



OPEN ACCESS

Validation of a machine learning algorithm for identifying infants at risk of hypoxic ischaemic encephalopathy in a large unseen data set

Anne L Murray ^{1,2}, Daragh S O'Boyle,² Brian H Walsh,^{1,2,3} Deirdre M Murray ^{2,3}

► Additional supplemental material is published online only. To view, please visit the journal online (<https://doi.org/10.1136/archdischild-2024-327366>).

¹Cork University Maternity Hospital, Wilton, Cork, Ireland
²INFANT Centre, Paediatric Academic Unit, Cork University Hospital, Wilton, Cork, Ireland
³Department of Paediatrics and Child Health, University College Cork, Cork, Ireland

Correspondence to
Dr Deirdre M Murray;
d.murray@ucc.ie

Received 6 May 2024
Accepted 21 August 2024

ABSTRACT

Objective To validate a hypoxic ischaemic encephalopathy (HIE) prediction algorithm to identify infants at risk of HIE immediately after birth using readily available clinical data.

Design Secondary review of electronic health record data of term deliveries from January 2017 to December 2021.

Setting A tertiary maternity hospital.

Patients Infants >36 weeks' gestation with the following clinical variables available: Apgar Score at 1 min and 5 min, postnatal pH, base deficit, and lactate values taken within 1 hour of birth

Interventions Previously trained open-source logistic regression and random forest (RF) prediction algorithms were used to calculate a probability index (PI) for each infant for the occurrence of HIE.

Main outcome Validation of a machine learning algorithm to identify infants at risk of HIE in the immediate postnatal period.

Results 1081 had a complete data set available within 1 hour of birth: 76 (6.95%) with HIE and 1005 non-HIE. Of the 76 infants with HIE, 37 were classified as mild, 29 moderate and 10 severe. The best overall accuracy was seen with the RF model. Median (IQR) PI in the HIE group was 0.70 (0.53–0.86) vs 0.05 (0.02–0.15), ($p<0.001$) in the non-HIE group. The area under the receiver operating characteristics curve for prediction of HIE=0.926 (0.893–0.959, $p<0.001$). Using a PI cut-off to optimise sensitivity of 0.30, 936 of the 1081 (86.5%) infants were correctly classified.

Conclusion In a large unseen data set an open-source algorithm could identify infants at risk of HIE in the immediate postnatal period. This may aid focused clinical examination, transfer to tertiary care (if necessary) and timely intervention.

BACKGROUND

Hypoxic ischaemic encephalopathy (HIE) is one of the leading causes of morbidity and mortality in infants worldwide.¹ It has an incidence of 1.5 cases per 1000 in high-income countries and 10–20 cases per 1000 in low-income and middle-income countries.^{1–3} Globally, 23% of infant mortality can be attributed to HIE and those who survive are at risk of adverse neurological outcomes, such as cerebral palsy and epilepsy, as well as behavioural and intellectual disabilities.^{4–6}

Therapeutic hypothermia (TH) is currently the only treatment available to mitigate the damage caused by HIE, and it has been repeatedly confirmed

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Early diagnosis of hypoxic ischaemic encephalopathy is essential for timely initiation of therapeutic hypothermia.

WHAT THIS STUDY ADDS

⇒ This study validates a previously developed machine learning algorithm using real-world data.
⇒ This algorithm may aid early diagnosis and can be used to quickly identify those who should have a detailed neurological or neurophysiological monitoring to detect and grade encephalopathy.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This algorithm may be useful to aid decision-making for timely transfer of infants to cooling centres for assessment and possible treatment.

to reduce cerebral injury and adverse neurological outcomes in cases of moderate and severe encephalopathy.^{17,8} Evidence suggests that TH is most beneficial when initiated within 6 hours of birth. Earlier initiation before 3 hours of life has been suggested to further enhance neurodevelopmental outcomes in affected infants.^{9–11} Therefore, the timely identification of HIE and accurate classification of its severity is of vital importance.

Diagnosis of HIE is difficult, as is the choice to initiate TH. Decision-making is often based on the neurological examination of the infant, which is known to change and evolve over time.^{12–14} Furthermore, the examination used may influence eligibility for TH, and variation between even experienced clinicians may impact the interpretation of the exam findings.^{15,16} According to the American College of Obstetricians and Gynecologists consensus statement on neonatal encephalopathy (NE), knowledge gaps still preclude a definitive test or set of markers that accurately identify an infant in whom NE is due to an acute intrapartum event.¹⁷

Machine learning (ML) can be defined as a branch of artificial intelligence in which computer software learns to perform a task by being exposed to representative data.¹⁸ Using training data, the ML algorithm is able to create a set of rules which can be used as predictors for new data with similar characteristics. These algorithms should then be tested with independent data not used in the development



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Murray AL, O'Boyle DS, Walsh BH, et al. *Arch Dis Child Fetal Neonatal Ed* Epub ahead of print: [please include Day Month Year]. doi:10.1136/archdischild-2024-327366

of the model.¹⁹ For our model, the training and testing phases have been previously published.^{20–21} The use of ML in clinical decision-making has the potential to positively affect the lives of patients and enhance our understanding of human disease.²²

Logistic regression (LR) and random forest (RF) are two types of ML algorithms. LR is commonly used when examining binary outcome and can be used to study the effect of multiple independent variables on an outcome of interest.^{23–25} RF is a form of ML that is becoming a common artificial intelligence (AI) tool in the medical field.^{24–26} Using multiple models when examining data sets is commonplace; if one algorithm outperforms the other in one metric, it may lose in another metric.²⁷

The aim of this study was to validate the previously reported ML algorithms to identify infants at risk of HIE immediately after birth using real-world, readily available clinical data.

METHODS

This was a retrospective review of electronic health record data of all term deliveries in Cork University Maternity Hospital during a 5-year period from 1 January 2017 to 31 December 2021. This hospital is a tertiary unit with approximately 7500 deliveries annually. Eligible infants were identified using the hospital's electronic health record. Inclusion criteria were all infants ≥ 36 weeks who had a blood gas drawn within 1 hour of birth. Infants were excluded if they were < 36 weeks gestation or were missing key data points for the decision support algorithm; Apgar Score at 1 min or 5 min, postnatal pH, base deficit (BD), or lactate. Infants < 36 weeks gestation were not included given current consensus guidelines on the use of TH extend only to infants ≥ 36 weeks gestation.¹⁷

Health records of all infants were examined for a diagnosis of HIE. Initial diagnosis was assigned following a contemporaneous prospective clinical exam based on a modified Sarnat Score.^{13–28} This exam consists of six domains, each assigned a stage of 1–3. Severe NE was defined as having ≥ 3 moderate/severe domains with more severe than moderate domains; moderate NE was defined as having ≥ 3 moderate/severe domains with more moderate than severe domains; and mild NE was defined as having one or more abnormal domains, but not meeting definition for moderate or severe NE. This was carried out by a trained registrar (resident or fellow) or consultant and the examination template did not change during the study period. All staff are provided training in this modified Sarnat Score by an experienced consultant when starting their rotation and it is local policy to assess for NE severity after 1 hour of age. Annually all encephalopathy cases are reviewed by two experienced senior clinicians, who were blinded to the outcomes of the infants, to ensure that fidelity of grading is maintained.²⁹ All infants who did not have a diagnosis of HIE were assigned to the non-HIE group and their diagnoses were also noted.

Previously trained open-source LR and RF prediction algorithms (<https://www.infantcentre.ie/predictionapp.html>) were used to calculate a Probability Index (PI) for each infant for the occurrence of HIE. These models were trained and assessed on separate data sets using the BiHIVE2 cohort of infants with signs of perinatal asphyxia with and without HIE matched against healthy controls.^{20–21} During development, multiple variables were examined to find those most predictive of HIE. The algorithms require the five variables listed above (Apgar score at 1 min or 5 min, postnatal pH, BD and lactate) to be inputted to calculate the PI for HIE. Each model provided a predicted probability for the development of HIE from 0 to 1, with higher values indicating a higher risk of HIE. The development data sets

and this validation data set are from the same/similar healthcare settings, with consistent criteria for the clinical diagnosis of NE.

Statistical analysis

Categorical data were described using frequencies and percentages. Continuous data were described using medians and IQRs. Differences between groups were evaluated with the Mann-Whitney U test for continuous non-parametrical data, or the χ^2 test for categorical data. The area under the receiver operating characteristics (AUROC) curve was used to investigate the predictive performance of the models. SPSS V.28 was used and a value of $p < 0.05$ was considered statistically significant. Calibration curves (online supplemental figure S1) were created using R studio V.24.04.02 following the Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)+AI guidelines (research checklist).

RESULTS

Over the 5-year study period, 1191 infants ≥ 36 weeks gestation had a blood gas measurement in the first hour of life. Of these, 1081 had a complete data set available. In total, 110 infants were excluded, all due to missing data. The groups differed by birth weight ($p = 0.04$), mode of delivery, with a higher proportion of prelabour caesarean sections recorded in the non-HIE group ($p = 0.005$), and by type of blood gas, with more arterial gases in the HIE population ($p < 0.001$) (table 1).

Of the 1081 infants, 76 (7%) had a diagnosis of HIE and 1005 (93%) did not have HIE. Of the 76 infants with HIE, 37 were clinically classified as mild, 29 as moderate and 10 as severe. 46 (60.5%) of the HIE group received TH (7 (19%) mild, 29 (100%) moderate and 10 (100%) severe) (table 2).

The primary diagnoses for the non-HIE infants were respiratory presentations, such as transient tachypnoea of the newborn and respiratory distress syndrome (44%), perinatal asphyxia without encephalopathy (23%), and suspected infection (16.5%). Twenty-six infants' diagnoses were classed as miscellaneous, and these included infants with rashes and neonatal admissions due to maternal health (table 1). The excluded infants had a similar range of diagnoses (online supplemental table S1).

For the RF model, the AUROC (CI) for prediction of HIE was 0.926 (0.893–0.959, $p < 0.001$) (figure 1). The median (IQR) PI in the HIE group was 0.70 (0.53–0.86) vs 0.05 (0.02–0.15), ($p < 0.001$) in the non-HIE group (figure 2). With the RF model, using a PI cut-off of 0.30 to optimise sensitivity, 936 of the 1081 (86.5%) infants were correctly classified by the algorithm (sensitivity=86.8%, specificity 86.6%). Of the infants with HIE, 66/76 (86.8%) were correctly identified. In addition, 135 of the 1005 were incorrectly classified as HIE (13.4%) (online supplemental table S2).

Using the LR model, the PI was calculated in the HIE versus non-HIE cohorts. The AUROC (CI) was 0.928 (0.892–0.964, $p < 0.001$) (figure 1). The median (IQR) PI for HIE was 0.71 (0.49–0.93) vs 0.05 (0.03–0.12), ($p < 0.001$) in the non-HIE group. Using a PI cut-off of 0.3, 955 of the 1081 (88.3%) infants were correctly classified by the algorithm (sensitivity of 84.2% and specificity of 88.7%). Of the infants with HIE, 64/76 (84.2%) were correctly identified. In addition, 114 of the 1005 were incorrectly classified as HIE (11.3%).

The RF model, using the cut-off threshold of 0.30, identified 78.4% of mild HIE, 93% of moderate HIE and 100% of severe HIE cases (figure 3). Of the eight cases of mild HIE that were incorrectly classified as non-HIE, none had seizures and none received TH. In the two cases of moderate HIE that

Table 1 Demographic details of the study population

Birth characteristics	All infants	HIE Median (IQR)	Non-HIE Median (IQR)	P value
Gestation	39.0 (37.7–40.3)	39.5 (38.2–40.6)	39.0 (37.7–40.3)	0.12
Birth weight (grams)	3420 (3020–3800)	3290 (2915–3650)	3440 (3030–3800)	0.04
Sex		N (%)	N (%)	
Male	612 (56)	40 (52)	572 (57)	0.47
Female	469 (43)	36 (47)	433 (43)	
Mode of delivery	n=1081(%)	n=76 (%)	n=1005 (%)	
Spontaneous vaginal	213 (19.7)	18 (23.7)	195 (19.4)	0.005
Operative vaginal	270 (24.9)	29 (38.15)	241 (23.9)	
Caesarean section	598 (55.2)	29 (38.15)	569 (56.5)	
Prelabour c-section	348 (32.1)	5 (6.6)	343 (34.0)	
Type of blood gas	n=1081 (%)	n=76 (%)	n=1005 (%)	
Arterial	20 (1.9)	13 (17.1)	7 (0.7)	<0.001
Venous	926 (85.6)	55 (72.4)	871 (86.6)	
Capillary	135 (12.4)	8 (10.5)	127 (12.6)	
Diagnosis		HIE N (%)	Non-HIE N (%)	
Mild HIE	–	37 (48.7)	–	
Moderate HIE	–	29 (38.2)	–	
Severe HIE	–	10 (13.1)	–	
Respiratory (TTN/RDS)	–	–	449 (44%)	
Perinatal asphyxia	–	–	231 (22.9%)	
Suspected infection	–	–	167 (16.5%)	
Low birth weight	–	–	37 (3.6%)	
Congenital anomalies	–	–	29 (2.8%)	
Haemorrhage (subgaleal, maternal)	–	–	27 (2.6%)	
Meconium aspiration	–	–	16 (1.5%)	
Cardiac	–	–	13 (1.2%)	
Neurological (non-encephalopathic)	–	–	5 (0.5%)	
Hypoglycaemia	–	–	3 (0.3%)	
Dysmorphic features	–	–	2 (0.2%)	
Miscellaneous	–	–	26 (2.58%)	
Apgar Scores	All infants	HIE	Non-HIE	
At 1 min	8 (5- 9)	2 (1- 4)	8 (5- 9)	<0.001
At 5 min	9 (7 -9)	5 (3- 7)	9 (8- 9)	<0.001
First postnatal gas	All infants	HIE	Non-HIE	
pH	7.25 (7.19–7.29)	7.14 (7.05–7.23)	7.25 (7.20–7.29)	<0.001
Base deficit	4.0 (2.0–6.8)	11.7 (7.9–15.9)	3.7 (1.9–6.1)	<0.001
Lactate	3.8 (2.2–6.8)	9.6 (7.4–12.8)	3.5 (2.1–6.2)	<0.001

HIE, hypoxic ischaemic encephalopathy; RDS, respiratory distress syndrome; TTN, transient tachypnoea of the newborn.

were incorrectly classified, one had clinical and electrographic seizures and also had an abnormal magnetic resonance imaging (MRI) of brain in the newborn period (single punctate area of high signal within deep white matter on conventional and diffusion imaging). The other moderate case had concern for clinical seizures but did not have electrographic seizures and had a normal MRI brain prior to discharge. The LR model has slightly worse ability to identify cases: identifying 73% of mild HIE, 93% of moderate HIE and 100% of severe HIE cases using the same PI threshold cut-off of 0.30. Of the two extra cases of mild HIE that were incorrectly classified, neither had seizures, and neither were cooled, but one had an abnormal MRI brain (single punctate area on diffusion restriction).

DISCUSSION

In this study we have validated the ability of our ML algorithm to correctly identify infants with HIE within the first hour of

life from a large sample of over 1000 infants with a range of diagnoses requiring admission to a tertiary neonatal unit. This algorithm has the potential to provide a user-friendly and readily accessible support tool in early decision-making in HIE. Using standard clinical variables available in the majority of high-income and middle-income settings, over 86.5% of infants could have been correctly identified as HIE or non-HIE within minutes of birth. Overall, the RF model performed better with higher sensitivity of 86.6% compared with 84.2% with the LR model (both at the PI cut-off of 0.3). The area under the curve of both models was similar and a statistically significant difference between the groups was seen with both models.

When developing our algorithms, we employed two different ML models to explore their capabilities in the early prediction of HIE. LR is a parametrics-based statistical model used for binary classification, predicting categorical outcomes based on logarithmic function.^{23 30} Second, we used an RF model, an

Table 2 HIE group by severity

	Mild n=37	Moderate n=29	Severe n=10	
LR—median PI (IQR)	0.52 (0.26–0.74)	0.78 (0.57–0.93)	0.99 (0.94–0.99)	p=0.003
RF—median PI (IQR)	0.64 (0.31–0.74)	0.72 (0.59–0.84)	0.91 (0.78–0.95)	p=0.003
Therapeutic hypothermia				
Yes	7 (19)	29 (100)	10 (100)	
No	30 (81)	0 (0)	0 (0)	
Death				
Yes	0 (0%)	1 (3.4%)	6 (60%)	
No	37 (100%)	28 (96.5%)	4 (40%)	
Seizures				
Yes	1 (2.7%)	9 (31%)	7 (70%)	
No	36 (97.3%)	20 (69%)	3 (30%)	

HIE, hypoxic ischaemic encephalopathy; LR, logistic regression; PI, Probability Index; RF, random forest.

algorithm that can learn non-linear interactions by averaging the prediction of multiple decision trees and creating a single prediction output.^{26–30} Both models can generate a continuous probability score.^{31–33} While the LR model had less false positives, it missed two more cases of HIE compared with the RF model. As the priority in clinical practice is to identify all possible cases at risk of HIE, our preference for future work with this ML algorithm is to use the RF model.

Our algorithms have been developed through the prior study of a large cohort of carefully characterised infants, examining multiple factors, with detailed maternal, perinatal and postnatal data.^{20–21} The most predictive model was found to include the condition of the infant at birth (Apgar Score at 1 min), and the clinical and biochemical response to resuscitation (Apgar Score at 5 min, and postresuscitation metabolic pH, BD and lactate). During development, the algorithms were trained using all blood gas types. The use of blood gas measurements and Apgar Scores in the prediction of HIE have been used in the past, but many studies use these as dichotomous variables. In reality, each variable provides a proportionate risk. Although dichotomising exposures is a common occurrence to aid interpretation of results, there is an increased likelihood of false positives along with decreasing the extent of variability within each group.^{34–35} ML techniques allow the incorporation of the strength of each marker to aid more accurate prediction.²⁰ While both the LR and RF models had good sensitivity for the prediction of HIE, both misclassified infants. Thus, we do not expect that clinicians would use the model's output in isolation, but rather as

a support tool that considers condition at birth and response to resuscitation when examining risk of HIE.

Our results show a statistically significant difference between HIE grades with both the LR and RF models. Although this highlights that our model could be used to aid with differentiation of the grade of HIE, ongoing use of serial clinical assessment and neuromonitoring is required for accurate assignment of HIE grade.

Both the LR and RF models described have the potential to be strong support tools for clinical decision-making, helping to identify infants who may benefit from specialist care and possibly neuroprotective strategies. Up to 60% of infants with HIE are born in non-tertiary facilities, and the challenges associated with clinical examination and its role in identification of infants with encephalopathy may be even more pronounced in centres with fewer cases.^{36–38} The tool is designed to act as a decision support tool, to identify infants at risk of HIE, triggering clinical assessment and senior review. This may aid appropriate risk stratification for transfer to tertiary and quaternary units. Unnecessary transfers represent a significant financial cost to health systems, with high costs also incurred by families, including financial, emotional and psychological costs.³⁹ By using our models, communication between referral and receiving sites could be greatly improved along with supporting transportation stratification, with the PI offering a supportive objective measure of risk of HIE. We hope that in the future, this model will be further validated in varied populations, to allow the tool to be used across many different settings. Our hope is that the model

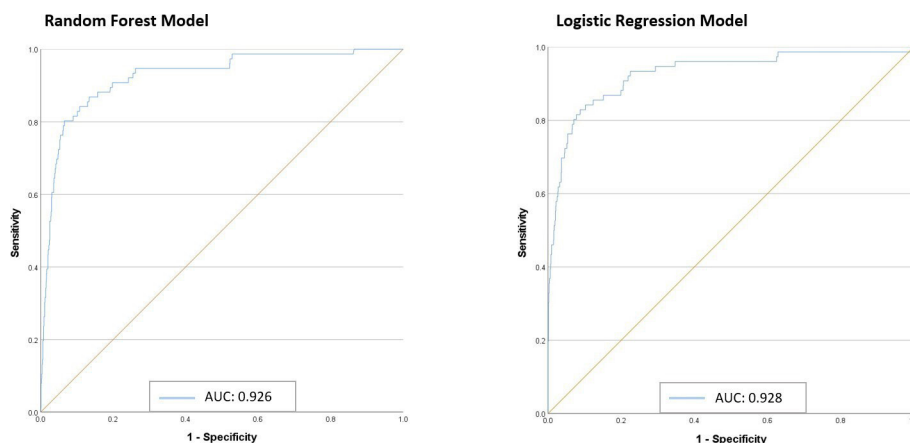


Figure 1 The area under the curve (AUC) of all study samples using the random forest model and the logistic regression model.

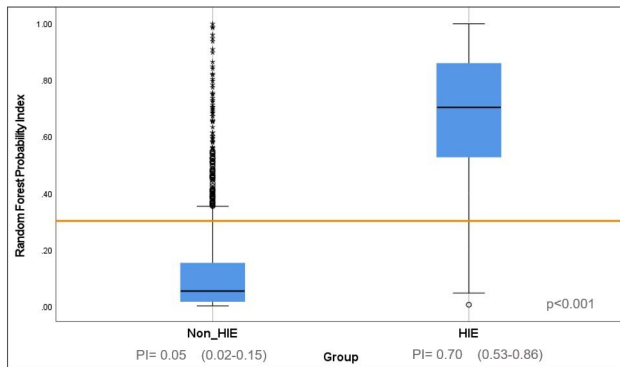


Figure 2 Median Probability Index (PI) in the non-HIE versus HIE groups. HIE, hypoxic ischaemic encephalopathy.

will be used to identify infants early who are at risk of HIE and prompt clinicians to perform a neurological examination and consider further action, to allow infants to receive appropriate referral and timely treatment if required.

The strengths of this study include the large cohort of non-HIE infants with a wide variety of aetiologies with which to validate the use of the algorithm. We have shown that the prediction remains high across all grades of HIE. The clinical variables used in this model are routine clinical measurements that are familiar and attainable. Our model, which can be used with little training or expertise, offers a valuable objective support tool for decision-making. Unlike using these clinical variables in a binary fashion (eg, $\text{pH} < 7.0$ or ≥ 7.0), the algorithm allows for the consideration of the weight of each variable itself, the weight of its value and the unique interaction between variables. By combining multiple clinical variables in this way, the algorithm can offer a more complete picture to aid identification of infants at risk of HIE. This may be especially helpful in smaller centres with less exposure to at-risk encephalopathic infants and may optimise early and targeted transfers, and ultimately, timely initiation of neuroprotective treatments, if indicated.

There were several limitations to this study. The recruiting centre was a tertiary facility in a high-income country. The model has not yet been validated in non-tertiary or international settings, or in middle-income or low-income settings. As this algorithm was designed to be used as a decision aid in the immediate postnatal period, we did not incorporate electroencephalographic (EEG) and MRI data into the diagnosis of HIE. Although this may have provided a more robust assessment of

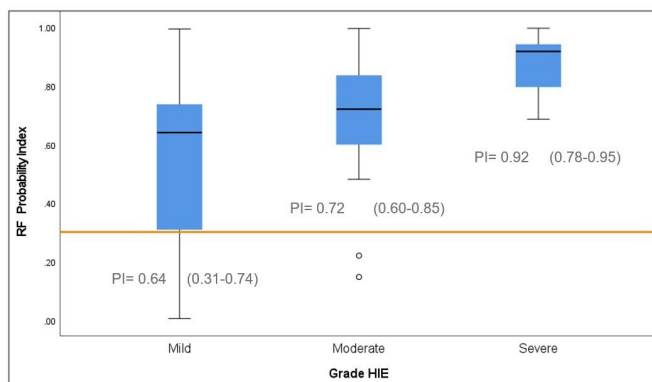


Figure 3 HIE group median Probability Index by severity of HIE. HIE, hypoxic ischaemic encephalopathy; PI, Probability Index; RF, random forest.

HIE, these data are not often available within the first hour of life, so only a clinical diagnosis of HIE was used in validating this prediction model. Our algorithm is open-source and available online. We encourage others to use it and validate its utility in their own local populations.

CONCLUSION

In a large unseen data set an open-source algorithm could identify infants at risk of HIE in the immediate postnatal period. When an infant has a PI Score ≥ 0.3 , this should prompt a focused clinical examination and may aid in identification of infants who require transfer to tertiary care and timely neuroprotective intervention.

Correction notice This paper has been corrected since it was first published. The authors noticed two number errors in the main text of the document. In the Results, paragraph 1, line 3, '122' has been corrected to 110. In Results, paragraph 2, line 1, '1008' has been changed to 1005.

Contributors Guarantor: DMM. DMM had full access to the data in the study and takes responsibility for its integrity and the accuracy of the data analysis. Concept and design: all authors. Acquisition, analysis or interpretation of data: all authors. Drafting of the manuscript: ALM. Critical revision of the manuscript for important intellectual content: BHW, DMM; Statistical analysis: ALM, DSO'B, DMM. Administrative, technical or material support: ALM, BHW, DMM. Supervision: BHW, DMM.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Ethical approval was granted prior to starting the study by the Clinical Research Ethics Committee of The Cork Teaching Hospitals (CREC Review Reference Number: ECM 4 (h) 13/04/2021).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Anne L Murray <http://orcid.org/0009-0008-3548-3969>

Deirdre M Murray <http://orcid.org/0000-0002-2201-9912>

REFERENCES

- Jacobs SE, Berg M, Hunt R, *et al*. Cooling for newborns with hypoxic ischaemic encephalopathy. *Cochrane Database Syst Rev* 2013;2013:CD003311.
- Kurinczuk JJ, White-Koning M, Badawi N. Epidemiology of neonatal encephalopathy and hypoxic-ischaemic encephalopathy. *Early Hum Dev* 2010;86:329–38.
- Lee BL, Gano D, Rogers EE, *et al*. Long-term cognitive outcomes in term newborns with watershed injury caused by neonatal encephalopathy. *Pediatr Res* 2022;92:505–12.
- Conway JM, Walsh BH, Boylan GB, *et al*. Mild hypoxic ischaemic encephalopathy and long term neurodevelopmental outcome - A systematic review. *Early Hum Dev* 2018;120:80–7.
- Volpe JJ. *Neurology of the newborn E-Book*. 2008.
- Lawn JE, Cousens S, Zupan J, *et al*. 4 million neonatal deaths: when? Where? Why? *Lancet* 2005;365:891–900.

- 7 Azzopardi DV, Strohm B, Edwards AD, *et al*. Moderate hypothermia to treat perinatal asphyxial encephalopathy. *N Engl J Med* 2009;361:1349–58.
- 8 Shankaran S, Laptook AR, Ehrenkranz RA, *et al*. Whole-body hypothermia for neonates with hypoxic-ischemic encephalopathy. *N Engl J Med* 2005;353:1574–84.
- 9 Gunn AJ, Thoresen M. Neonatal encephalopathy and hypoxic-ischemic encephalopathy. *Handb Clin Neurol* 2019;162:217–37.
- 10 Peliowski-Davidovich A, Canadian Paediatric Society, Fetus and Newborn Committee. Hypothermia for newborns with hypoxic ischemic encephalopathy. *Paediatr Child Health* 2012;17:41–6.
- 11 Thoresen M, Tooley J, Liu X, *et al*. Time is brain: starting therapeutic hypothermia within three hours after birth improves motor outcome in asphyxiated newborns. *Neonatology* 2013;104:228–33.
- 12 Thompson CM, Puterman AS, Linley LL, *et al*. The value of a scoring system for hypoxic ischaemic encephalopathy in predicting neurodevelopmental outcome. *Acta Paediatr* 1997;86:757–61.
- 13 Sarnat HB, Sarnat MS. Neonatal encephalopathy following fetal distress. A clinical and electroencephalographic study. *Arch Neurol* 1976;33:696–705.
- 14 Murray DM, Bala P, O'Connor CM, *et al*. The predictive value of early neurological examination in neonatal hypoxic-ischaemic encephalopathy and neurodevelopmental outcome at 24 months. *Dev Med Child Neurol* 2010;52:e55–9.
- 15 Walsh BH, El-Shibiny H, Munster C, *et al*. Differences in standardized neonatal encephalopathy exam criteria may impact therapeutic hypothermia eligibility. *Pediatr Res* 2022;92:791–8.
- 16 Arnaez J, Vega-Del-Val C, Hortigüela M, *et al*. Usefulness of video recordings for validating neonatal encephalopathy exams: a population-based cohort study. *Arch Dis Child Fetal Neonatal Ed* 2021;106:522–8.
- 17 Executive summary: Neonatal encephalopathy and neurologic outcome, second edition. Report of the American College of Obstetricians and Gynecologists' Task Force on Neonatal Encephalopathy. *Obstet Gynecol* 2014;123:896–901.
- 18 Abels E, Pantanowitz L, Aeffner F, *et al*. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J Pathol* 2019;249:286–94.
- 19 Obermeyer Z, Emanuel EJ. Predicting the Future - Big Data, Machine Learning, and Clinical Medicine. *N Engl J Med* 2016;375:1216–9.
- 20 Mooney C, O'Boyle D, Finder M, *et al*. Predictive modelling of hypoxic ischaemic encephalopathy risk following perinatal asphyxia. *Heliyon* 2021;7:e07411.
- 21 O'Boyle DS, Dunn WB, O'Neill D, *et al*. Improvement in the Prediction of Neonatal Hypoxic-Ischemic Encephalopathy with the Integration of Umbilical Cord Metabolites and Current Clinical Makers. *J Pediatr* 2021;229:175–81.
- 22 Schwalbe N, Wahl B. Artificial intelligence and the future of global health. *Lancet* 2020;395:1579–86.
- 23 Panesar SS, D'Souza RN, Yeh F-C, *et al*. Machine Learning Versus Logistic Regression Methods for 2-Year Mortality Prognostication in a Small, Heterogeneous Glioma Database. *World Neurosurg* 2019;2:100012.
- 24 Couronné R, Probst P, Boulesteix A-L. Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 2018;19:270.
- 25 Stoltzfus JC. Logistic regression: a brief primer. *Acad Emerg Med* 2011;18:1099–104.
- 26 Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
- 26 Breiman L. Random forest vs logistic regression: binary classification for heterogeneous datasets. *Mach Learn* 2001;45:5–32.
- 28 Paediatrics Fo. Candicy checklist for neonatal therapeutic hypothermia (cooling). Royal College of Physicians of Ireland (RCPI). 2017. Available: <https://www.nntp.ie/library/pdf/NNTP-Cooling-Criteria-Checklist.pdf>
- 26 Breiman L. Neonatal therapeutic hypothermia in Ireland. *Mach Learn* 2001;45:5–32.
- 30 Kirasich K, Smith T, Sadler B, eds. *Random forest vs logistic regression: binary classification for heterogeneous datasets*. 2018.
- 31 Dankowski T, Ziegler A. Calibrating random forests for probability estimation. *Stat Med* 2016;35:3949–60.
- 32 Luo Y, Tseng H-H, Cui S, *et al*. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR Open* 2019;1:20190021.
- 33 Bratko I, ed. *Machine learning: between accuracy and interpretability*. Vienna: Springer Vienna, 1997.
- 34 Altman DG, Royston P. The cost of dichotomising continuous variables. *BMJ* 2006;332:1080.
- 35 Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med* 2006;25:127–41.
- 36 Stetson RC, Brumbaugh JE, Weaver AL, *et al*. Association of outborn versus inborn birth status on the in-hospital outcomes of neonates treated with therapeutic hypothermia: A propensity score-weighted cohort study. *Resuscitation* 2021;167:82–8.
- 37 Jacobs SE, Morley CJ, Inder TE, *et al*. Whole-body hypothermia for term and near-term newborns with hypoxic-ischemic encephalopathy: a randomized controlled trial. *Arch Pediatr Adolesc Med* 2011;165:692–700.
- 38 Natarajan G, Pappas A, Shankaran S, *et al*. Effect of inborn vs. outborn delivery on neurodevelopmental outcomes in infants with hypoxic-ischemic encephalopathy: secondary analyses of the NICHD whole-body cooling trial. *Pediatr Res* 2012;72:414–9.
- 39 Lemmon ME, Donohue PK, Parkinson C, *et al*. Parent Experience of Neonatal Encephalopathy. *J Child Neurol* 2017;32:286–92.

Supplementary File

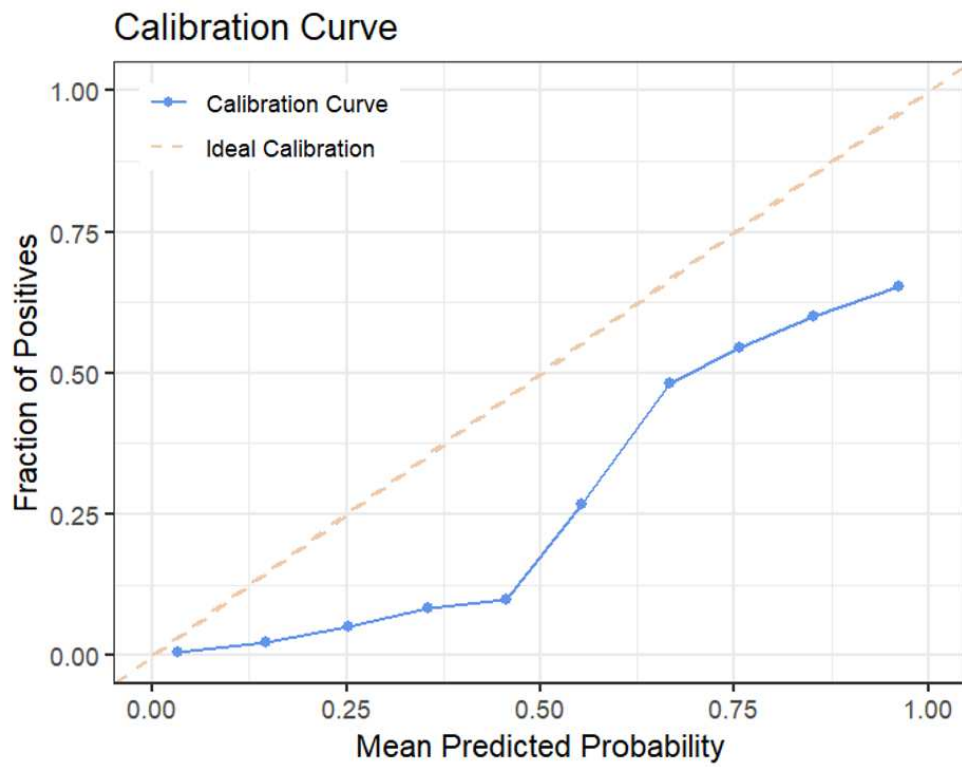


Figure S1: Calibration Curve for Random Forest Model

Table S1: Diagnosis of Excluded Infants	
Diagnosis	N=110
HIE	8
Perinatal asphyxia	26
Suspected infection	13
Respiratory	35
Neurological (other)	1
Cardiac	3
Low birth weight	6
Congenital anomalies/genetic	12
Meconium aspiration	3
Haemorrhage (subgaleal, maternal)	3

Table S2: Diagnosis of False Positive Cases (per Random Forest Model)		
Diagnosis	N=135	Median PI
Perinatal asphyxia	92 (68.1%)	0.47
Suspected sepsis	19 (14%)	0.50
Respiratory	18 (13.2%)	0.46
Neurological (other)	3 (2.2%)	0.56
Cardiac	2 (1.5%)	0.61
Subgaleal haematoma	1 (0.7%)	0.37

PI= Probability Index