

Can machine learning screen for predictive risk model for spontaneous preterm birth?

Dr Rachel Phoy Cheng Chun¹, Miss Hiu Gwan Chan¹, Assistant Professor Gilbert Yong San Lim², Associate Professor Devendra Kanagalingam³, Dr Pamela Partana³, Professor Kok Hian Tan¹, Professor Tiong Ghee Teoh¹, Dr Ilka Tan¹

¹KK Women's and Children's Hospital, Singapore, Singapore.

²SingHealth AI Office, Singapore, Singapore.

³Singapore General Hospital, Singapore, Singapore

Background

Preterm birth (PTB) is a major contributor to perinatal morbidity and mortality globally. Accurate identification of at-risk pregnancies is essential for timely intervention and improved neonatal outcomes. Recent advances in machine learning offer new opportunities for data-driven risk stratification and early prediction of PTB.

Objective

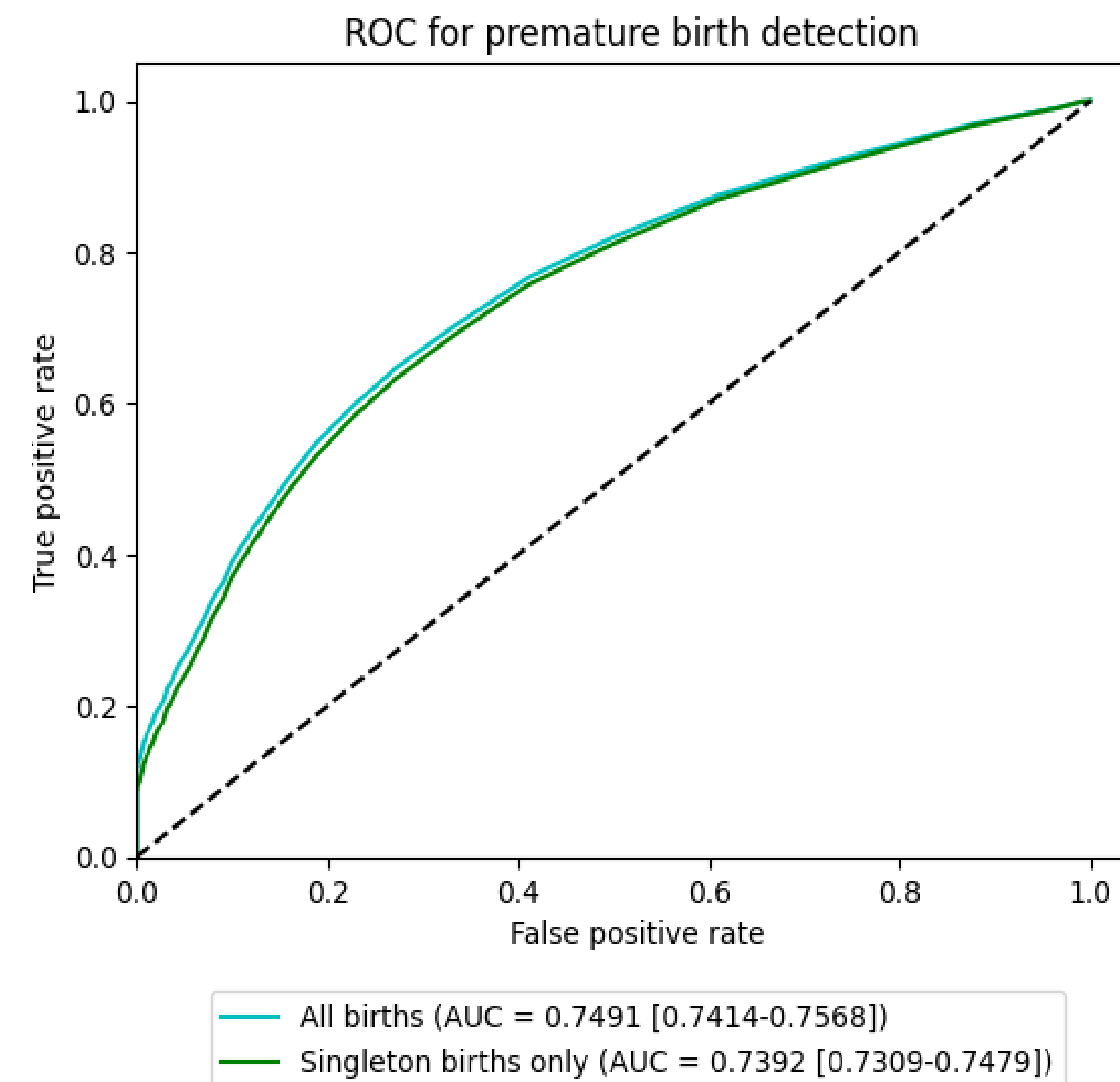
Analyse PTB trends in Singapore from 2017 to 2023, identify risk factors in this multi-ethnic population, and evaluate the effectiveness of a machine learning model-specifically XGBoost in predicting spontaneous PTB.

Design/Method

A retrospective analysis was conducted using data from the Singapore Registry of Births and the SingHealth cluster, encompassing all PTBs between 22+0 and 36+6 weeks of gestation from 1 January 2017 to 31 December 2023. Maternal and neonatal variables were extracted from electronic medical records. Statistical analyses included the Cochran-Armitage trend test and multinomial logistic regression to assess temporal trends and risk factors. For predictive modelling, an XGBoost algorithm was trained and validated to estimate the risk of spontaneous PTB based on clinical and demographic features.

Results

While the overall PTB rate in Singapore remained stable during the study period, the SingHealth population exhibited a modest decline in PTB rates (from 11.3% to 10.2%), primarily due to reductions in late spontaneous PTB. Spontaneous PTBs constituted approximately 60% of all cases. Key predictors identified by both statistical and ML analyses included Chinese ethnicity, maternal age ≥ 35 years, BMI ≥ 23 kg/m², unmarried status, primiparity, twin pregnancy, and maternal blood group AB. The XGBoost model achieved an AUROC of 0.75, demonstrating moderate predictive performance and validating the utility of ML approaches for individualised PTB risk assessment.



Conclusion

Despite stable national PTB rates, local risk factor profiles remain pivotal for targeted prevention. This study demonstrates that machine learning models, such as XGBoost, can effectively integrate complex, multi-dimensional clinical data to predict spontaneous PTB risk with moderate accuracy. The adoption of machine learning-driven risk stratification holds promise for enabling personalised prenatal care and optimising intervention strategies in diverse populations.