Appropriate health diagnostic tools for developing countries using mobile phones and inexpensive sensors

The aim of the initial 2 year Project was to demonstrate that it is possible to discriminate between human diseases using basic test samples, simple sensor technologies, coupled with rigorous data extraction methods. Breath and urine samples were to be analysed with simple sensor arrays to produce a 'fingerprint' of metabolites (small molecules that are the products of metabolism). Changes in these metabolites (and therefore changes in the fingerprint) are caused by changes in the patient's biochemistry in response to the disease.

Many of the health diagnostic tests used in developed nations are unavailable in low resource settings due to factors such as cost, training, power, refrigeration and so forth. There is, however, a pressing need for rapid, easy to use, and inexpensive diagnostics, particularly of infectious disease, in such settings. For example, the standard diagnostic test for pulmonary tuberculosis is visual inspection of a sputum sample under a microscope by a trained professional – a procedure developed more than 120 years ago! The chemometric approach adopted in this project, focussed on sampling urine or breath, has the potential to become a diagnostic product suitable to meet this demand and the aim of the Project was to demonstrate the initial proof of concept.

SIEF funding allowed the pursuit of five primary activities over the two years:

1. ***Sample collection*** - Establishment of strong working relationships with St Vincent's Hospital and the Royal Melbourne Hospital where three research nurses were deployed to recruit trial participants from the emergency departments. The Nossal Institute for Global Health (University of Melbourne) led the hospital liaison and managed the research nurses. Establishing the necessary approvals to undertake sample collection was a significant task involving organisational agreements, several ethics approvals, recruitment, workflow, patient selection criteria, databases, privacy policies as well as installation of special equipment. With this taken care of, the nurses identified patients presenting at the emergency departments willing to participate in the trial by providing breath and urine samples and answering a questionnaire. In excess of 8000 potential participants were screened from which slightly more than 250 met the trial criteria and samples were subsequently collected. Approximately half of the participants presented with an infection and the remaining half (the control group) were chosen for their absence of infection (e.g. coming to the ED with an injury or chest pain).
   Output: 250+ urine and breath samples and associated (de-identified) clinical data.
2. ***Breath analysis*** - Patient breath samples were concentrated and captured in special tubes which were then transported to CSIRO laboratories. Stored breath samples were later released from the tubes and analysed with three different models of eNose instruments – electronic devices for characterising odours. The samples were also analysed using gas chromatography mass spectrometer (GC-MS) - a large and expensive laboratory-based instrument able to identify the chemical constituents of the breath – which could validate the information from the eNose sensors.
   Output: chemical fingerprint data from breath samples – from eNose platforms and GC-MS
3. ***Urine analysis*** - Patient urine samples were sent to a different CSIRO laboratory where they were tested using a CSIRO patented chemiresistor sensor array able to measure liquid

samples. CSIRO's sensor array had been previously demonstrated to discriminate low-level contaminants in water samples for environmental monitoring applications and the SIEF funding allowed researchers to explore and adapt the technology for use with urine samples. Urine contains components that can interfere with the chemiresistor in undesirable ways and the majority of the project duration was devoted to developing a sensor and experimental configuration able to operate in the presence of such interferants.

Output: chemiresistor configuration adapted for urine measurements, chemical fingerprint data from urine samples using the CSIRO chemiresistor array.

4. ***Data analysis*** - CSIRO machine learning experts analysed the chemical fingerprint data and clinical information and tested the performance of various machine learning algorithms for correctly identifying whether a sample originated from a case (patient with an infection) or control (patient without infection). For each candidate algorithm in turn, the available data (cases and controls) were divided into "training" sets (the majority of the data) and the remainder of the data reserved in a test set. The training data is used to computationally optimise the algorithm for classifying case or control. To test the optimised algorithms, the reserved test data is presented to gauge whether it accurately classifies the data as case or control. There are many ways to divide the data into the training and test sets, and this set of results is used to rigorously calculate the statistical confidence for the algorithm's classifications ("cross-validation"). There were other experimental factors which further combinatorially multiplied the complexity, such as which and how many data features were selected from each sensor element. Intelligent algorithms were developed for data feature selection to avoid these computational complexity issues and at the same time to optimise classification results. The computational resources necessary become very large, and so this part of the project utilised a high-performance computing cluster.

   Output: The process successfully identified an algorithm and an information-theoretic method of data feature selection that greatly enhanced classification performance compared with the conventional form of analysis. These were encapsulated in the development of a feature selection and classification toolkit. Data from breath analysis was demonstrated to correctly classify cases and controls above a statistically significant level.
5. ***Smart Cable*** - The Nossal Institute for Global Health developed the Smart Cable, producing a fully functional prototype within the 2-year project. The Smart Cable is able to integrate a mobile handset (focussing on the Nokia Series 40 Feature Phone – there being 800 million in use globally – mainly in the regions of interest) with a variety of "pluggable" devices, which would include our biosensors, but also physiological measurements such as temperature, breathing rate, heart rate, weight etc. The prototype demonstrated the ability to capture sensor data, pre-process as necessary and relay to the mobile handset where it might be used in a medical decision support process and ultimately treatment selection and dosing. The team also developed a prototype of an accompanying software development kit (SDK) which allows handset applications to be generated in one programming format and then be compiled to run on a range of mobile phone makes. An offshoot from the Smart Cable team won the 2012 Microsoft Australia Imagination Cup with their StethoCloud device.

Summary

In the two years, the project was able to establish the team, approvals and processes, collect more than the anticipated number of real patient samples of breath and urine, and analyse these with chemiresistor (some samples) and eNose (all samples) sensors. The algorithm development activity successfully identified a technique for selecting data features and analysing these which provided a superior diagnostic classification than a conventional analysis not involving intelligent feature

selection. A prototype battery powered, mobile phone based integrated Smart Cable was demonstrated. Finally, the project team demonstrated discrimination between infected and un-infected patients using a chemometric fingerprint at a statistically significant level. These results give the team great encouragement that the approach can be refined and developed towards a viable point of care diagnostic which offers the attributes necessary for widespread deployment in low resource settings: robustness, simplicity of use, integration into locally relevant medical decision support protocols, and, most importantly, affordability.