

Machine Learning Guided Postnatal Gestational Age Assessment Using Newborn Screening Metabolomic Data in South Asia and Sub-Saharan Africa

Sunil Sazawal (✉ ssazawal@jhu.edu)

Centre for Public Health Kinetics

Kelli Ryckman

University of Iowa

Sayan Das

Centre for Public Health Kinetics

Rasheda Khanam

John Hopkins Bloomberg School for Public Health

Imran Nisar

Aga Khan University

Elizabeth Jasper

University of Iowa

Arup Dutta

Centre for Public Health Kinetics

Sayedur Rahman

PROJAHNMO Research Foundation

Usma Mehmood

Aga Khan University

Bruce Bedell

University of Iowa

Saikat Deb

Public Health Laboratory Ivo de Carneri

Nabidul Chowdhury

PROJAHNMO Research Foundation

Amina Barkat

Aga Khan University

Harshita Mittal

Centre for Public Health Kinetics

Salahuddin Ahmed

PROJAHNMO Research Foundation

Farah Khalid

Aga Khan University

Rubhana Raqib

International Centre for Diarrhoeal Disease Research

Alexander Manu

World Health Organization

Sachiyo Yoshida

World Health Organization

Said Ali

Public Health Laboratory Ivo de Carneri

Abdullah Baqui

John Hopkins Bloomberg School for Public Health

Fyezah Jehan

Aga Khan University

Usha Dhingra

Centre for Public Health Kinetics

Rajiv Bahl

World Health Organization

Muhammad Ilyas

Aga Khan University

Ambreen Nizar

Aga Khan University

Research Article

Keywords: Pre-term births, Machine Learning, Gestational age, new born screening.

Posted Date: January 20th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-143551/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Babies born early and/or small for gestational age in Low- and Middle-income countries (LMIC) contribute substantially to global neonatal and infant mortality. Tracking this metric is critical at a population level for informed policy, advocacy, resources allocation and program evaluation and at an individual level for targeted care. Early prenatal ultrasound is not available in these settings, gestational age (GA) is estimated using newborn assessment, LMP recalls and birth weight, which are unreliable. Algorithms in developed settings, using metabolic screen data, provided GA estimates within 1-2 weeks of ultrasound-based GA. We sought to leverage machine learning algorithms to improve accuracy and applicability of this approach to LMIC settings.

Methods: This study uses data from AMANHI-ACT prospective pregnancy cohorts in Asia and Africa where early pregnancy ultrasound estimated GA and birth weight are available and metabolite screening data in a subset of 1318 newborn are available. We utilized this opportunity to develop machine learning (ML) algorithms. Random Forest Regressor was used where data was randomly split into model-building and model-testing dataset. Mean absolute error (MAE) and root mean square error (RSME) were used to evaluate performance. Bootstrap procedures were used to estimate confidence intervals (CI) for RMSE and MAE. For pre-term birth identification ROC analysis with bootstrap and exact estimation of CI for area under curve (AUC) were performed.

Results: Overall model estimated GA, had MAE of 5.8 days (95%CI 5.6-6.3), which was similar to performance in SGA, MAE 6.3 days (95%CI 5.6-7.0). GA was correctly estimated to within 1 week for 70.9% (95%CI 67.9-73.7). For preterm birth classification, AUC in ROC analysis was 92.6% (95%CI 87.5-96.1; $p < 0.001$). This model performed better than lowa regression, AUC Difference 2.8% (95%CI 0.9-11.8%; $p = 0.021$).

Conclusions: Machine learning algorithms and models applied to metabolomic gestational age dating offer a ladder of opportunity for providing accurate population-level gestational age estimates in LMIC settings. These findings also point to an opportunity for investigation of region-specific models, more focused feasible analyte models, and broad untargeted metabolome investigation.

Background

Of 15 million preterm births annually, 90% happen in Low and Middle Income Countries (LMIC)^{1,2} contributing to 1 million deaths < 5 years, 35% of deaths < 28 days³. Further 23.3 million infants (19.3% of live births) are born small for gestational age (SGA) in LMIC. Reduction by 10.0% in these would reduce neonatal deaths by 254,600 deaths⁴. Identifying and tracking this metric is therefore critical for advocacy, for surveillance, research, evaluation of preventive strategies, and care of these high risk infants in LMIC. Activities essential to achieving United Nations Sustainable Development Goal 3 target 3.25^{5,6} (elimination of preventable under 5 years deaths by 2030)⁶.

The estimation of accurate gestational age at birth is essential for identifying both preterm and SGA births. Early ultrasound examination, considered as gold standard for gestational age (GA) assessment is unavailable due to high equipment cost and lack of trained manpower in most LMIC settings. Recall of last menstrual period (LMP)⁷ used in these settings is unreliable in estimation of GA⁸. Postnatal methods, birthweight and standardized scoring system (Dubowitz or Ballard scales) have poor reliability and high inter user variability limiting their usage⁹⁻¹².

There global health need for novel tools that could help monitor on a population scale these metrics in LMIC countries¹³. Algorithms developed in three North American settings using routine metabolic screen data to derive GA estimates, have been shown to provide accurate estimates to within 1–2 weeks of ultrasound based GA¹⁴⁻¹⁶. Limited data for external validity of these methods in LMIC populations^{17,18} demonstrated satisfactory performance but lower accuracy for GA predictions especially among SGA newborns in Africa and Asia. These models were based on conventional statistical modelling approaches like linear/logistic regression and discriminant analysis. These statistical methods mainly focus on inference from fitting of a project-specific probability model¹⁹. Recent advances in Machine Learning (ML) techniques and big data analysis, allows for efficient handling of large number of predictors while incorporating non-linear association and complex interactions. ML techniques are more robust in nature and they mainly deal with the prediction of outcomes by using general-purpose learning algorithms to find patterns in a dataset. Recently, ML approaches has been shown to help preterm identification in hospital setting²⁰.

We hypothesized that application of ML to metabolite profile datasets of AMANHI-ACT cohorts (Representing both South Asia and Sub-Saharan Africa) would potentially improve the prediction of GA as compared to conventional approaches previously reported. Additionally, method will not be dependent on the North American population datasets for generation of equation coefficients and enable regional adjustments in the future. Among various ML classifiers, we chose to use random forest as particularly well suited for clinical predictions²¹. We report the performance of our machine learning based GA estimation algorithms.

Methods

Study population

This study was undertaken using data from the Alliance for Maternal and Newborn Health Improvement (AMANHI) All children thrive (ACT), community based, prospective pregnancy and New-born cohorts from Pemba (Tanzania), Sylhet (Bangladesh) and Karachi (Pakistan). The rationale for these cohorts and associated bio-bank, procedures, and cohort characteristics have been described elsewhere^{22,23}. One of the objectives of the AMANHI study was to develop and validate programmatically feasible approaches to accurately assess the gestational age of babies after they are born. Briefly, women were enrolled in early pregnancy and followed through delivery and the postpartum period (supplementary Fig. 1). Written

informed consent was obtained from all participants prior to study enrolment for collection of maternal and new-born data and samples. GA was established by ultrasonography²⁴ at screening using the fetal crown rump length (if < 14 weeks gestation)²⁵ or biparietal diameter and femur length (if \geq 14 weeks)^{24,25}. All fetal biometry measurements were measured twice and then averaged for gestational age calculations^{24,27}. Birth weight (5 g sensitivity) was measured using standard newborn weighing scale (SECA corporation, Columbia, MD).

Sample collection and processing

The metabolic screening data from 1283 samples used for this analysis was generated as part of the AMANHI collaboration with Department of Epidemiology, College of Public Health, University of Iowa, for evaluating external validity of the GA estimation methods developed based on American samples^{14,18}. An overview is provided in consort flow diagram (Fig. 1) and protocol process flow (supplementary Fig. 1). Heel prick blood spots were obtained on a protein saver card (Whatman^R 903, GE healthcare,USA), within 24–72 hours of birth from newborns as per standard procedures. The 903 cards were air-dried and stored in air tight zip-lock bags with desiccant at -80 °C and shipped in dry ice to the University of Iowa they were labelled and examined for quality. Then samples were sent to State Hygienic Laboratory, Ankeny, Iowa, USA at regular intervals (ensuring processing before potency window). Sixty-six metabolites (supplementary Table 4) which included amino acid, acylcarnitine's, enzymes and hormones were analyzed using tandem mass spectrometry. Only singleton births were included in the final analysis since analyte values are associated with birth status²⁸.

Selection of the Machine Learning Algorithm

ML algorithms such as artificial neural network (ANN), decision tree (DT) and random forest (RF) were tested to check for the most appropriate tool for this analysis. Amongst these RF regressor was found to be the best algorithm suitable for this analysis²⁹. For implementing RF regressor the dataset was divided with bootstrapping into training and test datasets. The training dataset constituted of 50% of the total sample with equal representation from Africa and Asia.

Architecture of Random Forest Regressor

ML models were generated using metabolite profiles along with birth weight and gender. RF regressor model for the present analysis includes; 1) bootstrap sampling –with replacement and 2) random feature selection – m (10 for this model). In our RF regressor, 10 decision trees were generated from the training samples. This procedure was then repeated until M decision trees were created to form a randomly generated “Forest”. We denoted the hyper-parameters for RF regressor as “mtry” (a variable imputed in the model) at each split node for performing regression, the proposed default value of “mtry” is $p/3$ where p is the number of predictors³⁰.

Implementation of Random Forest Regressor and creation of test and training dataset

Sklearn.ensemble.³¹ (Random Forest Regressor package), a Python module was used for running the RF regressor. NumPy, Scipy and Pandas were used as python dependencies for running the module. R coding was used to create train and test datasets. Equal number of samples were assigned to test and training dataset using R sliders and R randomisation.

Models used for the analysis.

Four different models were used for the prediction of Gestational Age.

Model 1: only one variable per metabolite (for all 66 metabolites) from the profile of the blood metabolites (Supplementary Table 3) along with birthweight and gender.

Model 2: was designed to replicate model published by Ryckman et al. ¹⁴ which included the linear, squared and cubic values of the metabolites as predictors.

Model 3: was designed to replicate the model published by Wilson et al. 2017 ¹⁶ which included the linear, squared and cubic values of the metabolites along with birthweight and gender as predictors.

Model 4 (selected Model): Contained all the predictors used in the above three models. the details of the variables in models have been provided in supplementary table 3.

Best Model Selection

The best model was selected on the basis of root mean square error (RMSE) and mean absolute error (MAE).

The RMSE of a predicted model with respect to the estimated variable x_{model} has been defined as the square root of the mean squared error³².

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_{\text{obs},i} - X_{\text{model},i})^2}{n}}$$

Where, x_{obs} is observed values, x_{model} is modelled values at time i .

Mean absolute error (MAE) has been calculated as

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n}$$

where x_i is the prediction and y_i is the true value.

Confidence intervals for RMSE and MAE

Efficient computation of RMSE, MAE values and 95% confidence interval were estimated using bootstrapped procedures³³ (Python package, (Bootstrapped 0.0.2)³⁴) and for t with a fixed seed number of 1 using boot³⁵ and metrics packages in R.

ROC analysis for evaluating discriminatory ability of the ML based GA

For ROC analysis we used Stata 16.1 (StataCorp LLC, Texas USA) and Medcalc (MedCalc Software Ltd. Belgium). Generation of ROC curve and AUC estimation was performed and interpreted using standard methods^{36,37}. We estimated Youden index J ³⁸

$$J = \max\{sensitivityf[c] + specificityf[c] - 1\}$$

where c ranges over all possible criterion values. Graphically, J is the maximum vertical distance between the ROC curve and the diagonal line. Bootstrapped 95% CI for Youden index and its corresponding criterion value were estimated^{39,40}. 95% CI for sensitivities and specificities were also estimated for a range of fixed and pre-specified sensitivities/specificities³⁵ and 95% CI estimated using bootstrapping^{39,40}. Comparison of ROC curves estimating difference, confidence interval and p-value were also performed using bootstrap methods^{41,42}. For the Bootstrap estimation a fixed seed was used to enable replication of the analysis.

Results

General Characteristic of the Cohort

Data from all 1318 new-borns having new-born metabolic screen analytes were included in the current analysis. Of these 742 samples were from Africa and 576 from Asia (Pakistan and Bangladesh). Baseline characteristics of the sample are provided in Table 2. The distribution of male and female subjects in the cohort was almost in the ratio of 1:1. The mean GA as confirmed by ultrasound was 38.5 ± 1.68 weeks. (mean \pm SD). Sample included 153 (11.6%) preterm birth, 199 (15.1%) low birth weight and 271 (20.6%) small SGA new-borns. Birth weight in African new-borns (3240.75 ± 585.88 g) tended to be higher than Asian new-borns (2774.55 ± 513.99 g). We excluded 35 non-singleton pregnancies in the final analysis.

Table 1
Cohort Characteristics Of Infants Included In The Metabolic Screening Study

Heel Prick Samples	All sites Combined (Total cohort)	Asia (Pakistan and Bangladesh)	Africa (Tanzania)
	N = 1318	N = 576	N = 742
Gender			
Male	695 (52.7%)	268 (46.5%)	428 (57.6%)
Female	623 (47.3%)	308 (53.5%)	315 (42.3%)
Gestational Age Mean ± S.D			
≥ 37 weeks	1165 (88.4%)	492 (85.4%)	673 (90.7%)
< 37 weeks	153 (11.6%)	85 (14.6%)	69 (9.3%)
34–37 weeks	126 (82.4%)	71 (83.5%)	54 (78.3%)
<34 weeks	27 (17.6%)	14 (16.5%)	15 (21.7%)
Birthweight (Mean ± S.D)			
	3037.21 ± 601.67	2774.55 ± 513.99	3240.75 ± 585.88
Birth Weight Category, n(%)			
< 2500 gms	199 (15.1%)	153 (26.6%)	46 (6.2%)
> 2500 gms	1119 (84.9%)	423 (73.4%)	696 (83.8%)
Sga Status			
Yes	272 (20.6%)	91 (15.8%)	181 (24.4%)
Multiple Birth Status	35 (2.7%)	8 (1.4%)	27 (3.6%)
Newborn Sample Collected (Hrs), Mean ± SD			
	49.0 ± 16.2	52.1 ± 19.4	46.6 ± 12.7

Table 2
Mean Abs Error and RSME in weeks in final machine learning model

STATISTICS	Cohort		Africa		Asia	
	Overall	SGA	Overall	SGA	Overall	SGA
Training Data set 50% Pemba Samples (358) + 50% Asian samples (284)	Test Data Set-Pooled remaining 50% Pemba Samples (357) + 50% Asian samples (284)		Test Data Set - remaining 50% Pemba Samples (357)		Test Data Set- remaining 50% Asian Samples (284)	
MAE (95% CI)*	0.83 (0.75– 0.90)	0.87 (0.75– 1.00)	0.91 (0.81– 1.01)	1.15 (0.85– 1.46)	0.74 (0.64– 0.85)	0.80 (0.67– 0.93)
RSME (95% CI)*	1.25 (1.16– 1.36)	1.29 (1.15– 1.47)	1.33 (1.21– 1.48)	1.57 (1.30– 1.91)	1.15 (1.01– 1.33)	1.19 (1.03– 1.41)
1 week difference (%)*	70.97 (67.91– 73.79)	71.50 (65.77– 76.10)	69.57 (64.85– 73.53)	67.20 (50.80– 76.20)	71.40 (66.60– 75.11)	72.37 (65.47– 76.93)
2 weeks difference (%)*	90.26 (85.28– 95.35)	90.30 (81.96– 98.71)	92.83 (90.24– 94.74)	96.25(90.44– 97.63)	87.68 (86.73– 88.21)	82.1 9 (88.78– 89.94)
Training Data set 50% Pemba Samples (358) for Pemba and 50% Asia sample (284) for ASIA			Test Data set 50% Pemba Samples (358)		Test Data set 50% Asia Samples (284)	
MAE (95% CI)*			0.88 (0.84– 0.93)	1.06 (0.99– 1.13)	0.72 (0.66– 0.75)	0.71 (0.67– 0.84)
RSME (95% CI)*			1.14 (1.07– 1.23)	1.49 (1.45– 1.54)	1.11 (1.04– 1.18)	1.14 (1.09– 1.131)
*Bootstrapped, Entire Cohort redone MAE 0.84 (0.82–0.87), RMSE 1.19						
*Detailed description of the analytes used in the models have been given in supplementary information						

Comparison of performance of gestational age estimation models

Initially we evaluated 4 models for performance predicting gestational age (supplementary table 4), the model 1 with only base terms for analytes was least accurate RMSE 1.48, model 2 using variables in final Iowa regression model¹⁴ had RMSE of 1.38 and model 3 using variables in Ontario regression model¹⁵ has RMSE 1.28 (Supplementary table 2), the final all-inclusive model providing a RMSE of 1.25 (95%CI 1.16–1.36) was selected and evaluated further. For identification of preterm births, AUC of Model 4 was significantly better than model 1 [5.8% (95%CI 2.4–9.2; p = 0.0007)] and Model 2 [0.8% (95%CI 0.1–0.2; p = 0.017)] .

Overall model estimated gestation age, had a mean absolute error (MAE) of 5.8 days (95%CI 5.6–6.3), compared to gold standard ultrasound dating. Accuracy was slightly lower in Africa MAE 6.3 days (95%CI 5.6-7.0) than Asia MAE 5.1 days (95%CI 4.2–5.6) (Table 2). Contrary to the results from external validity of regression models^{16,18}, performance in SGA new-borns was not appreciably reduced, MAE 6.1 days (95%CI 5.3-7.0 days). GA was correctly estimated to within 1 week of ultrasound-assigned values for 70.9% (95%CI 67.9–73.7) overall, 69.6% (95%CI 64.8–73.5) in African and 71.3% (96%CI 66.6–75.1) in Asian new-borns. Estimations performed as well in SGA new-borns within 1 week 71.5% (95%CI 65.7–76.1) (Table 2).

To evaluate impact of using a regionally trained algorithm (an important future prospect), we repeated the analysis with machine being trained by African sample for Africa estimations and Asian sample for Asia estimations. The model performance in spite of reduced sample for training improved for both regions RMSE of 1.14 (Africa) and 1.11 (Asia). The precision of MAE improved rather than being reduced Africa 6.1 days(95%CI 5.8–6.5), Asia 5.0 days (95% CI 4.6–5.2) (Table 2).

Model Discrimination of Preterm birth

For ability to classify correctly preterm births ($GA \leq 37$ weeks), model in ROC analysis showed an area under curve (AUC) of 92.6% (BC 95%CI 87.5–96.1; p < 0.001). Criterion of ≥ 37 providing a sensitivity of 82.4% and specificity of 94.6% (Fig. 2). This model provided a significant improvement (difference in AUC 6.4% (95%CI 0.9–11.8%; p = 0.021) over predicting GA by regression models in the same dataset AUC 86% (95CI 84.2–88.1), (Fig. 3b). There was an 11.3% (95% CI 3.7–19.0; p = 0.004) difference in AUC between Africa and Asia (Fig. 3a). The AUC between SGA and non-SGA new-borns did not differ (difference in AUC difference 0.6% ; 95%CI -8.3–9.5%; p = 0.891) (Fig. 3c).

Performance across gestation age categories

Estimation of RMSE and MAE as well as cross tabulation of actual and predicted GA by 2 weekly categories (table 3), indicated that the accuracy of the current set of analytes was diminishing at both margins < 35 as well as above 40 weeks. Application of the model on the entire cohort again to stabilize the estimates, increased discordance below 35 weeks yielded concordance of 47% as compared to 63.1 (35–36 weeks), 57.7 (37–38 weeks), 62.9 (39–40 weeks) and 62.0 (> 40 weeks) (table 3).

Discussion

This study has highlighted promising application of ML methodology to birth weight and new-born metabolomic screening data for improving postnatal prediction of gestation age at birth and discriminating between preterm and term new-borns. It also demonstrated ability of using LMIC data for training ML models and not needing external estimators from developed country datasets. In LMIC setting of South Asia and Sub-Saharan Africa, GA estimates from ML model were within an average of 5.8 days of ultrasound based GA. The ML estimated GA enabled discrimination between pre-term and term births AUC 93% was significantly better than regression estimated GA AUC 86%. The optimal criterion of ≤ 37 weeks providing a sensitivity of 82.3% and specificity of 94.6%.

As against lower performance of previous approaches^{14,16-18}, in estimating GA in SGA sub population, our ML model estimates were within 6.1 days of ultrasound based GA. This also reflected in the finding of a similar proportion with estimated gestation being within 1 week of the ultrasound confirmed gestation, 71.0% overall vs 71.5% in SGA subgroup. Use of data with 50% each of Asian and African data for training the models, was associated with some variation in predictive accuracy for Asia (average of 5.2 days) compared to Africa (average of 6.4 days). Using region specific data for training reduced the variation 6.1 and 6.3 days respectively yet in spite of reduced sample size of training dataset, improved the precision (Table 2). With caution of being preliminary proof of principle, these findings provide a vision for future implementations, wherein region specific training datasets may improve global application of metabolomics based data for gestational age assessment.

Our study had a number of important strengths and also some limitations which need consideration while interpreting the results. The strengths included 1) a sampling frame which utilized samples from both South Asia and East Africa, home to most of the global mortality associated with preterm and SGA births, 2) the study design was nested in a well-described population-based cohort of pregnancy with WHO coordinated and harmonized protocols and SOP, 3) Active surveillance for early pregnancy identification with added measures (menstrual calendar, pregnancy), culminating in harmonized ultrasound based gestation assessment between 8–19 weeks of gestation and 4) Sample collection, storage and shipment SOP based on pilot QC, resulting in high quality of samples. The primary limitation of this study is the participation bias against early preterm and early deaths before sample collection window. Relatively small proportion of actual births in this sub-sample limits our ability to comment on model performance in these sub-groups. Our finding of lower accuracy in pre-term ≤ 34 weeks may either reflect lack of association of the metabolites in that sub-group, a function of lack of sample in that group and/or bias introduced by selective exclusion of early deaths. Additionally we were working with the limitation of small sample size as compared to the usual sample sizes in machine learning universe. We did try to use methods appropriate to accommodating smaller sample sizes, however would not have been protected against extreme chance affecting the sample. The ability to train the model and precision of estimates is somewhat reassuring but would need confirmation.

Preterm births and SGA account for a substantial burden of mortality in first 5 years^{3,4}. Tracking these metrics is therefore critical for advocacy, allocation of resources for surveillance, research, evaluation of preventive strategies, and care of these high-risk infants in low- and middle-income countries^{47,48}. At the

core of this is the estimation of gestational age at birth and being able to discriminate pre-term births accurately. Difference in GA at birth of a week impacts neonatal morbidity, mortality, and long-term outcomes significantly^{43,44}. Our findings provide evidence that ML gestational dating models improve upon the currently-used postnatal gestational age estimation methods^{7,9,11,12,14-18}. However while considering implementation of metabolic gestational dating approaches for robust population-level estimates, current challenges and future opportunities that machine learning brings to this domain need consideration. Heel prick samples for new-born screening are typically collected at least 24 hr after birth to accommodate postpartum fluctuations in analyte levels. This introduces a bias due to early deaths selectively occurring in pre-term births, further in LMIC settings most mother-infant pairs do not stay in hospital beyond 24 after delivery⁴⁵. In most LMIC new-born screening is not a standard practice and will entail challenges in sample collection and processing for metabolic screening, therefore scale up needs to include rethinking about development of cord-blood-specific models restricted to analytes less susceptible to fluctuations in the postnatal environment, establishing a profile of fewer selected metabolites that are measurable in less sophisticated equipment. While rethinking and investigating low-tech variations suitable to LMIC settings, also to consider are, newer high throughput trans proteome/metabolome platforms which are now becoming affordable (i.e. Seers Nano peptide technology⁴⁶). An untargeted metabolomic approach may improve our ability to estimate GA postnatally while also identifying infants at risk of a variety of conditions. Use of a broader spectrum of analytes may also help select a restrictive model for cord blood. Building on this study, use of ML methodology would positively influence development of all the above approaches, due to flexibility, ability to use regional data for ML and not requiring circling back to accumulating large datasets with new intended analyte profiles.

Summary

Towards implementing preterm birth surveillance initiatives⁴⁹ ML algorithms and models applied to metabolomic gestational age dating offer an opportunity ladder to provide accurate population-level gestational age estimates in LMIC settings. Further research should focus on application of ML enabling investigation and incorporation of region-specific models, evaluating broad untargeted metabolome or more focused feasible analyte pool with ML approaches. Derivation and optimization of cord blood metabolic profiles models predicting gestational age accurately would usher a new feasibility for use of this approach in LMIC settings.

List Of Abbreviations

LMIC	Low And Middle Income Countries
SGA	Small For Gestational Age
GA	Gestational Age
LMP	Last Menstrual Period
ML	Machine Learning
AMANHI	Alliance For Maternal And Newborn Health Improvement
ACT	All Children Thrive
ANN	Artificial Neural Network
DT	Decision Tree
RF	Random Forest
RMSE	Root Mean Square Error
MAE	Mean Absolute Error
CI	Confidence Interval
SD	Standard Deviation
WHO	World Health Organisation
QC	Quality Control

Declarations

Ethics approval and consent to participate

All study protocols for AMANHI cohorts were approved by ethical review committees of the WHO and appropriate institutional review board in each of the participating sites. Additionally, institutional/Local sample utilization committees approved shipment of samples to Iowa for metabolic screening assay. Written informed consent for ultrasonography and additional procedures was obtained from all participating mothers in their local or preferred language by study supervisors during enrollment. Mothers were asked for additional informed consent before obtaining a heel prick from the baby.

Consent for publication

Not applicable

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request

Competing interests

No competing interests

Funding

The AMANHI study was funded by the Bill & Melinda Gates Foundation through a grant to the World Health Organization. The funders have played no role in the drafting of the manuscript and the decision to submit for publication.

Authors' contributions

- a. SS, KR, AB, FJ, RB contributed to conceptualization, design and implementation, quality control, participated in analysis and interpretation of data.
- b. SS, S.Das, AD, RB Conceptualized Machine learning exercise and contributed to pooled data analysis and machine learning analysis
- c. RK, IN, EJ, UD, contributed to the implementation, analysis and quality control in the field.
- d. KKR, EJ, BB, contributed to laboratory analysis of samples
- e. UM, AD, NHC, AB, SR, S.Deb, SA, FK, RR, AM, SY contributed to field implementation, data collection and quality control.
- f. RB, AM, SY coordinated the study
- g. SS, SDas, HM, AD additionally contributed to first draft write-up.
- h. All authors reviewed and contributed to manuscript writing.

Acknowledgements

We acknowledge Sergey Feldman for his inputs regarding machine learning packages and algorithms. We acknowledge the contribution of the AMANHI study teams in three countries and support of local participating institutions (Public Health Laboratory-IdC and Ministry of Health in Zanzibar, PROJAHNMOH Research Foundation in Sylhet, Bangladesh and Aga Khan University in Karachi, Pakistan). We sincerely thank all the mothers and families for their participation time and contributions to this study. We also thank the support of local health systems and ethical review boards for their oversight.

References

1. Lawn J.E., Kinney M. Preterm birth: now the leading cause of child death worldwide. *Sci Transl Med.* 2014;6:263ed221
2. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn JE, Cousens S, Mathers C, Black RE. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet.* 2016 Dec 17; 388(10063):3027–3035.

- doi: 10.1016/S0140-6736(16)31593-8. Epub 2016 Nov 11. Erratum in: *Lancet*. 2017 May 13;389(10082):1884. PMID: 27839855; PMCID: PMC5161777.
3. Walani SR. Global burden of preterm birth. *Int J Gynecol Obstet*. 2020; **150**: 31–33.
 4. Lee AnneCC, Kozuki Naoko, Cousens Simon, Stevens Gretchen A, Blencowe Hannah, Silveira Mariangela F *et al*. Estimates of burden and consequences of infants born small for gestational age in low and middle income countries with INTERGROWTH-21st standard: analysis of CHERG datasets *BMJ* 2017; 358:j3677
 5. Liu L, Oza S, Hogan D, Chu Y, Perin J, Zhu J, Lawn JE, Cousens S, Mathers C, Black RE. Global, regional, and national causes of under-5 mortality in 2000-15: an updated systematic analysis with implications for the Sustainable Development Goals. *Lancet*. 2016 Dec 17;388(10063):3027–3035. doi: 10.1016/S0140-6736(16)31593-8. Epub 2016 Nov 11. Erratum in: *Lancet*. 2017 May 13;389(10082):1884. PMID: 27839855; PMCID: PMC5161777
 6. United Nations. Sustainable Development Goals. United Nations; New York: 2015. (accessed Sept 11, 2015). <http://www.un.org.proxy1.library.jhu.edu/sustainabledevelopment/summit/>
 7. Alexander GR, de Caunes F, Hulsey TC, Tompkins ME, Allen M. Validity of postnatal assessments of gestational age: a comparison of the method of Ballard *et al*. and early ultrasonography. *Am J Obstet and Gynecol*. 1992; 166:891–895. DOI: [https://doi.org/10.1016/0002-9378\(92\)91357-G](https://doi.org/10.1016/0002-9378(92)91357-G), PMID: 1550159
 8. Lynch CD, Zhang J. Th research implication for the estimation of gestational age dating method. *Paediatr Perinat Epidemiol*. 2007 Sep;21 Suppl 2:86–96
 9. Taylor R, Beyai S, Owens S, Denison F. The external ballard examination does not assess gestational age accurately in a rural field setting in the Gambia. *Arch of Disease in Childhood - Fetal and Neonatal Edition*.2010; 95:Fa103.
 10. Spinnato JA, Sibai BM, Shaver DC, Anderson GD. Inaccuracy of Dubowitz gestational age in low birth weight infants. *Obstet and Gynecol*. 1984; 63:491–495. PMID: 6700894
 11. Sanders M1, Allen M, Alexander GR, Yankowitz J, Graeber J, Johnson TR and Repka MX. Gestational age assessment in preterm neonates weighing less than 1500 grams. *Pediatrics* (1991). Sep;88(3):542-6
 12. Robillard PY1, De Caunes F, Alexander GR and Sergent MP. Validity of postnatal assessments of gestational age in low birthweight infants from a Caribbean community. *J Perinatol*.1992.Jun;12(2):115–9
 13. Wyber R, Vaillancourt S, Perry W, Mannava P, Folaranmi T, Celi LA. Big data in global health: improving health in low- and middle-income countries. *Bull World Health Organ*. 2015;93(3):203–208. doi:10.2471/BLT.14.139022
 14. Ryckman KK, Berberich SL, Dagle JM. Predicting gestational age using neonatal metabolic markers. *Am J Obstet Gynecol*. 2016; 214(4):515.e1–515.e13. doi:10.1016/j.ajog.2015.11.028
 15. Wilson LA, Murphy MS, Ducharme R, Denize K, Jadavji NM, Potter B, Little J, Chakraborty P, Hawken S, Wilson K. Postnatal gestational age estimation via newborn screening analysis: application and

- potential. *Expert Rev Proteomics*. 2019 Sep;16(9):727–731. doi: 10.1080/14789450.2019.1654863. Epub 2019 Aug 17. PMID: 31422714; PMCID: PMC6816481.
16. Jelliffe-Pawlowski LL, Norton ME, Baer RJ, Santos N, Rutherford GW. Gestational dating by metabolic profile at birth: a California cohort study. *Am J Obstet Gynecol*. 2016; 214(4):511.e1–511.e13. doi:10.1016/j.ajog.2015.11.029
 17. Murphy MSQ, Hawken S, Cheng W, LA Wilson, Lamourex M, Henderson M, Pervin J, Chowdhury J, Gravett C, Lackritz E, Potter BK, Walker M, Little J, Rahman A, Lackritz E, Chakraborty P, Wilson K. External validation of postnatal gestational age estimation using newborn metabolic profiles in Matlab, Bangladesh. *eLife*. 2019; 8: e42627
 18. Sazawal *et al.*,2021. Using AMANHI-ACT cohorts for external validation of Iowa new-born metabolic profiles based models for postnatal Gestational age estimation. In submission
 19. Bzdok, D., Altman, N. & Krzywinski, M. Statistics versus machine learning. *Nat Methods* 15, 233–234 (2018). <https://doi.org/10.1038/nmeth.4642>
 20. Rittenhouse KJ, Vwalika B, Keil A, *et al.* Improving preterm newborn identification in low-resource settings with machine learning. *PLoS One*. 2019;14(2):e0198919. Published 2019 Feb 27. doi:10.1371/journal.pone.0198919
 21. Khaled Fawagreh, Mohamed Medhat Gaber & Eyad Elyan (2014) Random forests: from early developments to recent advancements, *Systems Science & Control Engineering*, 2:1, 602–609, DOI: 10.1080/21642583.2014.956265
 22. AMANHI Study Group 2017. Development and validation of a simplified algorithm for neonatal gestational age assessment– protocol for the Alliance for Maternal Newborn Health Improvement (AMANHI) prospective cohort study. *J Glob Health*. 2017 Dec;7(2):021201. doi: 10.7189/jogh.07.021201.
 23. AMANHI Study Group 2017. Understanding biological mechanisms underlying adverse birth outcomes in developing countries:protocol for a prospective cohort (AMANHI bio–banking) study. *J Glob Health*. 2017 Dec;7(2):021202. doi: 10.7189/jogh.07.021202.
 24. Butt K, Lim KI. Guideline No. 388-Determination of Gestational Age by Ultrasound. *J Obstet Gynaecol Can*. 2019 Oct;41(10):1497–1507. doi: 10.1016/j.jogc.2019.04.010. PMID:3154803
 25. Ohuma, E.O., Papageorghiou, A.T., Villar, J. *et al.* Estimation of gestational age in early pregnancy from crown-rump length when gestational age range is truncated: the case study of the INTERGROWTH-21st Project. *BMC Med Res Methodol* 13, 151 (2013). <https://doi.org/10.1186/1471-2288-13-151>
 26. Salomon LJ, Alfirevic Z, Da Silva Costa F, Deter RL, Figueras F, Ghi T, Glanc P, Khalil A, Lee W, Napolitano R, Papageorghiou A, Sotiriadis A, Stirnemann J, Toi A, Yeo G. ISUOG Practice Guidelines: ultrasound assessment of fetal biometry and growth. *Ultrasound Obstet Gynecol*. 2019 Jun;53(6):715–723. doi: 10.1002/uog.20272. PMID: 31169958
 27. Aris T Papageorghiou, Eric O Ohuma, Douglas G Altman *et al.* International standards for fetal growth based on serial ultrasound measurements: the Fetal Growth Longitudinal Study of the

- INTERGROWTH-21st Project. *Lancet* 2014; 384: 869–79
28. Ochiai M, Matsushita Y, Inoue H, *et al.* Blood Reference Intervals for Preterm Low-Birth-Weight Infants: A Multicenter Cohort Study in Japan. *PLoS One*. 2016;11(8):e0161439. Published 2016 Aug 23. doi:10.1371/journal.pone.0161439
 29. Landset, S., Khoshgoftaar, T.M., Richter, A.N. *et al.* A survey of open source tools for machine learning with big data in the Hadoop ecosystem. *Journal of Big Data* 2, 24 (2015).
<https://doi.org/10.1186/s40537-015-0032-1>
 30. Breiman, L. Random Forests. *Machine Learning* **45**, 5–32 (2001).
<https://doi.org/10.1023/A:1010933404324>
 31. Pedregosa *et al.*, Scikit-learn: Machine Learning in Python, *JMLR* 12, pp. 2825–2830, 2011.
 32. T. Chai and R. R. Draxler. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, 7, 1247–1250, 2014. doi:10.5194/gmd-7-1247-2014.
 33. DiCiccio, Thomas J.; Efron, Bradley. Bootstrap confidence intervals. *Statist. Sci.* 11 (1996), no. 3, 189–228. doi:10.1214/ss/1032280214.
 34. <https://pypi.org/project/bootstrapped/>
 35. Canty A, Ripley BD (2020). boot: Bootstrap R (S-Plus) Functions. R package version 1.3–25
 36. Zhou XH, Obuchowski NA, McClish DK (2002) *Statistical methods in diagnostic medicine*. Wiley-Interscience.
 37. Hanley JA, McNeil BJ (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143:29–36.
 38. Shan G. Improved Confidence Intervals for the Youden Index. *PLoS One*. 2015 Jul 1;10(7):e0127272. doi: 10.1371/journal.pone.0127272. PMID: 26132806; PMCID: PMC4488538.
 39. Platt RW, Hanley JA, Yang H. Bootstrap confidence intervals for the sensitivity of a quantitative diagnostic test. *Statistics in Medicine* 2000; 19:313–322.
 40. Efron and Tibshirani, *An introduction to the bootstrap*. Chapman & Hall, London 436 p, 1993
 41. De Long *et al.* 1988. Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*. Vol. 44, No. 3 (Sep., 1988)
 42. Hanley JA, McNeil BJ (1983) A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology* 148:839–843.
 43. Parikh LI, Grantz KL, Iqbal SN, Huang CC, Landy HJ, Fries MH, Reddy UM. Neonatal outcomes in fetuses with cardiac anomalies and the impact of delivery route. *Am J Obstet Gynecol*. 2017 Oct;217(4):469.e1-469.e12. doi: 10.1016/j.ajog.2017.05.049. Epub 2017 May 31. PMID: 28578168; PMCID: PMC5793863.
 44. Boyle EM, Poulsen G, Field DJ, Kurinczuk JJ, Wolke D, Alfirevic Z, Quigley MA. Effects of gestational age at birth on health outcomes at 3 and 5 years of age: population based cohort study. *BMJ*. 2012 Mar 1;344:e896. doi: 10.1136/bmj.e896. PMID: 22381676; PMCID: PMC3291750.

45. Campbell OM, Cegolon L, Macleod D, Benova L. Length of Stay After Childbirth in 92 Countries and Associated Factors in 30 Low- and Middle-Income Countries: Compilation of Reported Data and a Cross-sectional Analysis from Nationally Representative Surveys. *PLoS Med*. 2016 Mar 8;13(3):e1001972. doi: 10.1371/journal.pmed.1001972. PMID: 26954561; PMCID: PMC478307
46. Blume, J.E., Manning, W.C., Troiano, G. et al. Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nat Commun* 11, 3662 (2020). <https://doi.org/10.1038/s41467-020-17033-7>
47. Farag TH, Koplan JP, Breiman RF, Madhi SA, Heaton PM, Mundel T, Ordi J, Bassat Q, Menendez C, Dowell SF. Precisely Tracking Childhood Death. *Am J Trop Med Hyg*. 2017 Jul;97(1):3–5. doi: 10.4269/ajtmh.16-0302. PMID: 28719334; PMCID: PMC5508885
48. Walani SR. Global burden of preterm birth. *Int J Gynaecol Obstet*. 2020 Jul;150(1):31–33. doi: 10.1002/ijgo.13195. PMID: 32524596. 41.March of Dimes. 2012. Partnership for maternal newborn and child health, save the children & WHO. Born Too Soon: The Global Action Report on Preterm Birth
49. WHO. Every Newborn Action Plan. WHO; 2014. 978924150744843. [http://www.un.org.proxy1.library.jhu.edu/sustainabledevelopment/summit/United Nations. Sustainable Development Goals. United Nations; New York: 2015. \(accessed Sept 11, 2015\).](http://www.un.org.proxy1.library.jhu.edu/sustainabledevelopment/summit/United Nations. Sustainable Development Goals. United Nations; New York: 2015. (accessed Sept 11, 2015).)

Tables

Due to technical limitations, table 3 is only available as a download in the Supplemental Files section.

Figures

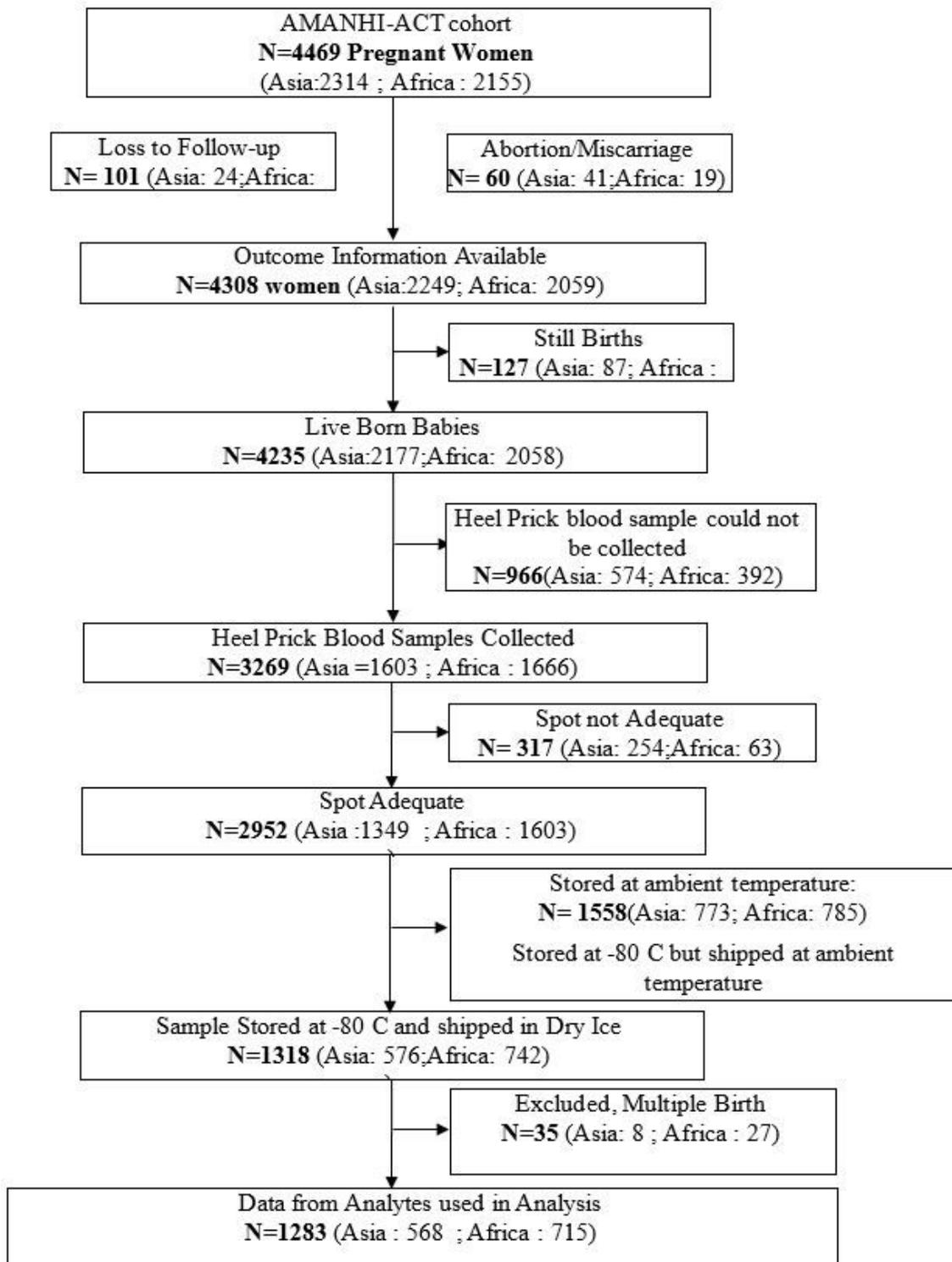


Figure 1

CONSORT COHORT FLOW DIAGRAM

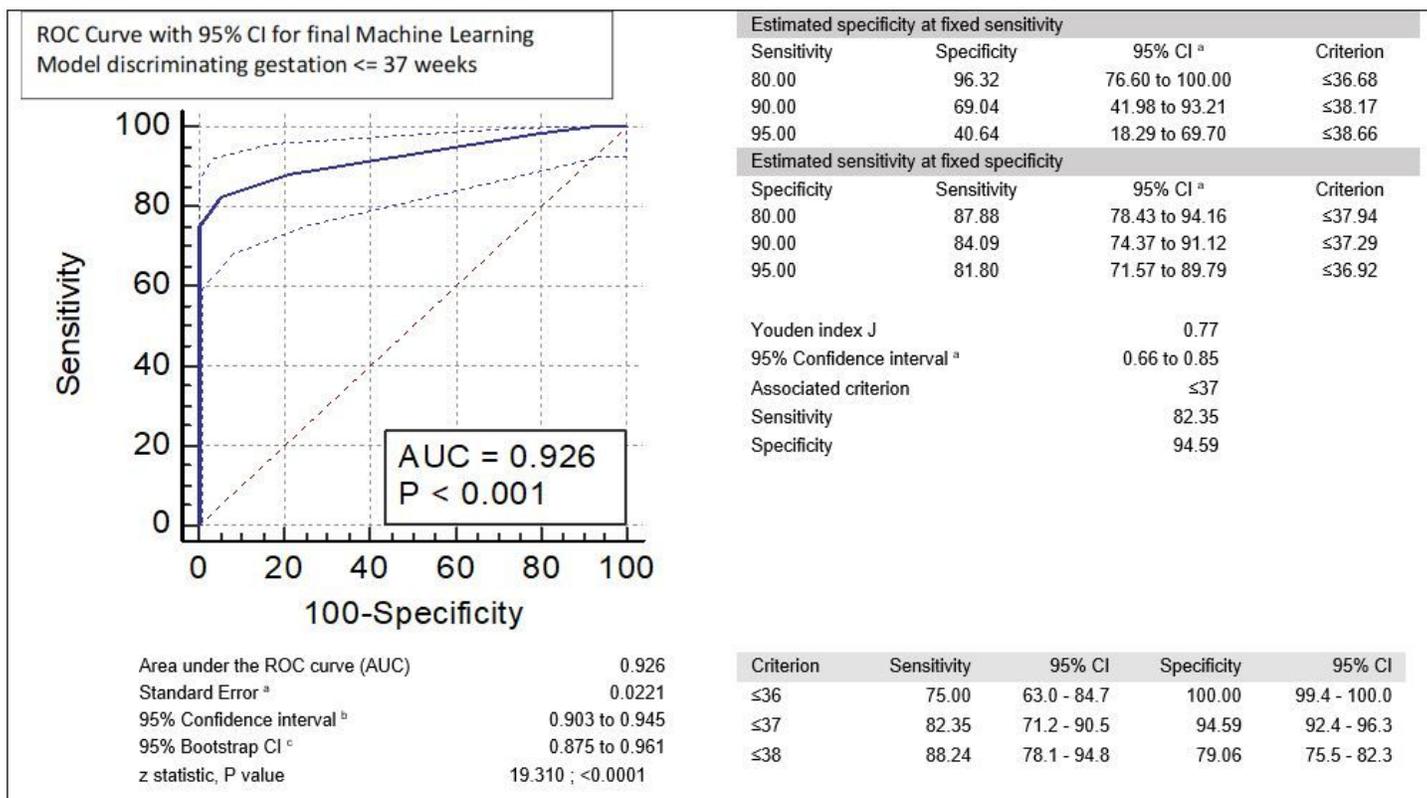
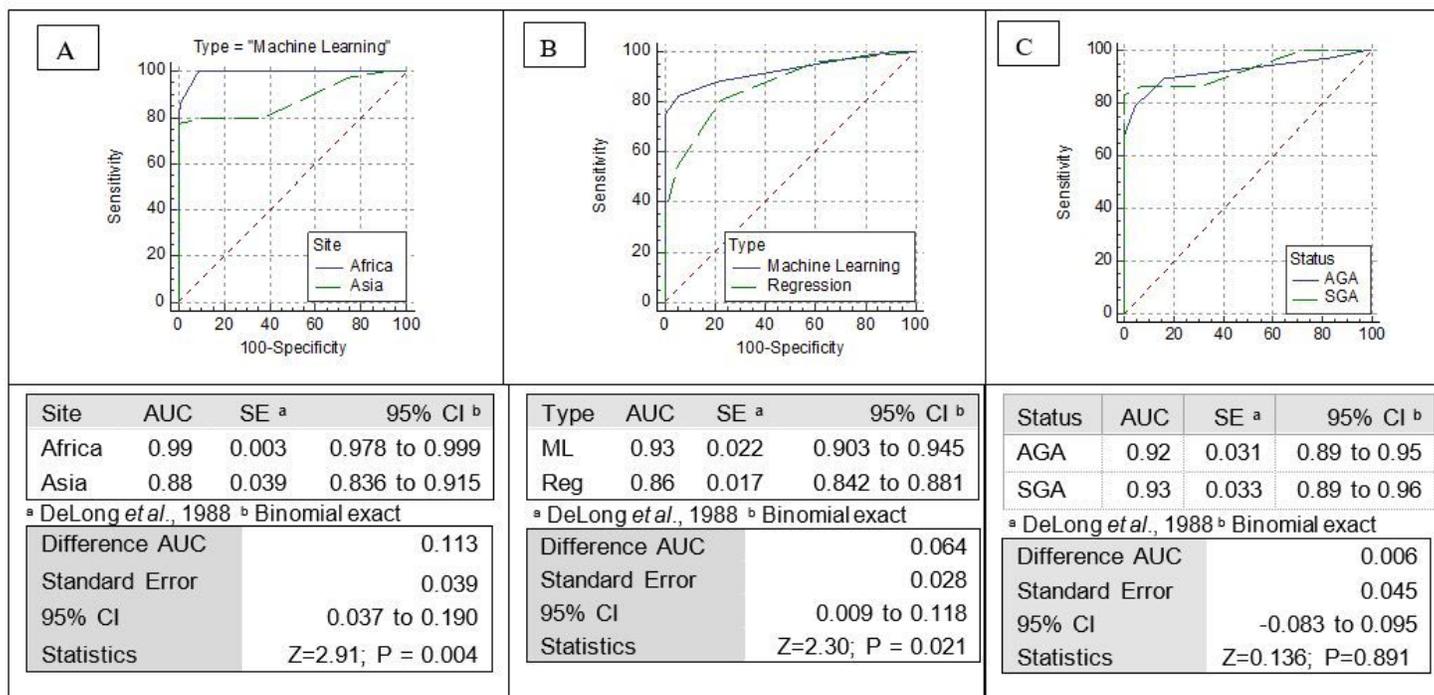


Figure 2

ROC analysis of Machine Learning Final Model in discrimination of gestation < 37 weeks



- A. By site
- B. With the estimates obtained from primary published regression analysis
- C. By SGA infants

Figure 3

ROC analysis comparing Machine Learning performance

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation.docx](#)
- [Table3.jpg](#)