Table of Contents

Main page

Aberystwyth

Project title: Principled Application of Evolutionary Algorithms

1st supervisor: Christine Zarges
2nd supervisor: Thomas Jansen
Research theme: T3 – novel mathematical, physical and computer science approaches
Project description: Evolutionary algorithms are general and robust problem solvers that are inspired by the concept of natural evolution. Over the last decades, they have successfully been applied to a wide range of optimisation and learning tasks in real-world applications. Recently, some researchers [1,2] argue that evolutionary computation now has the potential to become more powerful than deep learning: While deep learning focuses on models of existing knowledge, evolutionary computation has the additional ability to discover new knowledge by creating novel and sometimes even surprising solutions through massive exploration of the search space.

While evolutionary computation methods are often easy to implement and apply, to achieve good performance, it is usually necessary to adjust them to the problem at hand. The main goal of this project is to exploit recent theoretical advances that shed light on the fundamental working principles of evolutionary algorithms [3] in real-world applications. It will build upon recent momentum and progress in both, theory and applications of evolutionary algorithms and related randomised search heuristics and further contribute to bridging the gap between these two branches of evolutionary computation research [4]. Starting point for the investigations will be modern benchmarking frameworks, e.g., [5,6], and competitions that tackle important societal and industrial challenges, see [7,8] for examples. Possible application areas can be discussed with the supervisors. They include but are not limited to routing, scheduling, and planning problems, bioinformatics, as well as benchmarking and combinatorial optimisation in general.

References:

[1] Neural Networks Research Group at the University of Texas at Austin. Evolution is the new deep learning. http://nn.cs.utexas.edu/pages/evolutionary-ai/newdeeplearning/
[2] Risto Miikkulainen. https://venturebeat.com/2018/05/17/evolutionary-computation-will-drive-the-future-of-creative-ai/
[3] Benjamin Doerr, Frank Neumann (eds). Theory of Evolutionary Computation - Recent Developments in Discrete Optimization. Springer, 2020. http://doi.org/10.1007/978-3-030-29414-4
[4] COST ACTION CA15140: http://imappnio.dcs.aber.ac.uk
[5] IOH Profiler. https://iohprofiler.github.io
[6] Nevergrad. https://facebookresearch.github.io/nevergrad/
[7] ACM GECCO 2021 Competitions. https://gecco-2021.sigevo.org/Competitions
[8] IEEE CEC 2021 Competitions. https://cec2021.mini.pw.edu.pl/en/program/competitions

Project title: Machine learning for NP-hard minimization problems

1st supervisor: Simon Cox (Maths, Aberystwyth)
2nd supervisor: TBD
Research theme: T3 – novel mathematical, physical and computer science approaches
Project description: Given a flat, circular pizza and a group of N friends, how should one cut the pizza with the least amount of cutting so that everyone gets a piece of the same size? This somewhat improbable question highlights a more general class of problems associated with seeking least perimeter or least area partitions of two- or three-dimensional objects respectively. Such problems arise naturally in the study of the geometric structure of foams and emulsions and, like the better-known Travelling Salesman problem, are generally NP-hard.

Re-framing the question as determining the least-perimeter partition of a circle into N regions of equal area, some progress in determining the optimal solutions has been made for small N. There are proofs for N=1, 2, 3, conjectured solutions for N up to 42 [1], and evidence that defects tend to cluster towards the periphery of the partition at N over about 5,000 [2].

There are also many generalisations: partitioning regular polygons, the surface of various three-dimensional shapes such as the sphere and the torus, or three-dimensional space itself. This latter question is known as the Kelvin problem, for which a solution has been sought for almost 150 years.

We therefore have training data and a well-defined optimization function. The goal of the PhD project would be to develop a machine learning algorithm for such problems, with the aim of presenting solutions for large N with a high degree of confidence, which might inspire mathematical proofs, and possibly new solutions to the Kelvin problem.

References:
[1] S.J. Cox and E. Flikkema (2010) The minimal perimeter for N confined deformable bubbles of equal area. Elec. J. Combinatorics 17:R45. See also users.aber.ac.uk/sxc/two_d_clusters.html

[2] S.J. Cox, F. Morgan and F. Graner (2013) Are large perimeter-minimizing two-dimensional clusters of equal-area bubbles hexagonal or circular? Proc. Roy. Soc A 469: 20120392.

Project Title: Collaborative mapping of large scale outdoor environments

1st supervisor: Dr Myra Wilson
2nd supervisor: Dr Fred Labrosse
Research theme: T3 – novel mathematical, physical and computer science approaches
Project description: Large outdoors environments often present variable types of land and terrain, both natural and artificial.
For many applications (e.g. search and rescue, military, and farming), area mapping is essential but difficult to achieve using solely ground-based vehicles due to the variability of the terrain.

This project proposes the use of a swarm of Unmanned Aerial Vehicles (UAV's) (fixed wing or multi-copters) to provide a comprehensive coverage to precisely and accurately map the area in combination with ground-based vehicles where appropriate. Previous work has investigated the use of Evolutionary Algorithms (EAs) to provide reliable communication between UAV's and ground-based vehicles in simulated environments [1]. This project would build on this work in a novel direction on real hardware.

Each UAV will be equipped with a camera and a small computer (eg Raspberry Pi) as well as data communication hardware. The UAV will perform real-time image analysis to decide where to fly in relation to the ground to accurately capture its features. For example, the UAV would fly at a lower altitude above highly textured areas, or to fly along tracks [2]. The data will be relayed to a base station using radio communication, possibly using other UAV's as relays to the base station. An EA will be used to position the drones to maximise the radio coverage of the swarm. The data collected can be used to map the area, and the UAV's and ground-based vehicles can be directed towards areas of interest.

References:
[1] "UAV Flight Coordination for Communication Networks: Genetic Algorithms versus Game Theory", 2021, Giagkos, A., Tucci, E., Wilson, M.S., and Charlesworth, P. In: Soft Computing, p. 1-21, https://doi.org/10.1007/s00500-021-05863-6
[2] "Automatic driving on ill-defined roads: an adaptive, shape-constrained, color-based method" Ososinski, M. & Labrosse, F., 2015, In: Journal of Field Robotics. 32, 4, p. 504-533 30 p.

Project title: Using Artificial Intelligence to understand hand eye co-ordination in humans.
1st supervisor: Dr David Hunter
2nd supervisor: Dr Patricia Shaw
Research theme: T2 – biological, health and clinical sciences T3 – novel mathematical, physical, and computer science approaches
Project description: This project will involve creating Virtual Reality environments to track peoples' hand and eye motions while they interact with objects. Data from these experiments will be used to design and train an artificial neural network that can learn to accurately replicate hand and eye motions. Over the past decades researchers have shown training neural networks that mimic known features of human physiology can teach us much about how human visual processing functions and develops. This project will look at Reaching and Grasping, a highly complex visual task that requires close coordination of eye gaze and motor control. Unlike many other gaze-related visual tasks, reaching and grasping requires top-down intent driven control of visual processing and therefore allows us to study how high-level intent drives low-level feature-driven visual processing. The Higher Education Funding Council for Wales Research Capital Funding has recently invested in a new eXtended Realities laboratory in Aberystwyth that will allow us to accurately track both eye gaze (saccades) and hand motion.

Project title: Utilisation of recorded voice samples for developing a machine learning framework to predict pulmonary function and related diseases

1st supervisor: Dr. Faisal I. Rezwan, Department of Computer Science, Aberystwyth University, Aberystwyth, UK

2nd supervisor: Dr. S. A. Reza Nouraei, Robert White Centre for Airway Voice and Swallowing, Poole Hospital, Poole, UK

Prof. Judith Holloway, Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK

Prof. John Holloway, Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK

Prof. James Batchelor Clinical Informatics Research Unit, University of Southampton, Southampton, UK

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

Project description: It is well known that, when speaking to their asthmatic patients, health care professionals can hear abnormalities in respiratory tract and judge the severity, however, it is difficult to automate this process due to challenge in differentiating speech and breathing automatically in recorded voice data. Very few works have been done based on speech breathing, and traditional statistical methods are insufficient to automate the process.

Previously, we designed a threshold-based mechanism to separate speech and breathing from 323 voice recordings from 26 samples (healthy controls, asthmatics smokers) in controlled environment and developed three machine learning (ML) models: a regression model to predict lung function (percentage predicted forced expiratory volume in 1 second, FEV1%) (mean square error = 10·86), a multi-class classification model to predict severity of lung function abnormality (accuracy = 73%), and a binary classification model to predict lung function abnormality (accuracy = 85%). (doi: https://doi.org/10.1101/2021.05.11.21256997, under review: Frontiers in Digital Health)

However, while implementing the threshold method on independent samples, the efficacy of the method was found inefficient to separate speech and breathing parts and which lead to the challenge of automatically detect speech and breathing from recorded voice samples irrespective of the surrounding environment and device. A more robust mechanism is warranted to generalise this separation method to identify speech and breath part of recorded voice which can handle this as well as accent, language and culture, and further use them to predict respiratory conditions. Hence, ML can play an important role in developing both separation method and prediction models. Therefore, the aim of this project is to develop an automated and stable system, using ML methods, to detect speech and breathing from recorded voice samples and predict the respiratory conditions with high accuracy.

Objectives:

• To explore different ML methods that is capable of separating speech and breathing from recorded voice sample irrespective from controlled or uncontrolled environment. These methods will be validated on the 200 independent samples (recorded using mobile phones in uncontrolled environment) to check the general usability and handling the challenges of background noise.

• To perform feature extraction for developing prediction models to predict lung function and/or severity of lung abnormalities, once a suitable separation method selected.

• To implement the established workflow on recorded voice samples generated by our partners from UK, Australia, Brazil and Sri Lanka to observe the effect of accent, language and culture.

• To implement the developed ML framework for predicting related pulmonary function and diseases (such as: lung function, asthma from Australian and Brazilian samples, COPD from UK samples, and COVID-19 from Sri Lankan samples).

• To develop an online platform and/or a mobile app prototype based on the ML framework and make available for wider community. This will help to collect recorded voice independently

from public with appropriate questionnaires relating to respiratory health and pave the way for further analysis.
Conclusion: The findings from this project will not only address the challenge of handling voice recording data for appropriately separating speech and breathing to extract features to develop prediction models with high accuracy but also have potential in implementing as a smartphone application in the future, offering a convenient and straightforward way to assess respiratory health for individuals.

Bangor

Project title: X-ray simulation and deep learning: Application to Automatic segmentation of defects in CT images corrupted by artefacts

1st supervisor: Dr Franck P. Vidal / School of Computer Science and Electronic Engineering
2nd supervisor: Dr Simon Middleburgh / School of Computer Science and Electronic Engineering
Research theme: T3 – novel mathematical, physical and computer science approaches
Project description: The use of X-ray computed tomography (XCT) in precision engineering is becoming commonplace to assess the accuracy of a manufacturing process. The ISO 10360-11 standard has recently been issued to define metrological characteristics and methods making use of XCT. Two of the main challenges lie in extracting the 3D surfaces from XCT data and detecting manufacturing defects in images that are prone to artefacts such as beam-hardening, phase contrast, scatter radiation and partial volume effect. We have developed gVirtualXRay, a fast programming-library to simulate X-ray images on graphics processing units (GPUs) [1]. Being able to generate many X-ray images is useful in machine learning (ML). We deployed gVirtualXRay on Supercomputing Wales to train an optimisation algorithm reproduce real XCT images that where highly corrupted by beam-hardening and phase contrast [2]. Haiderbhai et al. Used gVirtualXRay to build a large training dataset of simulated images. This dataset is used to train a generative adversarial network (GAN), a ML approach to create synthetic images.
The segmentation of real images by ML requires a large amount of manual input to build training datasets. In this research, the PhD candidate will segment defects (e.g., cavities and cracks) from XCT scans and produce parametrised models to generate them in CAD models. Corresponding XCT scans will be simulated on Supercomputing Wales, with and without defects, with and without scanning artefacts, to build datasets in a controlled environment (the location and type of defects will be known, as well as the imaging artefacts). These datasets will be used to train and evaluate a convolutional neural network (CNN).
This research will benefit a large part of the manufacturing sector, where NDT by XCT or radiography is used. It will help material scientist detect manufacturing defects in XCT images corrupted by imaging artefacts. It will also enable the automation of defect detection in an industrial context.
National and international collaborators:
Dr Llion Evans, Director of the Centre of Excellence in Advanced Data-Driven Engineering Design, Swansea University
Prof Jean-Yves Buffiere, MATEIS Materials Science laboratory, INSA-Lyon, France
Dr Jean Michel Letang, CREATIS Medical Imaging Research laboratory, INSA-Lyon

[1] Vidal, F. P. *et al*. “Development and validation of real-time simulation of X-ray imaging with respiratory motion”. In: Computerized Medical Imaging and Graphics (2016), 49, pp. 1–15. doi:10.1016/j.compmedimag.2015.12.002.
[2] Vidal, Franck P; Mitchell, Iwan T; Letang, Jean Michel, "Use of fast realistic simulations on GPU to extract CAD models from microtomographic data in the presence of strong CT artefacts," in Precision Engineering (2021), doi: 10.1016/j.precisioneng.2021.10.014. Online ahead of print.
[3] Haiderbhai, M. et al. “pix2xray: converting RGB images into X-rays using generative adversarial networks”. In: International Journal of Computer Assisted Radiology and Surgery (2020), 15(6), pp. 973–980. doi:10.1007/s11548-020-02159-2.

Project title: Explainable Artificial Intelligence (XAI) using visualization
1st supervisor: Professor Jonathan C. Roberts / School of Computer Science and Electronic Engineering
2nd supervisor: Dr Panagiotis Ritsos / School of Computer Science and Electronic Engineering
Research theme: T3 – novel mathematical, physical and computer science approaches
Project description: Machines and algorithms are becoming smarter. Artificial intelligence algorithms are being used to understand human speech, drive cars, make decisions, compete at games, and recommend products. But how do they work? How do you know that they make the right decisions? How do you know you can trust them? Why did the algorithm choose one way, in comparison to another? Explaining and understanding AI is a huge challenge for today's society. If we, as humans, are to trust these `intelligent algorithms' then we need to know, understand, and explain how they work. Regulatory bodies, Governments, companies and so on are requiring clearer accountability and transparency in the decision processes [1]. Explanatory AI (XAI) is AI where the results can be explained and understood by humans. Data-visualization techniques can be used to explain different phenomena [2,3]. Visual explanations can be ideated that make explicit, ideas, processes, algorithms and so on; techniques that make explicit ideas and decisions that are implicitly encoded by the AI algorithm. This research will investigate different XAI challenges and develop visualization solutions. It will investigate facets to explain; techniques to capture data and metrics from the AI algorithms and processes; designs that visualize them and evaluate techniques to display uncertainty. A design-study approach will be applied [4] and new data-driven and human-computer interactive interfaces created and evaluated. It requires someone with skills in design as well as software engineering. All with the aim to develop trustworthy, justifiable, and contestable explanatory visualizations of AI.

References:

[1] Danding Wang, et al. 2019. Designing Theory-Driven User-Centric Explainable AI. Proceedings of the 2019 CHI. ACM, New York, NY, USA, Paper 601, 1–15. DOI:https://doi.org/10.1145/3290605.3300831

[2] J. C. Roberts, P. Butcher, A. Sherlock, and S. Nason, "Explanatory Journeys: Visualising to Understand and Explain Administrative Justice Paths of Redress," in *IEEE Transactions on Visualization and Computer Graphics*, doi: 10.1109/TVCG.2021.3114818.

[3] J. C. Roberts, P. D. Ritsos, J. R. Jackson and C. Headleand, "The Explanatory Visualization Framework: An Active Learning Framework for Teaching Creative Computing Using Explanatory Visualizations," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 24, no. 1, pp. 791-801, Jan. 2018, doi: 10.1109/TVCG.2017.2745878

[4] J. C. Roberts, C. Headleand and P. D. Ritsos, "Sketching Designs Using the Five Design-Sheet Methodology," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 1, pp. 419-428, 31 Jan. 2016, doi: 10.1109/TVCG.2015.2467271.

Project title: Combining Artificial Intelligence and Mixed Reality visualisation for immersive investigation of river flood risk

1st supervisor: Dr Panagiotis Ritsos / School of Computer Science and Electronic Engineering
2nd supervisor: Dr Sopan Patil / School of Natural Sciences
Research theme: T3 – novel mathematical, physical and computer science approaches
Project description: Hydrological models are widely used to understand, and forecast, a river's flood response to storm events. Combined with a 3-Dimensional representation of the landscape, these models can also aid in determining the areas that are vulnerable to flooding for different storm magnitudes. Recent advances in Artificial Intelligence (AI), and especially Deep Learning (DL), have resulted in the ability to provide efficient high-dimensional interpolators that can handle data of multiple dimensions and heterogeneous information, such as those encountered in hydrological modelling. Coupled with novel immersive technologies, such as Mixed Reality (MR) and Situated Analytics (SA), it is possible to visualize this hydrological information *in situ* and mediate its interaction with spatial data in the context of flood risk analysis.

In this project, our goal is to develop an AI-powered, MR visualisation framework for robust assessment of flood risk scenarios to inform urban planning and decision making that is underpinned by a physics-based hydrological model. Specifically, our project proposes the novel synergy of MR, AI and SA to depict hydrological model parameters onto and interactive, explorable, immersive 3D graphical representations of river basins, coastal zones, and adjacent land, both in lab environments and *in situ*. This approach will involve addressing challenges both in the domains of DL and MR, where novel techniques, for extracting, processing, and visualising hydrological data, in MR is required. These immersive, high-level abstractions of hydrological models and physical river basin data will be used to depict, and interactively explore, the impact of land management decisions (e.g., urban development, conservation) on river basin and coastal zone hydrology.

References:

Rydvanskiy, R.; Hedley, N. Mixed Reality Flood Visualizations: Reflections on Development and Usability of Current Systems. *ISPRS Int. J. Geo-Inf.* 2021, *10*, 82. https://doi.org/10.3390/ijgi10020082

P. W. S. Butcher, N. W. John, and P. D. Ritsos, "VRIA: A Web-based Framework for Creating Immersive Analytics Experiences," *IEEE Transactions on Visualization and Computer Graphics*, vol. 27, no. 07, pp. 3213–3225, Jul. 2021.

Patil, S. D., Y. Gu, F. S. A. Dias, M. Stieglitz, and G. Turk, "Predicting the spectral information of future land cover using machine learning", *International Journal of Remote Sensing*, 38(20), 5592-5607, 2017.

Project title: Multiple Stream Processing of Language - Big Data for NLP

1st supervisor: Dr. William Teahan / Senior Lecturer, School of Computer Science and Electronic Engineering

2nd co-supervisor: Dr. Cristopher Shank / Head of School, School of Languages Literatures and Linguistics

With wider research team: Professor Delyth Prys / Head of Languages Technologies, Canolfan Bedwyr. Dewi Jones / Chief Software Engineer, Language Technology Unit, Canolfan Bedwyr

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

Project description: Big Language Data involves natural language datasets (audio and text) that are extremely large (many terabytes in size). Automatically analysing Big Language Data presents many challenges. These include collection and cleaning, annotation, indexing and storage, as well as retrieval and analysis of huge quantities of data that has been created by humans. Scale is a huge challenge, because when the data is analysed, there can be a tendency for it to grow larger than its raw size. The research in this PhD will investigate new models and innovative approaches to these challenges. The work will focus on techniques of applying smart (AI) techniques to enable dynamic multiple stream processing that will allow for more effective processing of big language data. The novel approach will be applied to building efficient state-of-the-art NLP tools for various tasks such as text classification, emotion recognition and language segmentation, and further applied to previously unresearched novel NLP tasks in the computational linguistics and language technologies areas in collaboration with the project's partners.

References:

[1] Teahan, William John. "A compression-based toolkit for modelling and processing natural language text." *Information* 9, no. 12 (2018): 294.

[2] Agerri, Rodrigo, Xabier Artola, Zuhaitz Beloki, German Rigau, and Aitor Soroa. "Big data for Natural Language Processing: A streaming approach." *Knowledge-Based Systems* 79 (2015): 36-42.

[3] Shank, Christopher, Plevoets, Koen, Van Bogaert, Julie. "That/zero alternation and epistemic parentheticals: Using a diachronic multivariate constructional approach to examine the grammaticalization of *I think*, *I believe* and *I suppose*". Journal of Corpus Linguistics and Linguistic Theory (CLLT). 2016 De Gruyter.

[4] Shank, Christopher, Plevoets, Koen. & Cuyckens, Hubert. A Diachronic Corpus based Multivariate Analysis of "I think that" versus "I think zero" Polysemy and Synonymy. Corpus Methods and Applications in Cognitive Linguistics. Glynn, D & Robinson, J. (Eds). 2014 John Benjamins Publishing Co.

[5] Prys, D. (ed.), Jones, D., Prys, G., Watkins, G., Cooper, S., Roberts, J. C., Butcher, P., Farhat, L., Teahan, W. & Prys, M. Language and Technology in Wales: Volume I. 5 Oct 2021, Bangor: Prifysgol Bangor University. 120 p.

Bristol

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

1st supervisor: Prof Joel Goldstein

2nd supervisor: TBD

Department/Institution: Particle 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 10^16. 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.

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

1st Supervisor: Sudarshan Paramesvaran

2nd Supervisor: Henning Flaecher

Department/Institution: Particle 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, in 2025 the long awaited Run 3 of the LHC begins, 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: Fast Inference for Online Data Selection at DUNE
1st Supervisor: Jim Brooke
2nd Supervisor: Sudarshan Paramesvaran
Department/Institution: Particle Physics, University of Bristol
Research theme: T1 - data from large science facilities
Project description: The Deep Underground Neutrino Experiment (DUNE) is a next-generation long baseline neutrino experiment, which will measure neutrino properties to exquisite precision. In particular, DUNE aims to resolve the neutrino mass hierarchy, as well as discovering CP violation in the lepton sector, which could explain the matter-antimatter asymmetry of the Universe. The Bristol DUNE group is active in data-acquisition and online selection for the DUNE far detector, as well as physics analyses at protoDUNE - a prototype detector operating at CERN. We have developed an object recognition network based on YOLO (You Only Look Once) for fast, online identification of neutrino interactions in the DUNE data. Selecting data of interest is critical since the raw data rate will approach one zettabyte per year. In this project, we will study network acceleration and performance on a range of hardware platforms (including FPGA implementation using hls4ml), construct and operate a demonstrator system for use at protoDUNE, as well as study network performance against a range of neutrino phenomena. Opportunities for developing new ML applications within DUNE/protoDUNE physics analysis will also be available.

Project title: Machine learning and radio source multiplicity

1st supervisor: Prof. Mark Birkinshaw
2nd supervisor: TBC
Department/Institution: Astrophysics, University of Bristol
Research theme: T1 - data from large science facilities
Project description: This project aims to study the long-standing problem of how to associate distinct source components in radio sources with one another in the presence of a wider population of radio sources. Many radio sources found in interferometric surveys are composed of several distinct bright components (the "hot spots" and "central component" among others) within a low surface brightness envelope. Towards the flux density limit of a survey these bright components can appear without the envelope and can be hard to associate as parts of a single source. Eye-based techniques have been used to make this association since the very early days of radio surveys, basing the association on such things as component linearity, or brightness similarity, or symmetry, or asymmetric component structures pointing towards a common centre.

But these techniques are hopelessly inadequate in the era of large-scale surveys and will become completely untenable once the SKA comes into operation. Further difficulty arises at high redshift where many of the distinguishing features of multi-component sources cease to be valid - because of strong Compton losses, interactions with a clumpy intergalactic medium, and source youth. Progress should be possible through using a well-studied survey field as a training set to develop a machine learning algorithm that combines radio morphological and spectral information to develop a high-speed source-finder. Work would start using one of the deep VLA fields at L and S band, and, if successful, would attempt to extend to surveys at other frequencies. Similar techniques might be applicable at other wavebands, for example in identifying clumpy galaxies in the optical or protoclusters in the X-ray.

Cardiff

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 late 2022, 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: Harvesting information from hyperspectral images
1st supervisor: Wolfgang Langbein (PHYSX)
2nd supervisor: Francesco Masia (BIOSI)
3rd supervisor: Paola Borri (BIOSI)
Department/Institution: School of Physics and School of Biosciences/Cardiff University
Research theme:
T2 – biological, health and clinical sciences
T3 – novel mathematical, physical and computer science approaches
Project description: For centuries, light microscopy has been used by biologists and material scientists to unveil the physico-chemical properties of specimens. Scientists used their experience and knowledge to describe the sample by extracting qualitative information from visual assessment, and to recognise and separate different objects in the image. With the development of more sophisticated light sources, optics, labelling, and detectors, the last few decades have witnessed a rapid emergence of novel advanced imaging modalities.

Among those, hyperdimensional imaging (HDI) adds a further dimension to the spatial ones to provide a step change in information content. For example, adding the light spectrum (indeed HDI is often also called hyperspectral imaging, HSI), providing for each spatial pixel a wavelength dependent intensity. In fluorescence microscopy, detecting the light emitted by fluorophores tagged to a biomolecule, HSI enabled imaging many biomolecules at once and correlate their spatial distribution. In environmental monitoring, satellite HSI allows the measurement of deforestation, pollution, etc. However, the increased information contained in the multi-dimensional voxel precludes their visual analysis, calling for suited algorithms.

Driven by the necessity of harvesting such a wealth of information, we are developing efficient computational tools to extract salient quantities. Our analysis concept is based on a dimensionality reduction while exploiting prior knowledge such as physical constrains, to condense the HSI data into few quantitative properties. These are then used as input for machine learning approaches to identify and classify objects in the image. Owing to the generality of such methods, we were able to gain insight of both biological and inorganic specimens investigated with a range of imaging modalities, from vibrational microscopy (based on Raman [1] and Brillouin [2] scattering) to fluorescence lifetime imaging (FLIM) [3], as well as electron microscopy techniques such as low energy electron microscopy (LEEM) and scanning or transmission electron microscopy (SEM or TEM) energy dispersive x-ray (EDX).

In this project, the student will join research in our group towards the development of new algorithms for HSI data analysis, encompassing the following lines of investigation
Solve a major issue in Raman vibrational imaging by factorization into separate fluorescence and Raman signals
Extend the uFLIM-FRET method [3] to multiple donor-acceptor pairs and non-resonant processes.
Improve the analysis of Brillouin data [2] by introducing additional physical constrains to the component spectral shape
Explore and validate new methodologies for the unsupervised classification of objects extracted from analysis of HSI data [4]
Extend the application of the method to electron microscope techniques [5]

References:

[1] F Masia, A Glen, P Stephens, P Borri, W Langbein, *Quantitative chemical imaging and unsupervised analysis using hyperspectral coherent anti-Stokes Raman scattering microscopy*
Analytical Chemistry 85, 10820 (2013) – 10.1021/ac402303g
[2] F Palombo, F Masia, S Mattana, F Tamagnini, P Borri, W Langbein, D Fioretto, *Hyperspectral analysis applied to micro-Brillouin maps of amyloid-beta plaques in Alzheimer's disease brains*, Analyst, 143, 6095 (2018) – 10.1039/C8AN01291A
[3] F Masia, P Borri, W Dewitte, W Langbein, *uFLIM - Unsupervised analysis of FLIM-FRET microscopy data*, arXiv:2102.11002v2
[4] F Masia, I Pope, P Watson, W Langbein, P Borri, *Bessel-Beam Hyperspectral CARS Microscopy with Sparse Sampling: Enabling High-Content High-Throughput Label-Free Quantitative Chemical Imaging*, Analytical Chemistry 90, 3775 (2018) – 10.1021/acs.analchem.7b04039
[5] F Masia, W Langbein, S Fischer, JO Krisponeit, J Falta, *Low-energy electron microscopy -- intensity-voltage data factorization, sparse sampling, and classification*, submitted

Project title: Exploiting GAIA data and understanding the galaxies' past histories with machine learning

1st supervisor: Mikako Matsuura

2nd supervisor: Tim Davies

Department/Institution: Physics and Astronomy

Research theme: T1 – data from large science facilities

T3 – novel mathematical, physical and computer science approaches

Project description: GAIA has mapped more than billion stars in the Milky Way and nearby galaxies, building how galaxies have made current shapes.

Amongst stars, some of the stars have specific characters in their spectra or photometry, and these stars can be used to trace how the stars were formed at a specific era.

With this PhD project, the student is expected to explore GAIA data with machine learning, identify these specific stars, and pin down the location of star formation in specific ages.

Project title: Improving the sensitivity of quantum-enhanced interferometers through machine learning

1st supervisor: Prof Hartmut Grote
2nd supervisor: Dr Katherine Dooley
Department/Institution:

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

Project description: Laser interferometers have long been used to probe fundamental physics phenomena, from disproving the existence of the theorized ether well over a hundred years ago to making the first direct detections of gravitational waves in recent years. Given the extraordinary measurement precision that can be achieved with the latest technological advances in laser interferometer design, they can now also be used to search for potential quantization of spacetime, ultra-high frequency gravitational waves and dark matter.

A critical aspect of pushing the high-frequency sensitivity to new limits involves minimizing optical losses in the path of a squeezed light field injected at the output port of the interferometer. One such source of loss comes from imperfectly matched beam parameters. This project will require the student to use machine learning to train a set of a few dozen piezo-electric actuators to dynamically change the shape of a mirror. Other applications of machine learning can be explored, such as for adaptive control loops and complex noise analysis. The project will be at the cross-section of AI/ML and experiment.

References:

"An experiment for observing quantum gravity phenomena using twin table-top 3D interferometers" Classical and Quantum Gravity, 38(8):085008 (2021); https://doi.org/10.1088/1361-6382/abe757

"Direct limits for scalar field dark matter from a gravitational-wave detector" https://arxiv.org/abs/2103.03783

"First Demonstration of 6 dB Quantum Noise Reduction in a Kilometer Scale Gravitational Wave Observatory" Physical Review Letters 126(4), 2021; https://doi.org/10.1103/PhysRevLett.126.041102

Project title: Machine Learning and Artificial Neural Network Accelerated Computational Discoveries in Catalysts Materials for Hydrogen Generation and Storage

1st supervisor: Dr Bo Hou, School of Physics and Astronomy, Cardiff University

2nd supervisor: Prof. Philip Davies, School of Chemistry, Cardiff University

3rd supervisor: Dr Jingchao Zhang, NVIDIA AI Technology Center

Research theme:

T1 – data from large science facilities

T3 – novel mathematical, physical and computer science approaches

Project description: Catalysts have attracted growing interest due to their unique effects on chemical reactions, drug discovery and health care. In particular, catalysts materials innovation is the key for addressing Net Zero and Clean Growth challenges. Artificial intelligence has been referred to as the "fourth paradigm of science," and as part of a coherent toolbox of data-driven approaches, machine learning (ML) dramatically accelerates computational materials discoveries. As the machinery for ML algorithms matures, significant advances have been made not only by the mainstream AI researchers but also those who work in materials science. As a result, the number of ML and artificial neural network (ANN) applications in materials innovation and pharmaceutic molecular design is growing at an astounding rate.

We would like to work with an enthusiastic PhD student to develop new algorithms for machine learning to screen and understand the structural factors for catalysis materials design, focusing on hydrogen generation and storage. At the end of their PhD, they should be able to produce high-quality codes and high-performance computing skills as well as contribute solutions to Net Zero and Clean Growth challenges. Developing new materials design methods and big data indexing innovations (from literature pools) to extract the new catalysis materials design principles will be encouraged as part of the student's research and is an area for high-impact publication. This is not only a transdisciplinary project, but it also intends to foster a broader understanding of catalysis materials and AI data mining, which is fit for current challenges in energy technologies, from clean energy generation to low carbon footprint energy storage.

The student will gain fundamental knowledge in machine learning, state-of-the-art of catalysis science and technologies, as well as practical experiences in coding using high-performance computing facilities. Equipped with these skills, the student will be highly competitive and sought after both in industry and academia.

References:

1. Hou* et al., *Catalysts* 2021, 11(9), 1129. https://doi.org/10.3390/catal11091129
2. Zhang* et al., *Phys. Chem. Chem. Phys.*, 2021, 23, 19166-19172. https://doi.org/10.1039/D1CP01715B
3. Zhang* et al., *International Journal of Heat and Mass Transfer* 2021, 171, 121073. https://doi.org/10.1016/j.ijheatmasstransfer.2021.121073
4. Zhang* et al., *WIREs Computational Molecular Science* 2019, 10, e1450. https://doi.org/10.1002/wcms.1450
5. Zhang* et al., *Carbon* 2019, 148, 115-123. https://doi.org/10.1016/j.carbon.2019.03.046

Swansea

Project title: Topological excitations and colour confinement in gauge theories

1st supervisor: Biagio Lucini
2nd supervisor: Alma Rahat
3rd supervisor: Daniel Archambault
Department/Institution: Mathematics, Swansea
Research theme: T1 – data from large science facilities
Project description: The strong interaction is the fundamental force that holds together nuclear matter. This force is described by Quantum Chromodynamics, a non-Abelian gauge theory whose fundamental particles are six fermion flavours (the quarks) interacting with eight gauge bosons (the gluons). Quarks and gluons never appear in final states of the strong interaction. This phenomenon, for which we lack an explanation, is known as colour confinement. Elegant mechanisms of colour confinement have been proposed that are based on the prominent role of field configurations with non-trivial topological properties. However, the underlying arguments are semiclassical, and would hence need to be supported by first-principle calculations such as those performed in the framework of Lattice Field Theories. The latter in turn would need to be guided by topological arguments and optimisation procedures, and would benefit from advanced visualisation methods that enable us to inspect the relevant configurations. The goal of this highly interdisciplinary project is to develop a methodology for detecting configurations with non-trivial topological properties combining large-scale numerical simulations, optimisation procedures, methods in topological data analysis and visualisation, in order to assess their relevance in mechanisms of colour confinement.

Project title: Ultrafast pulse shaping via machine-learning-enabled multi-element adaptive optics and levitated optomechanics.

1st supervisor: Dr. Kevin O'Keeffe

2nd supervisor: Dr. James Bateman

Department/Institution: Physics Department, Swansea University

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

Project description: The ability to optimally focus and tailor the structure of ultrafast laser pulses is fundamental to many applications including surgery, high-resolution optical imaging, and nonlinear microscopy. Multi-element adaptive optics (AO) provides a powerful approach for manipulating the structure of ultrafast pulses but necessitates the development of new techniques capable of measuring and optimizing the laser field throughout the focal volume. The ability of machine learning to rapidly determine the solution to inverse problems makes it ideally suited to multi-element AO systems in which frequently an output, such as mode profile, is known but the required AO response, which has many degrees of freedom, is unknown. In this project a model-based Deep Neural Network will be developed to control the response of a multi-element AO system consisting of a spatial light modulator and deformable mirror and will be trained using data from a range of laser pulse diagnostics. Light scattered from a levitated nanoparticle, which is rapidly scanned through the focal plane using the optical dipole force and active stabilization, will allow the laser field throughout the focal volume to be characterized. The combination of a machine-learning-enabled multi-element AO system and rapid volume characterization will form a closed-loop shaping system with applications in 3D laser micro-fabrication, data storage and electron microscopy.

Project title: Analogue Computing for Advanced Machine Learning and Sensing

Joint supervisors: Sahar Basiri Esfahani, Gert Aarts

Department/Institution: Department of Physics, Swansea University

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

Project description: Today's learning algorithms mostly operate using silicon-based graphical processing units (GPU) in digital computers. These consume a considerable amount of power. Recently, analogue learning schemes have attracted great research interest [1,2]. They mimic the function of biological neurons that, we believe, are a perfect example of where nature has developed 'optimal neural design'.

In this project we will investigate how engineered quantum systems can be exploited to enhance machine learning (ML). In particular, the project targets important applications in analogue learning machines and sensing by using unique features of physical systems in the quantum regime. In this project we will develop a theoretical model, computational platform, and parameter estimation for a novel category of highly precise sensors and efficient learning machines which are experimentally realisable. Moreover, the project aims at using advanced ML algorithms [3] to design state-of-the-art sensors operating in the quantum regime.

This research combines current hot topics – ML and quantum physics – and provides the opportunity for an exceptional doctoral candidate to undertake research involving both theoretical and computational development within a research team composed of scientists in computational intelligence and quantum physics.

References:

[1] M. J. Kewming, et al. "Designing a physical quantum agent", *Phys. Rev. A* 103, 032411 (2021).

[2] N. Tezak and H. Mabuchi, "A coherent perceptron for all-optical learning", *EPJ Quantum Technol.* 2.1, 1-22 (2015).

[3] C. Klos, et al. "Dynamical learning of dynamics", *Phys. Rev. Lett.* 125, 088103 (2020).

Project title: Learning (from) lattice field theory
1st supervisor: Prof Gert Aarts
2nd supervisor: Prof Biagio Lucini / Prof Chris Allton
Department/Institution: Physics Department, Swansea University
Research theme: T1 – data from large science facilities
Project description: Currently, machine learning approaches to analyse large data sets are being developed in the context of lattice field theory. Conversely, lattice field theory methods may provide new insights in the development of machine learning algorithms (implementation of local symmetries, using QFT-based analysis). In this project we will explore these topics, with the goal of arriving at results for QCD.

Project title: Robust Parameter Optimisation for Image Segmentation

1st supervisor: Dr Alma Rahat
2nd supervisor: Dr Daniel Archambault
Department/Institution: Computational Foundry, Swansea University.
Research theme: T2 – biological, health and clinical sciences
T3 – novel mathematical, physical, and computer science approaches

Project description: Medical professionals rely on robust image segmentation algorithms to highlight anomalous tissues in a patient, e.g., cancer. However, image segmentation algorithms are typically calibrated on a limited number of training images through a tedious trial and error process. This gives limited context to the robustness against overfitting which makes it difficult to predict how well the segmentation algorithm will perform on unseen images. This could lead to an incorrect diagnosis. The goal of this project is to use a combination of visualization techniques (e.g., Tuner [1]) and Bayesian model calibration techniques (e.g., history matching [2]) to develop a system for robust parameter optimisation with a focus on image segmentation algorithms.

References:

[1] Torsney-Weir, T., Saad, A., Moller, T., Hege, H.C., Weber, B., Verbavatz, J.M. and Bergner, S., 2011. Tuner: Principled parameter finding for image segmentation algorithms using visual response surface exploration. IEEE Transactions on Visualization and Computer Graphics, 17(12), pp.1892-1901.

[2] Andrianakis, I., Vernon, I.R., McCreesh, N., McKinley, T.J., Oakley, J.E., Nsubuga, R.N., Goldstein, M. and White, R.G., 2015. Bayesian history matching of complex infectious disease models using emulation: a tutorial and a case study on HIV in Uganda. PLoS computational biology, 11(1), p.e1003968.

‘Additional projects’

Aberystwyth

Project title: Deciphering how stroke impairs spatial navigation during walking
1st supervisor: Otar Akanyeti (CS)
2nd supervisor: Federico Villagra (IBERS)
3rd supervisor: Sebastian McBride (IBERS)
Research theme: T3 – novel mathematical, physical, and computer science approaches
Project description: Stroke disrupts various cognitive processes including memory formation, orientation, executive function, attention, and judgement (Lugtmeijer et al., 2021). While relationship between post-stroke cognitive impairments and functional outcomes is well documented, much less is known about to what extent these impairments affect people’s ability to navigate in the real world, from low level locomotor control (e.g., maintaining balance) to high level path planning (e.g., detection and avoiding obstacles) (van der Ham et al., 2013). Diminished navigational skills has serious repercussions on independence and quality of life by increasing the risk of injury and falling.

The new PhD student will combine experimental tools including brain imaging, wearable eye tracking and motion analysis, and virtual reality to study how stroke (compared to age-matched controls) impairs spatial navigation and to create data-driven computational models/algorithms to elucidate stroke-impaired navigation strategies. These models/algorithms should be human-interpretable and biologically plausible so that they can help scientists generate new research hypotheses about stroke and navigation as well as rehabilitation interventions. While developing and testing these models, the student will use state-of-the-art system identification and machine learning approaches (e.g., NARMAX, fuzzy logic and evolutionary artificial neural networks).
The student will be part of a vibrant and multi-disciplinary research group (Aberystwyth Stroke Research Group, stroke.aber.ac.uk) including computer scientists, health and exercise physiologists, molecular biologists, psychologists, and neuroscientists. The project will be delivered in close collaboration with Prof Derek Jones and Dr Maryam Afzali from Cardiff University Brain Research Imaging Centre. In addition, the student will gain first-hand experience in interacting with real patients and health care professionals according to NHS ethical guidelines and data protection regulations.

We are looking for someone with analytical skills, and who is motivated with a strong passion for science. A good degree (2:1 minimum) in Computer Science or related fields, prior knowledge in data mining and modelling, and strong interest in human health and cognitive neuroscience is desired.

References:

Lugtmeijer, S., Lammers, N.A., de Haan, E.H.F., de Leeuw, F.-E., Kessels, R.P.C., 2021. Post-Stroke Working Memory Dysfunction: A Meta-Analysis and Systematic Review. Neuropsychology Review 31, 202-219.

van der Ham, I.J., Kant, N., Postma, A., Visser-Meily, J.M., 2013. Is navigation ability a problem in mild stroke patients? Insights from self-reported navigation measures. J Rehabil Med 45, 429-433.

Project title: Machine Learning and Traditional Celtic Music Composition
1st supervisor: John Gough (Physics)
2nd supervisor: TBD
Research theme: T3 – novel mathematical, physical, and computer science approaches
Project description: The aim of this project is to train a neural net to complete Jigs or Reels based existing databases of Irish music starting with the well-known collection O'Neill's 1001 Irish Gems. (This is a natural choice as there are pre-existing large databases of midi files.)

One example where this has been tried is in the AI Music Generation Challenge where the competition winning Irish jig was generated by neural net using "artificial critics" to select tunes according to on a metric and intervallic metric [B.L. Sturm, AIMC, 2021].

From an aesthetic point of view, the jig sounds unnatural and would never be played by traditional Irish musicians, and probably not even recognized as being Irish. The more natural approach, essentially the one followed by a human composer, would be to start off with a short motif (say 3 or 4 notes that sound good) then work out where this should go in a melody. This means that you have a subsequence and would want to work out where best to place it in a fixed structure, then fill in the remaining notes: in other words, to train the network to position a subsequence within a grid, then fill in the remaining elements.

The user input would be a short motif. This would repeat at various stages (mostly deterministic) and the real problem is to train the network to build the rest of the melody around it. From what I can see, most approaches involved getting the network to recognize patterns which are already well-known to musicians, rather than dealing with the real problem of extending a short motif to make a tune. RNN's and LSTM have been widely used, with emphasis on going both backwards and forwards over the sequences generated, however, we a central role would be given by the hierarchical structure [motif, 2-bars, 4-bar phrase (w/o cadences), 8-bar period (antecedent/consequent), sequence of periods]. This would be the novelty in the learning process.

Related work would include conversion of Welsh/Scottish traditional music to Irish form; learning from Scottish traditional music databases; categorization of tunes by region; comparison with original compositions.

Project title: Few shot learning for 3D plant phenomics
1st supervisor: Bernard Tiddeman
2nd supervisor: John Doonan
Research theme: T3 – novel mathematical, physical, and computer science approaches
Project description: The 3D analysis of plants has become increasingly effective in modelling the relative structure of organs and other traits of interest. Our previous work has developed a pipeline for 3D plant data capture (using multi-view images and novel feature point detection [1] and matching algorithms [2]), 3D segmentation using Deep Learning [3] (using our PatternNet algorithm [4]), and 3D measurement [5] (e.g. using RANSAC). A disadvantage of the current pipeline is the need for considerable amounts of training data for the deep learning segmentation and classification algorithms. Few-shot learning algorithms have been applied successfully in many computer vision problems to enable a classifier to generalize to new classes with very few examples. Recent work has developed a few-shot learning approach for 3D point clouds of indoor scenes / rooms. We have found 3D plant segmentation to be a much more challenging problem, with more complex branching structures and more noise, where previous supervised algorithms have failed. This PhD will investigate novel few-shot learning algorithms that are robust for this particularly challenging case.

References:
M. Ghahremani, Y. Liu and B. Tiddeman, "FFD: Fast Feature Detector," in IEEE Transactions on Image Processing, vol. 30, pp. 1153-1168, 2021.
Morteza Ghahremani, Yitian Zhao, Bernard Tiddeman, Yonghuai Liu,
Interwoven texture-based description of interest points in images,
Pattern Recognition, Volume 113,2021.
Ghahremani Morteza, Williams Kevin, Corke Fiona M. K., Tiddeman Bernard, Liu Yonghuai, Doonan John H., "Deep Segmentation of Point Clouds of Wheat", Frontiers in Plant Science, Volume 12, 2021.
Ghahremani M., Tiddeman B., Liu Y., Behera A., "Orderly Disorder in Point Cloud Domain", ECCV, 2020.
Morteza Ghahremani, Kevin Williams, Fiona Corke, Bernard Tiddeman, Yonghuai Liu, Xiaofeng Wang, John H. Doonan, "Direct and accurate feature extraction from 3D point clouds of plants using RANSAC", Computers and Electronics in Agriculture, Volume 187,2021.
Na Zhao, Tat-Seng Chua, Gim Hee Lee, "Few-shot 3D Point Cloud Semantic Segmentation", CVPR, 2021.

Project title: Anytime Analysis for Dynamic Optimisation Problems
1st supervisor: Thomas Jansen
2nd supervisor: Christine Zarges
Research theme: T3 – novel mathematical, physical, and computer science approaches
Project description: Many optimisation problems are too difficult to be solved efficiently by standard algorithms. Heuristic optimisation methods like evolutionary algorithms or simulated annealing are frequently applied in these situations. The theory of these heuristics still puts an emphasis on runtime analysis and is at odds with the way these heuristics are actually applied. This project addresses this gap by concentrating on anytime analysis targeting dynamic problems that change over time.
Building on existing anytime analysis results (sometimes also called fixed budget results [2]) as well as recent results from fixed target analysis [1], the project performs a systematic study of dynamic optimisation. The starting point are simple static unimodal and multimodal benchmarks [3] and simple different tools to construct dynamic optimisation problems from static ones. Starting from simple heuristics like random sampling and local search the tools and methods are developed to compare these baseline methods with more advanced methods, employing populations, crossover, and different approaches to deal with dynamic optimisation problems like hall of fame approaches or diploidy.

References:

[1] M. Buzdalov, B. Doerr, C. Doerr, and D. Vinokurov (2020): Fixed-target runtime analysis. In Proceedings of the 2020 Genetic and Evolutionary Computation Conference (GECCO 2020), ACM Press, pages 1295-1303. https://doi.org/10.1145/3377930.3390184

[2] T. Jansen (2018): Analysing stochastic search heuristics operating on a fixed budget. In B. Doerr, F. Neumann (Eds.): Theory of Evolutionary Computational Recent Advances. Springer, pages 249-270. https://doi.org/10.1007/978-3-030-29414-4_5

[3] T. Jansen and C. Zarges (2016): Example landscapes to support analysis of multimodal optimisation. In Proceedings of the 14th International Conference on Parallel Problem Solving From Nature (PPSN XIV). Springer, LNCS 9921, pages 792-802. https://doi.org/10.1007/978-3-319-45823-6_74

Project title: Modelling seafloor change

1st supervisor: Hannah Dee

2nd supervisor: Jan Hiddink

Department/Institution: Department of Computer Science Aberystwyth University and Ocean Science Bangor University

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

Project description: The purchase of a new underwater robot (a Sparus II submarine from Iqua Robotics) with 2- and 3-dimensional imaging capabilities opens up new avenues for detailed spatio-temporal monitoring of sea floor state via robotic sensing. This project would be interdisciplinary, developing skills in AI, robotics, and marine science. The robot will be programmed to follow a survey path at a distance to the seafloor and capture datasets (sidescan SONAR and visible spectrum imaging) of the same area at multiple time points. These datasets would be large and unique, with computational challenges to solve in three main areas:

- Registration, both between modalities and across times.
- Object detection and recognition in underwater domains.
- The extraction of image features from 2d and 3d which correspond to change in the marine environment.

When a vessel is anchored the mooring chain may be dragged repeatedly across the seabed in an arc around the anchor as the vessel moves with tidal motion (Davis et al. 2016). This chain-scour is likely to impact the sediments and it is necessary to quantify the impacts of this anchor footprint on benthic communities. Broad et al. (2020) found that 90% of studies of anchoring and mooring focused on recreational vessels and identified that empirical investigations of anchor scour stemming from large merchant ships are scant and require urgent attention. The aim of this research is to gain an understanding of the effects of disturbance by anchoring activity of merchant vessels on the soft-sediments at Red Wharf Bay. First the student will develop a method for using AIS Ship Data to estimate anchoring intensity. We will use AIS vessel tracking data (obtained through an existing collaboration with Global Fishing Watch) to identify the area impacted by anchors and anchor chains. Anchored vessels in tidal areas will move around the anchor in a half or full-circle, which means that the position of the anchor and the paths of the anchor chain can be inferred from the circle of AIS locations that is transmitted by anchored vessels. Areas with different intensities of anchoring activity will then be surveyed four times using the Sparus II AUV over the period of a year to quantify differences in sediment topography. We expect sediment topography to change throughout the year in anchored sites, and during storms in unanchored sites.

- Repeated surveys using side scan sonar and video
- Compare difference in topography between surveys
- Relate difference to anchoring activity between surveys
- If possible, capture imagery of anchoring in action, which is impossible using other methods because you cannot get close using a boat or divers.

The project would enable new work to be done in underwater imaging and also marine science.

Project title: Prediction of facial growth for children with cleft lip and palate using 3D data mining and machine learning

1st supervisor: Richard Jensen

2nd supervisor: TBD

Research theme: T2 – biological, health and clinical sciences

T3 – novel mathematical, physical, and computer science approaches

Project description: Approximately 150 children are born in England and Wales each year with complete unilateral cleft lip and palate (cUCLP). Despite improvements in clinical outcomes in the UK over the past 15 years, between 20-25% children with cUCLP have poor facial growth compared to 3% of the non-cleft Caucasian population. Poor facial growth results in poor aesthetic appearance and poor dental occlusion which can negatively impact on a child's psychosocial development with long-lasting effects. It is not clear why only some children with cUCLP have poor growth, nor why facial growth outcomes vary between surgeons and centres. A number of explanations has been advanced including extrinsic factors such as poor surgery in cleft palate repair during infancy, surgical technique and timing, and intrinsic factors such as the congenital absence of the upper lateral incisor, or the shape of the infants' upper arch, indicating a genetic cause. The relationship of the upper dental arch to the lower arch reflects mid-face growth and can be assessed as early as 5 years using the 5-year index. Children with cleft lip and palate in the UK have been treated in regional specialist centres since 2000 and facial growth is routinely assessed between the ages of 5 and 6 years in this way. It is also routine for cleft centres to take and keep a dental model of the upper arch of infants with cUCLP before they have any surgery.

This project would involve the development of techniques for both 3D data mining and machine learning for the scanned models of infants with cUCLP, in order to determine which features are most predictive of facial growth outcome and if a predictive model can be learned. The maxillary arch models taken from infants prior to their first surgical procedure will be used along with the 5-year index score to develop models via machine learning and identify important regions. In particular, the identification of an intrinsic neonatal arch shape that is predictive of detrimental facial growth would give an opportunity to explain prognosis and manage expectations more easily with parents. It would also facilitate research on the development of new techniques for earlier treatment of poor facial growth and more personalised care for individual patients.

Bristol

Project title: Machine Learning for CMS Trigger

1st Supervisor: Sudarshan Paramesvaran
2nd Supervisor: Jim Brooke
Department/Institution: Particle Physics, University of Bristol
Research theme: T1 – data from large science facilities
Project description: A major upgrade of the Large Hadron Collider from 2024 will increase the rate of proton collisions to an unprecedented 10 billion per second. A sophisticated data processing system - the “L1 Trigger” - will process each event to identify interesting collisions for storage and further analysis; events not selected are lost forever. The L1 Trigger is constructed of custom hardware processors using high-speed optical links and FPGA devices. It can receive data at 60 Tb/s and must process each event in less than 10μs. In this project we will develop fast machine learning algorithms to identify interesting collisions, such as production of a pair of Higgs bosons, for the CMS L1 Trigger. The physics performance of these networks will be studied and optimised using Monte-Carlo simulation. The trade-off between physics performance and inference latency and resource usage will be studied for a range of network types. Ultimately, candidate algorithms will be implemented in FPGAs and demonstrated in real processor boards.

Cardiff

Project title: Revealing the Hidden Light: Understanding the cold interstellar medium in galaxies

1st supervisor: Dr Matthew Smith
2nd supervisor: Prof. Stephen Eales
Department/Institution: School of Physics and Astronomy, Cardiff University
Research theme: T1 – data from large science facilities
T3 – novel mathematical, physical and computer science approaches
Project description: Half the light ever emitted by stars in the Universe has been absorbed by cosmic dust. This increases to >90% emitted 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 dark regions.

Our group in Cardiff is leading several international teams to obtain the best possible data in the submillimetre and millimetre to complement our existing maps from the Herschel Space Observatory. This includes large surveys ongoing on the James Clark Maxwell Telescope (JCMT) in Hawaii, IRAM in Spain, the Large Millimeter Telescope (LMT) in Mexico, and ALMA in Chile, to observe M31, M33 and other local galaxies (for example the DOWSING & Dustpedia surveys). In particular, the HASHTAG survey on the JCMT will produce the first ground-based sub-mm image of Andromeda increasing our resolution by over a factor of ~5. The student will have the opportunity to join these survey teams and help with the data collection.

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.

Project title: Simulation-based Inference of gravitational waves signals from black holes and neutron stars

1st supervisor: Dr Vivien Raymond

2nd supervisor: Prof Stephen Fairhurst

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

Research theme: T1 – data from large science facilities

Project description: Black holes and neutron stars are the densest objects in the universe, well beyond what we can produce in a laboratory and at the very edge of our understanding of physics. They lead to puzzling physical consequences, in particular regarding the behaviour of space and time. When they collide, they produce the most violent events in the universe, shaking space and time and creating gravitational waves: ripples in the fabric of spacetime which propagate away at the speed of light. Gravitational-waves were observed for the very first time in September 2015, when two colliding black holes were detected by the LIGO-Virgo collaboration. Since then, several signals have been observed, and we were able to characterise the black holes and neutron stars at the source of those gravitational waves.

This characterisation currently involves stochastic sampling methods with a very high computational cost, and simplified assumptions of the detectors' properties. This project will leverage modern advances in likelihood-free inference methods, and in particular simulation-based inference, to solve this inference problem accurately. Our approach will adapt automatically to changing features in the detector noise, allow for new data to be continuously included, and will be applicable to the upcoming new generation of gravitational-wave detectors. Gravitational-wave sources are laboratories where we can measure in neutron stars the equation of state of matter at densities otherwise unattainable, and test General Relativity in the strong field regime. Inference of their extragalactic population enables new understandings of the Universe's structure of matter, and independent measurements of the Universe's expansion.

Project title: Inferring the gas mass surface density of dense star-forming clouds at high-angular resolution using machine learning

1st supervisor: Nicolas Peretto

2nd supervisor: Tim Davis

3rd supervisor: Paul Clark

Department/Institution: Cardiff University

Research theme: T1 – data from large science facilities

Project description: H_2 column density (N_{H2}), or equivalently mass surface density, is a fundamental parameter of star formation models. On galaxy scales, it is believed to be a good predictor of star formation rates (e.g. [1]). On parsec scales and below, it is often used to identify sub-structures, such as filaments and cores, i.e. the direct progenitors of individual stars. Hence, our ability to derive accurate mass surface is central to our understanding of star formation throughout the Galaxy and beyond. For the past 10 years, H_2 column density images of star-forming clouds have been mostly obtained using Herschel continuum data (e.g. [2]). While extremely powerful, Herschel column density images have the drawback to be, at best, at 18" resolution, which is too low to probe the internal structures of most star-forming clouds within the Milky Way. On the other hand, molecular line observations of these clouds obtained with some of the most powerful (sub-)millimetre telescopes around the world such as ALMA and NOEMA provide a detailed view of these clouds at arcsecond resolution (e.g. [3]). However, at the moment, deriving H_2 column density images from such observations is very time consuming as it requires the modelling of every single spectrum, each ALMA/NOEMA observations typically holding several tens of thousands of them. In order to speed up this process, the student will use machine learning, extending the analysis we published in [4]. In that study we showed how one can predict H_2 column densities from molecular line emission using the Random Forest algorithm.

The student will use the Herschel mass surface density and dust temperature images of Galactic star-forming clouds obtained at 18" resolution alongside ALMA integrated N_2H^+(1-0) intensity and dust continuum images at ~1" resolution as an input to the Random Forest algorithm. N_2H^+(1-0) is the main contributor to reproducing the high-density parts (A_v>10) of the H_2 column densities [4], so the proposed combination of data will allow us to accurately recover the H_2 column density images of our cloud sample at the same angular resolution as our ALMA/NOEMA data. For this project to work, we require a training dataset. For this, we will use single-dish large-scale N_2H^+ mapping of nearby star-forming clouds recently performed within two IRAM 30m large programmes, ORIONB (PI: J. Pety) and LEGO (PI: J. Kauffmann). The physical resolution of these observations (27" at 400pc -> ~0.05pc) are very similar to those of our clump sample (1" at 4kpc -> ~0.03pc). If needed, we will also use synthetic observations obtained from numerical simulations produced locally in Cardiff (by Dr P. Clark) to train the algorithm. The outcome of this PhD project could be a real game-changer in the field of Galactic star formation.

References:

[1] Kennicutt, 1998, ApJ, 498, 541

[2] Andre et al, 2010, A&A, 518, 102

[3] Peretto et al., 2013, A&A, 555, 112

[4] Gratier et al. 2021, A&A, 645, 27

Project title: The rise of supernovae: a machine learning application
1st supervisor: Dr Cosimo Inserra
2nd supervisor: TBD
3rd supervisor: TBD
Department/Institution: School of Physics and Astronomy / Cardiff University
Research theme: T1 – data from large science facilities
T3 – novel mathematical, physical and computer science approaches
Project description: Supernovae are catastrophic stellar explosions shaping the visible Universe and affecting many diverse areas of astrophysics. Supernovae arising from massive stars, referred to as core-collapse supernovae, play a major role in many intriguing astronomical problems since they produce neutron stars, black holes, and gamma-ray bursts. We are now living in the golden era of transient astronomy, with roughly 20000 transients discovered per year. The advent of the Legacy Survey of Space and Time (LSST) at the Vera Rubin Observatory will boost the number of yearly discoveries by a factor of 100, providing an unprecedented wealth of data, especially soon after the supernova explosion.

The rise-time of supernovae (from explosion to peak luminosity) has always revealed important connections between the observed transients and their progenitor properties. Extrapolating early data points via machine learning techniques, based on a prior using physical knowledge of the supernova explosion via means of extra hyperparameters, will reveal the progenitor star, the explosion mechanism and the stellar explosion scenario and environment as a function of supernova type and redshift.

Project title: Hunting stellar explosions via artificial intelligence
1st supervisor: Dr Cosimo Inserra
2nd supervisor: TBD
3rd supervisor: TBD
Department/Institution: School of Physics and Astronomy / Cardiff University
Research theme: T1 – data from large science facilities
T3 – novel mathematical, physical and computer science approaches
Project description: Supernovae are catastrophic stellar explosions shaping the visible Universe and affecting many diverse areas of astrophysics. Supernovae arising from massive stars, referred to as core-collapse supernovae, play a major role in many intriguing astronomical problems since they produce neutron stars, black holes, and gamma-ray bursts. We are now living in the golden era of transient astronomy, with roughly 20000 transients discovered per year. The advent of the Legacy Survey of Space and Time (LSST) at the Vera Rubin Observatory will boost the number of yearly discoveries by a factor of 100, creating a new data challenge. Task-specific algorithms employed until now for transient's classification have limitations in taming the zoo of transients.
The main project goal is to develop an Artificial Intelligence tool (deep learning algorithm) that can process time-series (e.g. luminosity evolution) and non-time-series (e.g. environment information) and that can identify core-collapse supernovae in two weeks from explosions, which is when we can retrieve crucial information about the progenitor nature. A secondary goal is to build such an AI tool in a way that is scalable enough to be applied to the environment of compact stars mergers producing gravitational waves. This application can predict the merger type (what objects are merging and their masses) and allow for rate and population studies at far distances.

Project title: The darkness that lurks in galaxies: machine learning approaches to revealing black holes and dark matter.

1st supervisor: Timothy Davis

2nd supervisor: Edward Gomez

Department/Institution: Physics and Astronomy

Research theme: T1 – data from large science facilities

T3 – novel mathematical, physical and computer science approaches

Project description: Understanding the formation and evolution of galaxies is a vital part of modern astrophysics. Several of the key components of our galaxy evolution paradigm are “dark” however, meaning they are hard to detect and characterise using directly emitted light. From central black holes to vast dark matter halos, only by obtaining a robust picture of these “dark” constituents of galaxies can we truly understand the physical processes driving galaxy evolution. The kinematics of the gaseous components of galaxies provide a key probe of these dark components. In the upcoming decade, next-generation observatories will reveal such motions in millions of galaxies, but our tools for interpreting this data are not fit for the "big data" era. A student taking on this project would develop innovative fast, unsupervised (or self-supervised) kinematic modelling techniques based on physics-aware convolutional auto-encoders, and apply them to new data from state-of-the-art telescopes around the world to help usher in the new era of data-intensive astronomy.

References:

Dawson et al. 2021, Monthly Notices of the Royal Astronomical Society, Volume 503, Issue 1, pp.574-585

Davis et al. Nature, 2013, 494, 328-330

Swansea

Project title: Multimodal analysis of Anatomical and Functional features to enhance the understanding of Brain Processing Phenomena: A Machine Learning Approach.
1st supervisor: Dr Su Yang
2nd supervisor: Prof Xianghua Xie
Collaborator: Dr Jiaxiang Zhang (Cardiff University)
Department/Institution: Department of Computer Science
Research theme: T2 – biological, health and clinical sciences
T3 – novel mathematical, physical and computer science approaches
Project description: This proposal looks to explore the combination of human MRI/fMRI and EEG/MEG -based biomarkers associated with changes in the human brain and develop statistical techniques to use these features for a better understanding of the healthy lifespan process. Neuroimaging allows to study the brain anatomical composition, the architecture of brain fibre tracts, and the brain metabolic energy consumption, which support the brain function and cognitive processing. A personalized fingerprint of brain functional and anatomical features could then be created through the combined analysis of multimodal neuroimaging data.

An important application of neuroimaging techniques for anatomical and brain functional mapping is during the surgical intervention of intractable epilepsy. For example, structural anomalies can be mapped using the magnetic resonance imaging (MRI), while the electroencephalogram (EEG) is essential for studying the abnormal fluctuations of brain activity during the epileptic crisis. Particularly, EEG is non-invasive, affordable, portable, and indispensable for accurately tagging the crisis periods when combined with video-monitoring. Magnetoencephalography (MEG) technique is also useful in epileptic studies as it shares many similarities with EEG, and it has been argued that it provides more accurate solutions. It was observed that the combination of EEG and MEG produce improved results. The study of the anatomical and functional connectivity is also critical in improving the understanding of the neural information processing.

We initially will take advantage of data from ADNI and HCP datasets [1][2] to test standard methods and the novel techniques, as these datasets have been already evaluated in many peer-reviewed studies. As part of the project, data collection may be considered also, along with developing novel machine learning algorithms (such as deep nets) for improved prediction/classification. This project is closely linked to a few pilot studies conducted by the supervisors [3][4]. This project will provide an excellent platform for the qualified PhD candidate to undertake a series of professional trainings, from experimental analysis to theoretical exploration of novel methods.

References:
[1] “ADNI | ACCESS DATA.” [Online]. Available: http://adni.loni.usc.edu/data-samples/access-data/. [Accessed: 28-Oct-2021].
[2] “Connectome - HCP Lifespan Studies.” [Online]. Available: https://www.humanconnectome.org/lifespan-studies. [Accessed: 28-Oct-2021].
[3] Su Yang, Jose Miguel Sanchez Bornot, Ricardo Bruña Fernandez, Farzin Deravi, Sanaul Hoque, KongFatt Wong-Lin and Girijesh Prasad, “Detection of Mild Cognitive Impairment with MEG Functional Connectivity Using Wavelet-Based Neuromarkers,” *Sensors*, vol. 21, no. 18, pp. 1–18, 2021.

[4] S. Yang, J. Miguel, S. Bornot, K. Wong-lin, G. Prasad, and S. Member, “M / EEG-based Bio-markers to predict the Mild Cognitive Impairment and Alzheimer ’ s disease : A Review from the Machine Learning Perspective,” *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2924–2935, 2019.

Project title: Multi-view data integration to predict nanomedicine efficacy; driving the future reality?

1st supervisor: Dr Lewis Francis, Associate Professor; Medicine
2nd supervisor: Professor Deya Gonzalez; Medicine
3rd supervisor: Professor Paul Rees; Engineering
Department/Institution: School of Medicine; Swansea University
Research theme: T2 – biological, health and clinical sciences
T3 – novel mathematical, physical and computer science approaches

Project description: After decades of biochemical, biotechnological and biomedical innovation a plethora of Nano scale drug delivery approaches exist which complement the spectrum of pharmaceutical compound libraries as well as future gene therapy approaches. Due to their versatility, heterogeneity and ability to adapt both drugs and intrinsic payloads to particular cohorts of patients, Nanomedicines are predicted to revolutionise future personalised medicine approaches. One example is their ability to synergise with pharmacological and physical co-treatments, integrating to multi modal combination therapy regimens.

In Swansea we have developed high throughput multi-view datasets, characterising both synthetic (polymerosomes) and natural (Extra cellular vesicle [EV]) nanoscale therapeutic approaches. Consisting of bio-chemical, mechanical and physiological variables these datasets have been further boosted by characteristic genomic, transcriptomic, proteomic, metabolomic, and lipidomic profiles for both cargo and target tissue mechanism of action studies.

This project will utilise a systems biology approach to map the biological and methodological causes of heterogeneity in these datasets, developing statistical methods for data integration. These approaches will require integrative machine learning models, to better use the vast volumes of heterogenous information in the deep understanding of these biological systems facilitating the development of predictive models. The major focus will be to develop omics and clinical data integration techniques for various essential analysis such as prediction, clustering, dimension reduction and association as a precursor to integrated multi modal learning to capture the complex mechanism of Nanomedicine systems. Indeed, integrated omics and multi-modal layers will provide more sensitive and comprehensive details on highly heterogeneous Nanomedicine populations (by means of size and biogenesis). Utilising bioinformatics tools for normalization and processing of multi-view data will facilitate more insights into the intricate cellular and molecular therapeutic modulation catalysed by these classes of future medicines.

The successful applicant will join the Reproductive Biology and Gynaecology Oncology research group in Swansea's Medical School in collaboration with Prof Paul Rees in Swansea's College of Engineering. 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: Visual Analytics for Public Health Network Analysis
1st supervisor: Daniel Archambault
2nd supervisor: Alma Rahat
Department/Institution: Computer Science Swansea University
Research theme: T2 – biological, health and clinical sciences
Project description: In public health settings, social networks are often encoded as multivariate graphs with both static and dynamic information. The human actors in these networks have demographic and survey information associated with them along with their social ties. Given this information, the analyst wants to understand how the information associated with the nodes and the social ties influence behaviours (under-age drinking, mental health, non-suicidal self-injury etc). In this project, how network analytics and ML can be supported by visual analysis in this setting.