8. IEEE, 2022.

Simpkin, Chris, Ian Taylor, Daniel Harborne, Graham Bent, Alun Preece, and Raghu K. Ganti. "Efficient orchestration of Node-RED IoT workflows using a vector symbolic architecture." Future Generation Computer Systems 111 (2020): 117-131.

Simpkin, Chris, Ian Taylor, Graham A. Bent, Geeth de Mel, Swati Rallapalli, Liang Ma, and Mudhakar Srivatsa. "Constructing distributed time-critical applications using cognitive enabled services." Future Generation Computer Systems 100 (2019): 70-85.

Swansea University

Project title: AI based approaches multi-dimensional functional genomics in cancer patients

1st Supervisor: Prof Steve Conlan Medical School
2nd Supervisor: Dr Lewis Francis Medical School

Department/Institution: Swansea University Medical School

Research theme: T2: biological, health and clinical sciences

Project description: This project will advance the fundamental understanding of the functional genomics status and dynamic responses in cancer environments, with project opportunities including single cell and spatial functional genomes including transcriptome and epigenome analysis.

Project examples include

Exploring the spatial epigenomics of a 3D cellular models to understand the impact of pharmacological interventions on histone methylation (e.g. H3K4Me3, H3K9me2, H3K27me3). Epigenome profiling in response to interventions will be assessed at the cellular level using the Microscopic Imaging of Epigenetic Landscapes technique involving high content multi-parameter data acquisition and analysis to reconstruct high-resolution spatial epigenome models.

Single cell and spatial analysis of cancer and pre-cancerous functional genomes obtained using 10X Genomics platforms to generate data from cancer and immune cell complements of ascites (ATAC and RNA seq) or spatial analysis of tumour biopsies (10 X Visium, RNA seq) from cancer patients.

Cell painting and/or data mining approaches developed in Swansea will be applied to overcome challenges for high content analysis including feature extraction and data analysis, and interpretation requiring the use of AI technologies (using Swansea's new ATOS supercomputing capability). The successful applicant will develop and implement AI based strategies for the high-content data generated from cellular models of tumour microenvironments/cancer patient samples using advanced and computationally expensive algorithms.

The successful applicant will join the Reproductive Biology and Gynaecology Oncology research group in Swansea's Medical School

The successful applicant will be involved in data acquisition and analysis and should have a degree in molecular biology or computer science or similar.

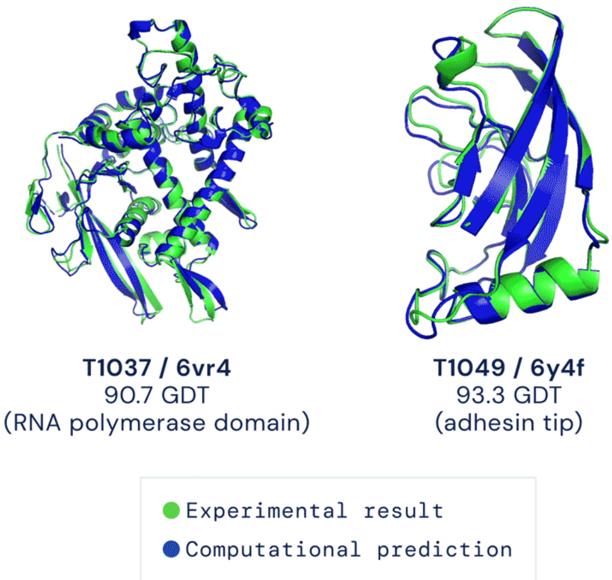
Project title: Protein Structure Prediction via Deep Learning

1st Supervisor: Lin Yuanbo Wu (Senior Lecturer in AI)
2nd Supervisor: Xianghua Xie (Professor in Computer Science)
3rd Supervisor: Johathan Mullins (Associate Professor in Biomedical Science)

Department/Institution: Department of Computer Science; Institute of Life Science

Research theme: T2 - biological, health and clinical sciences

Project description: Proteins are the basic composition units of life activities and have a vital impact on life being. The structure of proteins is the key to study life cycles, major diseases and drug discovery. However, since the affirmation of proteins, humans are limited in acquiring profound understanding in proteins, which is manifested in the insufficient ability to predict the structure and function of proteins, let alone modifying proteins to meet the needs in a variety of practices. Until 2021, Google DeepMind developed AlphaFold2 [1,2] based on deep learning to improve the accuracy of protein structure prediction and they achieved the prediction accuracy as high as 90%. The emergence of AlphaFold2 partially solves the puzzle of protein structures and shows the great promise from deep learning towards this field. However, AlphaFold2 still has certain limitations. For instance, AlphaFold2 cannot accurately predict the structure of multi-domain proteins, and it is powerless for proteins with dynamic structures. However, the structural prediction of these proteins is highly important to the domains of biology and medicine. In this end, this is an urgent problem to be solved. Therefore, the study of protein structure prediction remains a valuable topic that requires sustained investment and should be paid more attention in our near future.



Since AlphaFold2 is not completely open source, this research group has completely reproduced and verified the inference and training code of AlphaFold2 [2]. They have completed the downloading/sorting of the training dataset and made the understanding of protein unfolding more profoundly than before. Based on such approachable principle, it enables researchers to make certain improvements based on AlphaFold2. After adopting the open-source weights provided by DeepMind, the current reproduced algorithm can more accurately predict the structure of proteins in many types. In other words, being able to train AlphaFold2 completely is the basis for the subsequent improvement work, and the training part of the code has also been reproduced and verified. Thus, AlphaFold2 can be preliminarily trained.

Short-term goals:

By improving the key technique of AlphaFold2, we aim to improve the accuracy and strive for winning the first place in the cameo protein structure prediction competition. Along with this process, we will have a deeper understanding of protein structure prediction and accumulate useful experience for follow-up work.

Long-term goals:

By exploring energy equations and differential equations that can better reflect the realistic chemical components as many medical experts expect, we aim to address a few challenges raised by complex protein structural prediction (such as multi-domain proteins, dynamic structural proteins). We will promote the collaboration with researchers in biology and medicine to provide technical support in their needs and accelerate potential research directions in relevant fields.

References:

[1] Highly accurate protein structure prediction with AlphaFold. Nature 596, pages 583–589, 2021.
[2] https://github.com/deepmind/alphafold

Project title: Development of a plasma lens for Laser hybrid Accelerator for Radiobiological Applications (LhARA) with an advanced computational approach

1st Supervisor: Prof Stefan Eriksson
2nd Supervisor: Dr Richard Hugtenburg
3rd Supervisor: Dr Christopher Baker

Department/Institution: Physics / Medical Physics at Swansea University

Research theme: T2 - biological, health and clinical sciences and T3 - novel mathematical, physical and computer science approaches

Project description: In 2020 cancer was responsible for over 15% of worldwide deaths, and 50% of treatments require radiotherapy. Typically, radiotherapy is performed with MeV-energy X-rays, but particle (proton & ion) beam therapy offer significant advantages: with less damage to healthy tissue and more energy deposited within cancerous tissues. The nascent LhARA (Laser hybrid Accelerator for Radiological Applications) collaboration [1] proposes novel ion-beam technology to reduce facility costs, improve patient access, and improve treatment performance. A novel laser source will generate a high intensity ion beam with a wide range of ion species, but high divergence. The programme will develop new ion-beam optics based on extreme (large volume, high density) non-neutral, single component, plasmas which will shape the ion beam before extraction and acceleration to radiobiological applications which will use new technologies for diagnosing the beam delivery. The proposed technology can be viewed as a radical departure from convention and computational methods will have an essential role in guiding the development of LhARA.

In this project, we will simulate the plasma in the ion-beam optics and compare with both preliminary experimental results [2] and new data from experiments at Swansea University. We will conduct further numerical work to guide the apparatus development in LhARA and more fully understand the observations once the experimental campaign begins. Owing to the extreme nature of the plasma, and the high computational expense of suitable Particle-In-Cell (PIC) simulations [3], an efficient way to explore the relevant parameter space is needed. Here, we propose leveraging machine learning and AI techniques. The techniques developed for analysing the PIC simulations and optimising the apparatus design parameters will be employed to analyse experimental observations, once available. The experimental apparatus will then be stepwise refined from well-understood plasma conditions towards the final operating regime.

LhARA takes a patient centred approach and has stakeholders in clinical research from the outset. At Swansea University our LhARA group has involvement from Medical Physics, and we plan to embed interaction with patient groups and outreach in this project. The student will become a part of a multidisciplinary team from the Faculties of Science and Engineering and Medicine, Health and Life Science at Swansea University and the Department of Physics, University of Manchester.

References:

[1] Aymar et al., “LhARA: The Laser-hybrid Accelerator for Radiobiological Applications” Front. Phys. 8 567738 (2020).

[2] Nonnenmacher et al., “Anomalous Beam transport through Gabor (Plasma) Lens Prototype” App. Sci. 11 4357 (2021). [3] https://txcorp.com/vsim/