OBUS – Estimated Fetal Weight Data

Overview

The extended FAMLI dataset (FAMLI2_enrolled, FAMLI2, and FAMLI3) was used to train and evaluate fetal weight models. Different versions of the combined dataset are described in [0.1 OBUS Data Description]. The results shown in the [3.4 Fetal Weight Evaluation Report] were based on models developed using the v9.3 dataset (see Table 2 in [0.1 OBUS Data Description]). We will thus only describe this final dataset and its construction here.

Label Definition

For the EFW model, there are some differences in the way the ground truth labels are determined compared to conventional features, because fetal weight is itself an estimated value, not a direct measurement. The ground truth for the EFW algorithm is the fetal weight value estimated formulaically from fetal biometric measurements: biparietal diameter (BPD), head circumference (HC), abdominal circumference (AC), and femur length (FL). There are a number of published formulas, but the one chosen for this study is the Hadlock 4-component formula [1]:

$$\begin{aligned}\log_{10}(EFW) = \quad & 1.3596 - 0.00386 \bullet AC \bullet FL + 0.0064 \bullet HC \\ & + 0.00061 \bullet BPD \bullet AC + 0.0424 \bullet AC + 0.174 \bullet FL\end{aligned}$$

Note that the Hadlock formula was developed based on pregnancies of Caucasian women and does not always translate to other populations [2].

The ground truth biometric measurements are based on standard plane ultrasound images of the relevant fetal anatomical components. These images are included in the FAMLI dataset, either as standalone images or fly-to videos. The biometric measurements are typically performed manually on the ultrasound acquisition devices with the aid of embedded software tools. Each biometric measurement is performed multiple times, and the final values are taken as the average of the measurements. Most ultrasound devices also calculate an EFW formula internally based on these biometric measurements and provide the estimated fetal weight as an output, which is recorded in the metadata spreadsheets.

Data filtering for reliable ground truth

Two levels of filtering are applied to ensure data quality. First, among multiple biometric measurements, any measurement that is an outlier (*i.e.*, value deviates too far from the median measurement) is discarded. Second, if the difference between the device-provided EFW and the EFW calculated from the Hadlock formula deviates by more than 2% of the Hadlock value, the entire exam is discarded. Based on this quality filtering, the data available for training and evaluation of the model is reduced from the total data available. The results of this filtering can be seen in Table 1 below.

Data Distributions

Training, validation, and testing splits were generated by the GHL machine learning team by stratifying on estimated fetal weight such that each split had roughly the same fetal weight distribution as the entire dataset. The statistics and purposes of these data subsets are provided in Table 1.

Subset	Patients	Exams	Videos	Purpose
Training	3,608	12,047	215,304	Training model weights via backprop
Validation	481	1,652	29,289	Optimizing hyperparameters
Testing	721	2,501	27,991	Estimating performance

Table 1. Data splits by patients, exams, and videos

References

[1] F. Hadlock, R. Harrist, R. Sharman, R. Deter and S. Park, "Estimation of fetal weight with the use of head, body, and femur measurements—A prospective study," *Am. J. Obs. & Gyn.,* vol. 151, no. 3, pp. 333-337, 1985.

[2] J. Ma, D. Cheng, Z. Zhang, B. Cai and X. Xu, "Evaluating the accuracy of sonographic fetal weight estimations using the Hadlock IV formula in a Chinese population," *Quantitative Imaging in Medical Surgery,* vol. 13, no. 6, pp. 3726-3734, 2023.