**Title:** An investigation into falsely classified samples in pattern recognition models

**Contact**: Gernot Liebchen <gliebchen@bournemouth.ac.uk>; Shahin Rostami <srostami@bournemouth.ac.uk>

**Project Description:**

* Numerous techniques exist for the optimisation and training of Artificial Neural Networks (ANNs) (e.g. the Widrow-Hoff least mean squares algorithm, back propagation, and neuro-evolution techniques), however, the accuracy of the prediction is heavily dependent on the training samples and associated labels themselves.
* The desired accuracy for an ANN often falls within certain margins, i.e. the Goldilocks principle, where it does not overfit or underfit the data samples. With this balance in accuracy, a high percentage samples are often misclassified.
* There is an opportunity for optimisation within the Goldilocks principle, such that marginal improvements can be made to the overall accuracy of the model. The importance of this optimisation can be demonstrated with the example of an ANN used for detecting brain tumours. Even a 1% increase in accuracy could mean that 10 in 1000 patients can benefit from the tumour being detected early.
* Additionally, data quality issues can lead to misleading conclusions. This is especially so if unsupervised algorithms are applied. Incorrect data items are often identified by outlier detection mechanisms. However, outliers may themselves be accurate instances with new unseen underlying characteristics. The problem is that the absence of data quality issues cannot be proven. Therefore, investigating characteristics of these data quality issues is important to identify yet unidentified data quality issues.
* The process and the lessons learnt from this project will be directly applicable to other data sets. This will feed into further investigations in the underlying characteristics of data quality issues.
* The outcome of this project will also offer preliminary results and insight into potentially fundamental issues which will lead onto the proposal of novel workflow for CI researchers.
* The recruited RA will be responsible of producing the above outcome. They will receive guidance from the two academics in charge of this project.

**Research Assistant:** an experienced **Graduate** **Research Assistant** with good analytical skills, a statistics background and knowledge of Matlab, Python or R. The project will run between Jan 2017 and March 2017.

**Hourly rate**: £12.12

**Number of hours**: 96 hours