**P1-BU-2020, Project title: Automated optimisation of industrial X-ray Computed Tomography**

**1st supervisor:** Dr Franck Vidal / School of Computer Science and Electronic Engineering

**2nd supervisor:** Dr Simon Middleburgh/ School of Computer Science and Electronic Engineering

This PhD will investigate how to automate the parametrisation of non-destructive testing (NDT) with Computerised Tomography (CT) for customised components. High Performance computing will be used to scan and tune multidimensional parameters, which is challenging using today’s algorithms. Predictive analytics and machine learning techniques will be deployed to accelerate the process by reducing the number of simulations required. The PhD will extend current state of the art, where we have already demonstrated that it is possible to model scanning parameters to simulate highly accurate CT data acquisition processes. An additional aim is to automatically design the geometry of the holder on which the scanned object will be placed during the CT scan, by combining high performance X-ray simulation on GPU with mathematical optimisation. This approach will generate a large amount of scanning parameters and corresponding simulated data. This multi-disciplinary PhD combines simulation, optimisation, machine learning and material science at Bangor University and Swansea University. The potential impact of this research is to reduce the time needed to perfom a successful NDT examination by porting some of the time-consuming manual experiments (finding the right parameters, and finding the right holder geometry) into a fully-automated computerised framework.

**P2-BU-2020, Project title: Learning from Badly Behaving Data**

**1st supervisor:** Prof Lucy Kuncheva / School of Computer Science and Electronic Engineering

**2nd supervisor:** Dr Franck Vidal / School of Computer Science and Electronic Engineering

Focusing on deep learning systems, this PhD will investigate the modern data challenge of data “behaving badly”. In addition to coming in massive volumes, data can be streaming, drifting, partially labelled, multi-labelled, contaminated, imbalanced, wide, and so on. A prime example of considerable interest is image and video analysis where the same object, person, or animal must be detected, learned, identified and then re-identified in the subsequent image collection or video stream. To solve this problem, we should look into semi-supervised learning in the presence of concept drift, adaptive learning, transductive learning, and more. Deep learning neural networks may prove valuable at the stages when large labelled data sets have been accumulated. Given that multiple objects of interest may be present within the same image, methods from the area of restricted set classification should be explored. This project will seek to offer novel and effective solutions for “badly behaving” data. Where possible, we will aspire to offer theoretical grounds for those solutions to ensure transferability across application domains. A curious potential application is identification of individual animals in a herd or a group for the purposes of non-invasive monitoring. Such an application will cross over to the area of environmental studies, specifically ecosystem conservation and behavioural ecology.

**P3-BU-2020, Project title: Developing Artificial Intelligence Techniques to Improve Hydrological Model Regionalisation**

**1st supervisor:** Dr Sopan Patil / School of Natural Sciences

**2nd supervisor:** Dr Panagiotis Ritsos / School of Computer Science and Electronic Engineering

The focus of this PhD is to develop AI techniques that can help improve hydrological model regionalisation. Specifically, the research will investigate novel use of AI and information visualization to interactively relate hydrological model parameters to the physical properties of river basins. Hydrological models are essential tools for simulating streamflow in river basins and are widely used for forecasting floods and droughts. However, appropriate application of hydrological models requires *a priori* calibration of parameters using historical measured streamflow data. To make matters worse, previous research has shown that hydrological model parameters are not strongly correlated to the physical properties of river basins (e.g., topography, soils, land use). This limits the ability to regionalise hydrological models, i.e. estimate model parameters at ungauged river basins or modify parameter values if land use changes in a river basin. Recent advances in Artificial Intelligence (AI), specifically in Deep Learning, 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. Our approach will involve development of Deep Learning techniques to extract high level abstractions in hydrological models and physical river basin data, which can be used to test the impact of land management decisions on river basin hydrology. This abstraction will be made available to relevant stakeholders via an interactive visualization interface to facilitate the investigation of multiple hydrological and land-use change scenarios using interpolation, classification and, where possible, generalization. Our training dataset will include data from >1000 river basins across the UK, and the coupled AI-hydrological modelling workflow will be streamlined to operate on HPC framework. This research will advance the field of AI through application of novel techniques for hydrological model regionalisation and help improve the assessment of land management decisions on flood/drought risk.

**P4-BU-2020, Project title:** Automated data cleaning, analysis and visualization from smartphone captured data for climate change

**1st supervisor:** Dr Simon Willcock / School of Natural Sciences

**2nd supervisor team:** Dr William Teahan, Prof Jonathan Roberts / School of Computer Science and Electronic Engineering

This research will investigate AI techniques to automate data cleaning, analysis and visualization of results from smarphone captured data science in the area of polulation growth, sustainability and climate change. It builds on collaborations with Natural Sciences and Computing, and require bespoke AI, Natural Language Processing (NLP) and visualisation solutions to be developed. Faced with population growth and climate change, sustainability has become one of the most important global challenges. To address this, we need high resolution spatiotemporal data on the state of natural systems and how we as a society are impacting them. Whilst artificial intelligence (AI) techniques are well established for the former (e.g. providing standardised hourly/daily/weekly data and analyses ranging from site-specific sensors up to remote satellites), most societal data are of extremely limited spatial and/or temporal resolution. To address this, we have pioneered the use of smartphone-based surveys in two countries with large sustainable development challenges: Bangladesh and Cambodia (<https://msds.tools/>). Our existing data demonstrates that real-time social data collection at large scales is now feasible and affordable. However, in order to up-scale these surveys, AI techniques to automate data cleaning, analysis and visualization of results are required, building on and enhancing bespoke AI, Natural Language Processing (NLP) and visualisation solutions developed at Bangor (see references below). Here, we will develop these techniques further. Machine learning (ML) algorithms automating data cleaning will be developed. Novel AI techniques will also be developed to analyse different types of social data, including quantitative (e.g. how much water was collected), qualitative (e.g. free text of how the water was used), spatial (e.g. locations of where the water was collected, or a GPS track of how to get there) and image (e.g. a photo of the water source) data. Such analyses will bring about an unprecedented level of survey data availability and so will require advanced visualization techniques to a) understand the results ourselves, but also to b) share these results back to the survey respondents via their smartphone. Finally, we will further advance survey methods by developing an AI approach to iteratively evolve dependent on the previous answers from that particular respondent – making future widespread, generalized surveys adapt based on earlier responses to become increasingly tailored to each individual over time.

**P5-BU-2020, Project title: Programme & Curriculum Analytics**

**1st supervisor:** Dr Cameron Gray / School of Computer Science & Electronic Engineering

**2nd supervisor:** Dr Dave Perkins / School of Computer Science & Electronic Engineering

This research will investigate appropriate methods of Machine Learning algorithms, tools and processes and apply them to examine how educational design influences both achievement and experience in modules and programmes. From there, models will be generated to predict the success and popularity of any proposed curriculum or programme changes. A necessary part of the project will also be examining how best to communicate any insight generated to academics to promote necessary change and best practice. There is future scope to meld this information with other forms of Learning Analytics (LA) using ensemble methods in order to more accurately predict student metrics. Indeed, Learning Analytics is already gaining ground within UK institutions, with their renewed focus on data, possibilities of using big data, and using it to drive decisions and influence policies. Research within the LA sphere is now shifting from prediction of student achievement and retention to other aspects of student engagement.