**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.

**AIMS**

1. To evaluate if automated feature selection, feature engineering and ML approaches applied to a large dataset with minimal human input can produce models with that predict disease outcomes as well, or better than conventional approaches with established risk factors.
2. To test if measures of infant growth can improve the prediction of HIE.
3. Determine and compare the performance of a range of binary classifiers in predicting HIE.

**METHODS**

**Outcome**

Hypoxic-ischaemic encephalopathy (HIE) was defined as having definite seizures, hypertonia, jitteriness, hypotonia, abnormal reflexes, or abnormal cry; after having a low 5 minute Apgar score (<7)27.

**Data preparation**Pregnancies were ordered chronologically and split into two equal subsets for training (infants born 1959-1962) and testing purposes (infants born 1963 to 1965). All variables were identified as either antenatal (measurable before 37 weeks’ gestation), growth (birth measures of growth), and intrapartum (measures only available at or after 37 weeks, up to the point to delivery), and classified as either unordered categorical, ordinal or numerical.

**Feature selection**

Two broad feature selection approaches were used. The first took established risk factors for HIE, based on *a-priori* proposed risk factors (Table 1)10,26(‘Conventional analysis’). The second approach used automated data driven methods to rank features by their predictability of HIE (‘AutoML analysis’).

**Model assessment**

Models were trained using the training data and each of the feature sets and applied to predict HIE status in the testing cohort. Receiver-operator curves were prepared to compare model discrimination and derive the area under the curve measure with 95% confidence interval using the pROC R-package (v1.16.2). Comparisons between features collected during the three timepoints (antenatal, antenatal and growth, antenatal and intrapartum) were performed to test if the addition of growth or intrapartum measures improved model prediction.

**Conventional analysis**

Training data were cleaned and harmonised where possible with the measures previously proposed 10,26. A logistic regression model was developed using the training data and evaluated using the testing data. The primary model included only antenatal variables (Table 1), but this was extended to additionally include foetal growth or intrapartum variables. This analysis was performed using Stata v16.

**Automated feature selection and feature engineering**

Agnostic feature selection approaches were applied to the three sets of predictors (antenatal, antenatal and growth, antenatal and intrapartum) and ranked using the training data. First, variables were discarded if they contained >5% of missing data values (28 out of 518) leaving a potential 490 exposure data fields for the prediction models. Second, unordered categorical variables were recoded as dummy variables. Third, the training data was used for feature selection using a range of algorithms from the scikit-learn Python package (v0.23.2) as follows with default parameters except where specified. These methods were chosen to cover a range of commonly used methods (tree, regularisation and recursive elimination) as outlined in the scikit-learn documentation.

*Reverse feature elimination (RFE)*

During the first iteration all input variables were included as predictors in a logistic regression model trained using five-fold cross validation. On each subsequent iteration the five weakest predictors (determined by the smallest absolute coefficient) were eliminated, with iterations continuing until only one predictor remained. The effect of each predictor on the cross-validation mean AUC was used to rank features (lowest rank is most predictive).

*Elastic-Net and Lasso regression*

The logistic regression with L1 and L2 regularisation penalties was trained using five-fold cross-validation to determine the optimal value of alpha. The Elastic-Net mixing parameter (l1\_ratio, representing the ratio of L1 to L2 penalty) was set to 0.5. The penalty term shrinks weak predictors to zero which were subsequently eliminated from downstream analyses. The remaining features were ranked by their absolute regression coefficient (smallest is least predictive).

*Extra trees classifier*

The feature importance metric obtained from an extra-trees classifier with five-fold cross-validation was used to rank features (smallest is least predictive).

*Linear support vector classifier (SVC)*

The linear SVC was trained with five-fold cross-validation using the AUC metric. An L1 penalty term was applied to shrink weak coefficients to zero. The coefficients were taken as a measure of feature importance (absolute value; smallest least predictive).

**Binary classification**The following models were trained and evaluated using default hyperparameters and the top n=20, n=40 or n=60 predictors of each feature selection method: logistic regression, random forest, naïve Bayes and neural network (one hidden layer with number of nodes equal to number of predictors using the rectified linear activation function and Adam optimiser implemented in Tensorflow v1.15).

**RESULTS**

**DISCUSSION**

Using freely available historical data, we have shown that agnostic feature selection, feature engineering and ML algorithms are able to predict the risk of HIE and perinatal death to a similar level of precision as expert-developed clinical prediction models despite minimum data preparation of the cohort. In addition, measure of prediction did improve in many of the models after addition of growth measures (did they?) but not in all. While the ML models were able to match the conventional analysis, it should be noted that the risk factors used in this (and the categories etc used) were derived from data measures in a different population some years late.

Internal limitations of this work include missing data (selection bias); although only 390 infants were not included in any of the conventional or ML models. While the data was derived from a large cohort study considered to be of high methodological quality, it is clearly limited by the age of the data; although the risk factors proposed in Badawi’s work (derived from an Australian population) some decades later were still strong predictors of outcome; suggesting that the main underlying causes of perinatal death and HIE remain significant over this time frame. In addition, for this to provide a valid use to clinicians it need to be tested on recent, but also routinely collected data; with likely less cleaning and more missing data points than is present in this research dataset.

MATT CAN YOU ADD YOUR ‘AREAS’ HERE?

This work is consistent with other prediction work, suggesting that poor birth condition can be predicted, although in this study we have attempted to test if ‘raw’ data, mostly unprepared by the research team, could be used by the ML model to aid clinicians. The advantages of this would mean application across the multiple data standards and systems in place within healthcare services, and the development of adaptive risk scores to aid decisions around and before birth. Women in the highest decile of risk had risks of between 20% and 40% of their infant developing brain injury around birth (compared to the background risk of only 0.5%); with this risk predictable from antenatal measures. If replicable using live data, this would give the mother and clinicians valuable data to help guide management as we know interventions exists to reduce or modify the risk. We also wanted to test if growth measure, at the time measures after the birth of the infant but now measurable in-utero using ultrasound would add prediction to the model. In the conventional model growth did add additional value for HIE (p=0.0263) and perinatal death (p=0.005), but ?NOT? in the ML models. This may reflect other measures of growth or correlated of it (e.g. number of outpatient appointments) stored in the antenatal record and further work using antenatal measures should interrogate this.

**Conclusion**

In this work, on a historical cohort, machine learning models with minimum data preparation were able to match and, in some examples, exceed the prediction of conventional analysis in predicting which infants would develop HIE after birth. Some predictions improved when measures of growth were included, supporting the role for routine antenatal measures of growth during pregnancies using modern imaging techniques. Routine growth measures, and automated ML models on other routinely collected health data may provide an additional tool to obstetric services to help identify infants at high risk of brain injury around birth and help target additional observation or interventions.

**Data and open-source code availability**

Stata, R and Python3 code are available from <https://ieugit-scmv-d0.epi.bris.ac.uk/ml18692/hie-ml>.

The CPP data files and documentation are available for download from the National Archives Catalog (https://www.archives.gov/research/electronic-records/nih.html)

**Acknowledgments**This study was funded by the NIHR Biomedical Research Centre at University Hospitals Bristol and Weston NHS Foundation Trust and the University of Bristol. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. This work was also funded by the UK Medical Research Council as part of the MRC Integrative Epidemiology Unit (MC\_UU\_00011/4 and MC\_UU\_00011/6).

**Competing interest**

TRG receives funding from GlaxoSmithKline and Biogen for unrelated research.

**REFERENCES**

1. Kingdon, C. *et al.* Choice and birth method: mixed-method study of caesarean delivery for maternal request. *BJOG* **116**, 886–895 (2009).

2. Berger, B., Schwarz, C. & Heusser, P. Watchful waiting or induction of labour--a matter of informed choice: identification, analysis and critical appraisal of decision aids and patient information regarding care options for women with uncomplicated singleton late and post term pregnancies: a . *BMC Complement. Altern. Med.* **15**, 143 (2015).

3. Odd, D. E. *et al.* Risk of low Apgar score and socioeconomic position: a study of Swedish male births. *Acta Paediatr.* **97**, 1275–1280 (2008).

4. Gynaecologists, R. C. of O. and. Easy Baby Counts. https://www.rcog.org.uk/eachbabycounts.

5. Health, D. of. New ambition to halve rate of stillbirths and infant deaths. https://www.gov.uk/government/news/new-ambition-to-halve-rate-of-stillbirths-and-infant-deaths.

6. Odd, D., Heep, A., Luyt, K. & Draycott, T. Hypoxic-Ischaemic Brain Injury: Delivery Before Intrapartum Events. in *Joined European Neonatal Societies Congress* (2015).

7. Draycott, T. *et al.* Does training in obstetric emergencies improve neonatal outcome? *Bjog* **113**, 177–182 (2006).

8. Chiossim, G. Timing of Delivery and Adverse Outcomes in Term Singleton Repeat Cesarean Deliveries. *Obs Gynecol* **121**, (2013).

9. Martinez-Biarge, M., Madero, R., González, A., Quero, A. & García-Alix, A. Perinatal morbidity and risk of hypoxic-ischemic encephalopathy associated with intrapartum sentinel events. *Am J Obs. Gynecol2* **206**, 148.e1–7 (2012).

10. Badawi, N. *et al.* Intrapartum risk factors for newborn encephalopathy: the Western Australian case-control study. *Bmj* **317**, 1554–1558 (1998).

11. Gülmezoglu, A., Crowther, C., Middleton, P. & Peatley, E. Induction of labour for improving birth outcomes for women at or beyond term. *Cochrane Database Syst Rev.* (2012) doi:10.1002/14651858.CD004945.pub3.

12. Campbell, M. K., Ostbye, T. & Irgens, L. M. Post-term birth: risk factors and outcomes in a 10-year cohort of Norwegian births. *Obstet. Gynecol.* **89**, 543–548 (1997).

13. Azzopardi, D. V *et al.* Moderate hypothermia to treat perinatal asphyxial encephalopathy. *N Engl J Med.* **361**, 1349–1358 (2009).

14. Odd, D. E., Gunnell, D., Lewis, G. & Rasmussen, F. Long-term Impact of Poor Birth Condition on Social and Economic Outcomes in Early Adulthood. *Pediatrics* **May 9**; **eFi**, e1498-504 (2011).

15. *Ten Years of Maternity Claims: An Analysis of NHS Litigation Authority Data*. (2012) doi:978-0-9565019-2-9.

16. Murray, C. J. L. *et al.* Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet (London, England)* **380**, 2197–2223 (2012).

17. Odd, D. E., Lewis, G., Whitelaw, A. & Gunnell, D. Resuscitation at birth and cognition at 8 years of age: a cohort study. *Lancet* **373**, 1615–22 (2009).

18. Odd, D. E., Gunnell, D., Whitelaw, A. & Lewis, G. The association between birth condition and neuropsychological functioning and educational attainment at school age. A cohort study. *Arch Dis Child* (2010).

19. Odd, D. E., Rasmussen, F., Gunnell, D. J., Lewis, G. & Whitelaw, A. A Cohort Study of Low Apgar Scores and Cognitive Outcomes. *Arch Dis Child Fetal Neonatal Ed* **93**, F115-20 (2008).

20. Goeree, R., Hannah, M. & Hewson, S. Cost-effectiveness of induction of labour versus serial antenatal monitoring in the Canadian Multicentre Postterm Pregnancy Trial. *Can Med Assoc J* **9.**, 1445–50. (1995).

21. Molina, G. *et al.* Relationship Between Cesarean Delivery Rate and Maternal and Neonatal Mortality. *JAMA* **314**, 2263–2270 (2015).

22. Induction of labour. *NICE Clin. Guidel. (July 2008)*.

23. National Institute for Health and Clinical Excellence (NICE). Inducing labour (CG70). (2008).

24. Hardy, J. B. The Johns Hopkins Collaborative Perinatal Project. Descriptive background. *Johns Hopkins Med. J.* **128**, 238–243 (1971).

25. Niswander, K. & Gordon, M. T. *The Women and Their Pregnancies: The Collaborative Perinatal Study of the NINDS*. (USGov. Printing Press, 1972).

26. Badawi, N. *et al.* Antepartum risk factors for newborn encephalopathy: the Western Australian case-control study. *Bmj* **317**, 1549–1553 (1998).

27. Odd, D., Lewis, G., Whitelaw, A. & Gunnell, D. J. Resuscitation at birth and cognition at 8 years of age: a cohort study. *Lancet* **9**, 1615–1622 (2009).

**Table 1. Established risk factors**

|  |  |  |
| --- | --- | --- |
| Antenatal Factors | Growth Measures | Intrapartum Factors |
| * Maternal age (<20, 20-24, 25-29, 30-34, >35) * Parity 0, 1,>1 * Maternal Employment * Health Insurance * Maternal race * FHx of seizures (recurrent non-febrile seizures) * FHx of neurological disorder (excludes seizures) * Infertility Treatment * Maternal Hypertension * Maternal height (<160, 160-164, >164) * Maternal Thyroid Disease * Pre-eclampsia * Antenatal bleeding (mod or severe) * Viral Illness * Alcohol (some, none, unknown) * Birthweight centile (>90th, 10-90th, 3rd-9th, <3rd) * Sex * Abnormal placenta * Late or no antenatal care * Multiple births | * Birth weight centile (>90th, 10-90th, 3rd-9th, <3rd) | * Gestation (37-42) * OP presentation * Maternal Pyrexia * Maternal Intrapartum Event (Haemorrhage, convulsions, uterine rupture, snapped cord, out of hospital birth) * Membrane rupture >12 hours * Blood Pressure abnormalities – Captured above * Nuchal cord * Cord prolapse * Onset of labour (spontaneous, induced, none) * Mode of delivery (Spontaneous, induced vaginal, elective CS, emergency CS, breech manoeuvre) * Shoulder dystocia * Epidural Anaesthetic * Breech Presentation * ROM>12 hours |

**Table 3. Traditional vs ML prediction of pregnancy outcomes**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Outcome | Antenatal Factors |  | Antenatal and Growth Factors |  |  | Antenatal and Intrapartum Factors |  |  |
|  | AUC (95% CI) | Proportion in highest decile | AUC (95% CI) | Proportion in highest decile | p-value\* | AUC (95% CI) | Proportion in highest decile | p-value\* |
| Hypoxic-Ischaemic Encephalopathy | | | | | | | | |
| Conventional Analysis | 0.71 (0.64-0.77) | 17 (28.8%) | 0.73 (0.67-0.79) | 22 (37.3%) | 0.09 | 0.7 (0.64-0.77) | 18 (30.5%) | 0.68 |
| ML (L-Regression) | 0.71 (0.65-0.77) | 18 (26.5%) | 0.73 (0.66-0.79) | 16 (30.2%) | 0.3 | 0.74 (0.67-0.8) | 17 (30.9%) | 0.16 |
| p-value\*\* | 0.85 |  | 0.9 |  |  | 0.24 |  |  |
| ML (Random Forest) | 0.63 (0.57-0.7) | 16 (23.5%) | 0.59 (0.52-0.65) | 11 (20.8%) | 0.61 | 0.65 (0.58-0.72) | 16 (29.1%) | 0.18 |
| p-value\*\* | 0.01 |  | 0 |  |  | 0.15 |  |  |
| ML (Neural Net) | 0.62 (0.55-0.68) | 12 (17.6%) | 0.62 (0.54-0.7) | 11 (20.8%) | 0.9 | 0.61 (0.52-0.69) | 14 (25.5%) | 0.81 |
| p-value\*\* | 0.01 |  | 0 |  |  | 0.03 |  |  |
| ML Model (Adanet) | 0.69 (0.63-0.76) | 21 (30.9%) | 0.71 (0.64-0.79) | 22 (41.5%) | 0.18 | 0.69 (0.61-0.77) | 17 (30.9%) | 0.91 |
| p-value\*\* | 0.63 |  | 0.83 |  |  | 0.78 |  |  |
| Perinatal Death | | | | | | | | |
| Conventional Analysis | 0.61 (0.57-0.65) | 30 (18.2%) | 0.66 (0.61-0.71) | 57 (35.2%) | 0 | NA | NA | NA |
| ML (L-Regression) | 0.75 (0.72-0.79) | 64 (35%) | 0.56 (0.46-0.66) | 10 (23.3%) | 0.3 | NA | NA | NA |
| p-value\*\* | 0 |  | 0.29 |  |  | NA |  |  |
| ML (Random Forest) | 0.65 (0.61-0.68) | 52 (28.4%) | 0.61 (0.52-0.69) | 13 (30.2%) | 0.32 | NA | NA | NA |
| p-value\*\* | 0.41 |  | 0.02 |  |  | NA |  |  |
| ML (Neural Net) | 0.66 (0.62-0.71) | 50 (27.3%) | 0.53 (0.46-0.6) | 6 (14%) | 0.03 | NA | NA | NA |
| p-value\*\* | 0.03 |  | 0.39 |  |  | NA |  |  |
| ML Model (Adanet) | 0.71 (0.67-0.75) | 60 (32.8%) | 0.57 (0.49-0.65) | 5 (11.6%) | 0.79 | NA | NA | NA |
| p-value\*\* | 0.01 |  | 0.05 |  |  | NA |  |  |

**\* Compared to Antenatal Factors Model**

**\*\* Compared with conventional mod**