Ma. Sheila A. Magboo • Vincent Peter C. Magboo

Department of Physical Sciences and Mathematics
College of Arts and Sciences
University of the Philippines Manila

Feature Importance applied to Late Intrauterine Growth Restriction

Introduction

Late intrauterine growth restriction (IUGR), associated with peri-natal hypoxic events and suboptimal neurodevelopment, is a leading cause of perinatal mortality and unexplained stillbirths. These risks could be reduced through prompt antenatal diagnosis and treatment with timely delivery. Our objective is to predict late IUGR on publicly available cardiotocographic (CTG) dataset using machine learning (ML).

Methods

We used antepartum CTG recordings from Azienda Ospedaliera Universitaria Federico II (Napoli, Italy) taken from IEEE Dataport. It contains fetal heart rate indices at different time scales and domains. Pre-processing steps done were outlier detection, normalization, and feature selection as a majority vote among recursive feature elimination, correlation analysis with target and among predictor variables. Following 70:30 split and 10-fold cross validation, several ML models (AdaBoost, support vector machine (SVM), logistic regression (LR), Naïve Bayes (NB), k-nearest-neighbors (kNN), decision tree (DT), and random forest (RF)) were applied followed by performance assessment and feature importance generation. Three experiments were assessed: outlier detection only, feature selection only and with outlier detection and feature selection.

Results

The best models were LR and SVM with highest accuracy (81-82%) in experiment 1, and highest accuracy (84-85%), recall (85-89%) and F1-scores (82-84%) for experiment 2. In experiment 3, metrics did not improve for all models. Our results highlight the importance of feature selection as it provides superior performance. The top important features relevant to late IUGR prediction were: DR_T5_s5[bpm], DR_T40_s1[bpm], both from complexity domain, and MF[ms2/Hz] from frequency domain.

Conclusion

Late IUGR is a leading cause of perinatal mortality for which prompt antenatal diagnosis could be utilized to reduce the risk. The best models were obtained by LR and SVM with superior performance, showing robustness and flexibility as predictive models. We further stressed the importance of feature selection to improve performance and to identify relevant features in predicting late IUGR. Obstetricians used CTG to identify fetal abnormalities and is a deciding factor for medical intervention to the baby. As CTG interpretation is subjective, ML models can be utilized as a decision-support tool. It is recommended that further development of other ML models be applied for better classification performance.

Keyword: cardiotocography, late IUGR, feature importance.