OBUS – Multiple Gestation Overview

Summary

The Multiple Gestation (MG) model predicts whether a pregnancy consists of one or multiple fetuses based on blind ultrasound sweeps of the gravid abdomen. It is important to detect multiple gestations as early as possible: (1) They are inherently more risky than single gestations during pregnancy, labor, and delivery, so may be referred to a higher level of care and/or monitored more frequently; and (2) algorithms targeting other features, such as gestational age and fetal presentation, will likely be less accurate for multiple gestations.

A basic assumption is that the same model can perform well across the gestational age range present in the dataset. This may not be true for two reasons: (1) there are very few MG exams in the very low age gestational age range; and (2) the appearance of very young fetuses where distinct anatomical structures may not yet be visible is very different from that of fully formed fetuses. Validation results show the model struggles at low GA; it is plausible that the little low-GA data in training is actually hurting overall performance.

The model functions as a binary classifier, with single gestation represented as the negative class, and multiple gestation as the positive class. The input is the standard series of six blind sweep videos from an exam. Please refer to [0.1 OBUS Data Description]. and the output is a prediction (in the form of a score between 0 and 1) of whether the pregnancy is a single gestation or multiple gestation. A score threshold must be applied to determine the multiple gestation classification.

Accuracy assessment

The accuracy of the multiple gestation (TWIN) classification model is assessed using several machine learning metrics. The relevant metrics are (a) accuracy (TN+TP)/(TN+FP+FN+TP); (b) sensitivity TP/(TP+FN); (c) specificity TN/(TN+FP); (d) precision (also called purity and positive predictive value) TP/(TP+FP); and (e) AUC-ROC (area under the ROC curve), which is independent of classification threshold.

We strove to achieve a sensitivity ~90% and specificity ~95% in detecting multiple gestations. The gestational age range at which this performance is relevant has not been fully defined; thus, we assess the performance on the whole test set, with a distribution of gestational ages roughly matching that of the overall data set.