The concurrent occurrence of multiple chronic diseases in a single patient is known as multimorbidity and it is the most common chronic condition affecting more than 50% of people over the age of 65 (Campbell-Scherer 2010). Patients with multimorbidity have complex treatment needs that are not met by the single index disease-oriented approach. Inappropriate single index disease-driven treatment in patients with multimorbidity can result in undesirable sequelae, including harmful interventions. Chronic inflammation is strongly implicated with multiple chronic diseases associated with multimorbidity and those associated with aging in particular. The elderly are the highest consumers of prescription medications and more than 50% of patients attending clinics have multiple morbidity.

The CDC (http://www.cdc.gov/arthritis/data_statistics/comorbidities.htm) reported that in patients with arthritis (primary index disease) 24% had cardiovascular disease, 19% respiratory conditions, 16% diabetes and 24% had depression. Thus there is an unmet need for better diagnostics and tailored therapies for patients with multimorbidity. There is a need to reclassify or stratify and treat complex multimorbidities better. Inflammatory diseases or auto-inflammatory diseases occur when the bodies immune system attacks its own tissues. Inflammation has been reported as a common occurrence in many patients exhibiting co-morbid or multimorbid disease. Examples include comorbidity such as cardiovascular disease risks in rheumatoid arthritis patients (Mariana JK 2010) and inflammation has also been associated with depression and Alzheimer’s Disease. Identification of biomarkers for stratifying patients with respect to inflammation and comorbidity or multimorbidity to better sub-group inflammatory diseases is the aim of this project.

Objectives of the research project:
Within the Northern Ireland Centre for Stratified Medicine we have recruited 2800 patients (total to be recruited 7000) with diseases that are related to inflammation and we are undertaking deep phenotyping and omics (transcriptomics, whole genome sequencing etc) analyses of samples from these patients. We aim to develop better tools to stratify patients recruited with a primary index disease but many also have co- or multimorbidity. So we are developing tools to better diagnose patients by examining the relationships between patients that have multiple co-morbidities associated with an inflammatory etiology. We aim to do this by identifying biomarkers that can sub-classify, ultimately better diagnose and also better inform the treatment regimens most suitable for such patients. This requires the identification of biomarkers or features associated with disease state and response to specific treatments in each patient. This requires accurate collection of the clinical and omics’ data associated with the patients and subjecting this data to advanced computational analyses. This project will involve using advanced computational methods to analyse data from the patients recruited at NICSM and public data.

Machine Learning is a sub group of Artificial Intelligence methods in Computer Science. And it has been widely used for prediction/classification problems in Bioinformatics (Tarca et al 2007, Baldi & Soren 2001, Savajardo et al 2011, Shukla 2010).

Objectives of this research project are:

1. To use high-throughput multi-omics data and clinical data available in the public data repositories (e.g. GEO, ArrayExpress) and those available from patients recruited at Northern Ireland Centre for Stratified Medicine under different inflammatory related disease work packages, to extract features which can be used to develop Machine Learning based models for classifying sub-groups of inflammatory diseases.
2. To perform high-throughput data analysis (Variant-Calling, Expression Profiling etc) on above mentioned datasets.
3. To identify/extract most relevant features from the above datasets which can robustly predict the sub-groups of patients with inflammatory disease(s).
4. To develop Machine Learning based models which can be applied to patient data for precision diagnosis of inflammatory diseases(s) and correlate outcomes with potential optimal treatments.

**Methods to be used:**

Data from each disease area (Cardiovascular, Rheumatoid Arthritis, Diabetes, Alzheimer/Mental Health/Depression patient cohorts) will be standardised and coded with agreed core criteria applied to each patient data set. Data will be converted from categorical variables to indicator variables and an analytic hierarchy process based meta-learning algorithm will be used to investigate and identify the most suitable supervised and/or unsupervised classification algorithm to derive a clinical decision output.

Quality control procedures will be applied to patient data sets, along with processes to determine data quality and identify missing data. Datasets will be appropriately cleaned, the missing data will be imputed as appropriate and translated into standard analysis-ready datasets.

Transformed patient and clinician features will be analysed using a diversity of computational tools (including supervised and unsupervised classification algorithms) to develop intelligent data analytic outputs that identify primary features/biomarkers that correlate with specific patient outcomes (disease state and response to treatment). As there is unlikely to be a single best model that is suitable for all disease cohort datasets, different classifier systems or combinations of classifiers (hybrid classifiers) will be investigated. We have already used such an approach for microbiome data associated with inflammatory bowel disease (public datasets) in another project. The analyses will include use of various Machine Learning algorithms such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), Hidden Markov Models (HMM), Random Forests (RF), Decision Trees and their ensemble to develop a computational model for predicting sub-groups (stratification) of inflammatory diseases. The ability of each model to generate informative diagnostic, prognostic and predictive features with putative utility for clinical decision making will be investigated.

**Skills required of applicant:**

Scripting/programming in Linux/Shell, Python and R.

Good statistical knowledge.

**References:**


**How to apply and scholarships:**

Please navigate through the weblink [https://www.ulster.ac.uk/research/phdresearch-degrees/annual-competition](https://www.ulster.ac.uk/research/phdresearch-degrees/annual-competition) and the various tabs on its left for further information concerning “how to apply”, “scholarship opportunities” (for home students and international students), application deadline, etc.