

ISBI 2026 - Ranking and Scoring Criteria of the FETUS Challenge

Key requirements for the ranking process:

The final rankings were determined by a weighted combination of three components:

Classification score: 40% weight, **Segmentation score:** 40% weight, and **Processing Time score:** 20% weight. Considering that the objective of this task is to comprehensively evaluate the risk of fetal congenital heart disease through both classification and segmentation, we deemed classification and segmentation to be more critical than processing time, higher weights are assigned to these two aspects compared to processing time. This comprehensive evaluation aligns with the challenge's objective to assess both segmentation, classification, and computational performance, ensuring a multi-dimensional assessment of participants' solutions.

The overall score is calculated as follows:

$$S_{total} = 0.4 * S_{cls} + 0.4 * S_{seg} + 0.2 * S_{time}$$

Where S_{cls} represents classification score, S_{seg} represents segmentation score, and S_{time} represents processing time score. Figure 1 illustrates the overall score framework of our challenge.

1. Classification Score: Class imbalance is an inherent issue in this task. To comprehensively evaluate classification performance, we adopt the mean F1-score across all classes as the final classification score. The calculation formula is as follows:

$$S_{cls} = \frac{1}{N} \sum_{i=1}^N F1_i$$

Where N denotes the total number of classes.

2. Segmentation Score: As shown in Figure 1, in the semi-supervised segmentation task, segmentation score is calculated based on the segmentation performance of **13 anatomical structures** across **four standard cardiac planes**, with the final segmentation score obtained by averaging the results across all structures and planes.

For each cardiac plane, the score is calculated based on two evaluation metrics — Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD) — which are equally

weighted (50%). Considering that all anatomical structures are of comparable clinical importance, each structure contributes equally to the overall segmentation performance. The segmentation score is calculated as follows:

$$S_{seg} = \frac{1}{M} \sum_{s=1}^M S_{(s)}$$

Where

$$S_{(s)} = \frac{DSC_{(s)} + NSD_{(s)}}{2}$$

Here, s denotes the anatomical structure (e.g., atria, ventricles, great vessels, etc.), and M represents the total number of anatomical structures. The final segmentation score is obtained by averaging the results across all M structures within the four standard cardiac views (four-chamber view, left ventricular outflow tract view, right ventricular outflow tract view, and three-vessel–trachea view).

3. Processing Time Score: The total time was defined as the duration from the first to the last test case for each participant. To control for extreme values, the time usage was clipped relative to the baseline time as follows:

- If a participant's time was $\leq 2/3$ of the baseline time, their time usage was set to – $(2/3 \times \text{baseline time})$.
- If a participant's time was ≥ 2 of the baseline time, their time usage was set to – $(2 \times \text{baseline time})$.
- Otherwise, it was set to – (total time).

After min-max normalization and score transformation was applied to yield a [0, 100] score.

The processing time score is calculated as follows:

$$S_{time} = \frac{(-t_{user}) - (-T_{max})}{(-T_{min}) - (-T_{max})} * 100$$

Where:

- t_{user} is the participant's (clipped) total time usage,
- $T_{min} = 2/3 \times \text{baseline time}$ is the fastest acceptable case (upper bound);

- $T_{max} = 2 \times \text{baseline time}$ is the slowest acceptable case (lower bound).

The baseline runtime will be released together with the dataset.

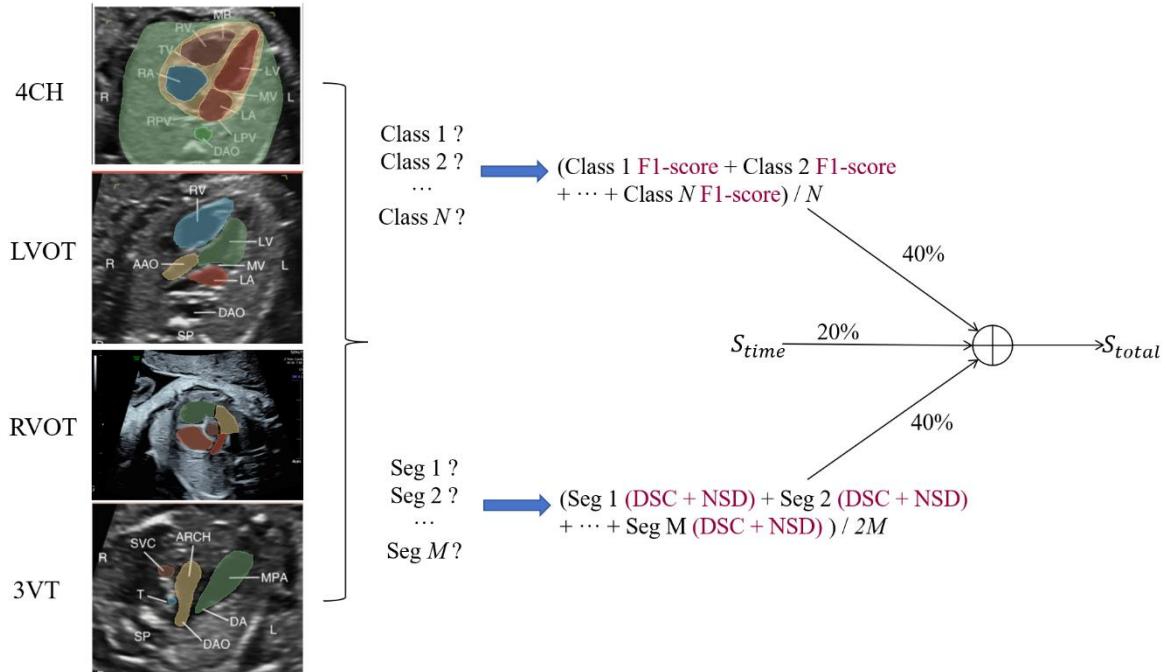


Figure 1. Each case includes B-mode fetal cardiac ultrasound images from four standard views: the four-chamber view (4CH), left ventricular outflow tract view (LVOT), right ventricular outflow tract view (RVOT), and three-vessel–trachea view (3VT). For each case, segmentation and classification results should be provided separately for each of the four views, with an independent classification result generated for every view.

Awards and Prizes:

We will provide cash prizes for the top 5 teams. A certificate will be awarded to the top 10 teams. The top 10 performing methods (teams) will be announced publicly and invited to give oral presentations during the ISBI 2026 conference.