Development and validation of a machine learning model for prediction of shoulder dystocia

Stanford University School of Medicine, Palo Alto, United States

Objective
We sought to develop a machine learning (ML) model for prediction of shoulder dystocia (ShD) and to externally validate the model accuracy and potential clinical efficacy in optimizing the use of cesarean delivery (CD) in the context of suspected macrosomia.

Methods
We used electronic health records (EHR) from the Sheba Medical Center in Israel to develop the model (derivation cohort) and EHR from the University of California San Francisco Medical Center to validate the model accuracy and clinical efficacy (validation cohort). Subsequent to inclusion and exclusion criteria, the derivation cohort consisted of 686 deliveries [131 complicated by ShD], and the validation cohort of 2,584 deliveries [31 complicated by ShD]. For each of these deliveries, we collected maternal and neonatal delivery outcomes coupled with maternal demographics, obstetric clinical data and sonographic biometric measurements of the fetus. Biometric measurements and their derived EFW were adjusted (aEFW) to the date of the delivery. A ML pipeline was utilized to develop the model.

Results
In the derivation cohort, the ML model provided significantly better prediction than the current paradigm: using nested cross validation the area under the receiver operator characteristics curve (AUC) of the model was 0.793 ± 0.041, outperforming aEFW and diabetes (0.745 ± 0.044, p-value = 1e-16). Similarly, in the validation cohort the model outperformed aEFW and diabetes (AUC = 0.866 vs. 0.784, p-value = 0.00007). Additionally, in the validation cohort, among the subgroup of 273 women carrying a fetus with aEFW above 4,000 g, the aEFW had no predictive power (AUC = 0.548), and the model performed significantly better (0.775, p-value = 0.0002). A risk-score threshold of 0.5 stratified 42.9% of deliveries to the high-risk group that included 90.9% of ShD cases and all cases accompanied by maternal or newborn complications. A more specific threshold of 0.7 stratified only 27.5% of the deliveries to the high-risk groups that included 72.7% of ShD cases, and all those accompanied by newborn complications.

Conclusion
We developed a ML model for prediction of ShD andexternally validated the model performance in a different cohort. The model predicted ShD better than EFW+ maternal diabetes and was able to stratify the risk of ShD and neonatal injury in the context of suspected macrosomia.