**PREDICTING HIE: CONVENTIONAL RISK FACTORS COMPARED TO AGNOSTIC MACHINE LEARNING ALGORITHMS**

**INTRODUCTION**

While mothers report that the wellbeing of their unborn infant is the single biggest priority for them1 there is little evidence to guide them or the professionals looking after their baby2. The prediction of which infants will become compromised around birth is poorly understood3, and has been identified as a priority for the RCOG4 and the UK Department of Health5. We have presented some work that shows that modelling of risk is feasible6 and we know that simple interventions can improve neonatal and maternal7,8 outcomes.

One significant cause of perinatal brain injury is perinatal asphyxia, leading to hypoxic-ischaemic encephalopathy (HIE). HIE is often devastating, with life-long impacts for the infant13 and their family, as well as costing society millions of pounds in medical compensation, lost earnings and welfare support14. In addition to the direct impact on infants and families, obstetric practice represents the biggest proportion of legal claims against the NHS15 and even small improvements in outcomes would yield substantial health benefits for individuals and economic benefits for healthcare services. Indeed, perinatal asphyxia is the 12th biggest cause of disability life years worldwide16 (i.e. a bigger impact than diabetes mellitus or tuberculosis), and even those infants with mild asphyxia have worse measures in cognition, movement and social metrics when compared to their peers17–19 and the true impact of this and other post-term related pathologies, and the economic implications, are unclear20. However interventions, such as induction of labour or operative delivery, can be employed if the risks of continuing the pregnancy are higher than delivery: for either the mother or the infant21. This lack of clear data on the perinatal risks and long term outcomes of these infants likely contributes to the variation in management of mothers with post-term babies11 and current NICE guidelines recognise this, and suggest that a research priority is to “identify babies at particularly high risk of morbidity and mortality who will benefit from induction and therefore avoid induction for babies who do not need it”22.

Risk factors for perinatal asphyxia and encephalopathy have been derived by a number of papers; although one of the most cited remains the work by Badawi *et al*10,26. This work identifies 35 potential risk factors for encephalopathy in an Australian population (Table 1) and together have been cited over 700 times (Data extracted Web of Science 17/12/2019). Analysing such data with machine learning (ML) algorithms to predict health outcomes is currently of great interest. However often algorithms require cleaned data, careful feature selection/engineering and significant training and expertise to develop, before they are able to meet, or exceed, clinical prediction. Recently more automated approaches to ML have become available, collectively known as automated machine learning (AutoML). We investigate the potential of such approaches to simplify the development of a clinical prediction model for HIE.

This work is based on the Collaborative Perinatal Project (CPP) 24. Collection of data was from 14 units across the United States and showed little evidence of selection bias25. The dataset includes data on approximately 60,000 pregnancies, and 58,000 live born infants born between 1959 and 1965. Data was collected throughout the prenatal period, labour and delivery, postpartum and as the child grew.