

# Fetal Head Circumference Measurement from Ultrasound Images

Lab Report

**Abstract**—This study explores automated fetal head circumference (HC) measurement from 2D ultrasound images using the HC18 dataset. We compare three approaches: classical image processing with OpenCV ellipse fitting, Random Forest regression, and U-Net deep learning segmentation with different loss functions. The OpenCV baseline achieves 0.71mm MAE by directly fitting ellipses to annotations. Random Forest regression achieves 4.18mm MAE using extracted features. For U-Net segmentation, we compare MSE, BCE, and BCE+IoU loss functions, finding that BCE+IoU loss is essential for learning meaningful segmentation masks.

## I. INTRODUCTION

Fetal head circumference measurement is essential for monitoring fetal growth during pregnancy. Manual measurement from ultrasound images is time-consuming and subject to inter-observer variability. This work explores automated approaches using the HC18 dataset, which contains ultrasound images with annotated ellipses marking the fetal head.

## II. EXPLORATORY DATA ANALYSIS

The HC18 dataset contains 999 training images with ground truth annotations and 335 test images. Each image has an associated pixel size (mm/pixel) for converting measurements to real-world units.

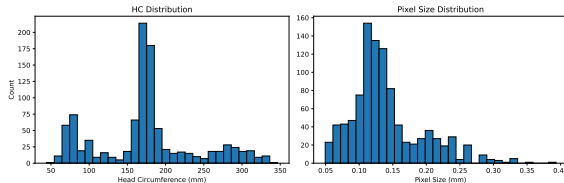


Fig. 1. Distribution of head circumference values and pixel sizes in the training set.

The HC values range from 44mm to 330mm with a mean of 197mm (Fig. 1). Sample images show the ultrasound scans with corresponding ellipse annotations (Fig. 2).

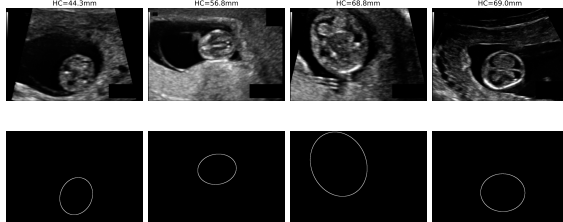


Fig. 2. Sample ultrasound images (top) with annotations (bottom).

## III. METHODS

### A. OpenCV Ellipse Fitting (Baseline)

The annotation images contain white ellipse outlines. We extract the ellipse using contour detection and fit an ellipse with `cv2.fitEllipse()`. The circumference is calculated using Ramanujan's approximation:

$$C \approx \pi(a+b) \left( 1 + \frac{3h}{10 + \sqrt{4-3h}} \right) \quad (1)$$

where  $h = \frac{(a-b)^2}{(a+b)^2}$  and  $a, b$  are semi-major and semi-minor axes.

### B. Random Forest Regression

Features extracted from annotations include ellipse dimensions (width, height), aspect ratio, contour area, perimeter, center position, and pixel intensity statistics (mean, std). A Random Forest regressor with 100 trees predicts HC directly from these 10 features.

### C. U-Net Segmentation

We implement a 3-level U-Net architecture for semantic segmentation of the fetal head region. The encoder uses convolutional blocks with channels 32, 64, 128, and 256 (bottleneck). Each block consists of two  $3 \times 3$  convolutions with batch normalization and ReLU activation. The decoder uses transposed convolutions for upsampling with skip connections from the encoder.

We compare three loss functions:

- **MSE Loss:** Mean squared error between predicted and ground truth masks.
- **BCE Loss:** Binary cross-entropy for pixel-wise classification.
- **BCE + IoU Loss:** Combination of BCE and Intersection-over-Union loss with  $\alpha = 0.5$ :

$$\mathcal{L} = 0.5 \cdot \mathcal{L}_{BCE} + 0.5 \cdot \mathcal{L}_{IoU} \quad (2)$$

After segmentation, HC is computed from predicted masks using ellipse fitting and Ramanujan's formula.

## IV. RESULTS

Table I summarizes the performance of all methods on the validation set (20% of training data).

The OpenCV baseline achieves the lowest MAE (0.71mm) because it directly measures the annotated ellipse, closely matching the ground truth calculation. Random Forest regression (Fig. 3) achieves 4.18mm MAE, showing that learned features can approximate HC measurement with reasonable accuracy.

TABLE I  
MODEL PERFORMANCE COMPARISON

Method	MAE (mm)	IoU
OpenCV Ellipse Fitting	0.71	—
Random Forest Regression	4.18	—
U-Net (MSE Loss)	—	0.00
U-Net (BCE Loss)	—	0.00
U-Net (BCE + IoU Loss)	62.16	0.14

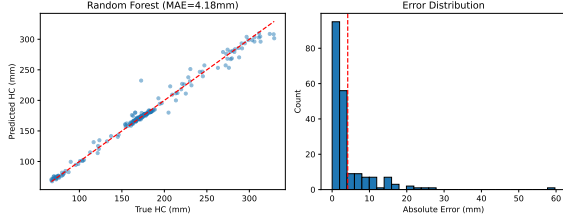


Fig. 3. Random Forest: predicted vs true HC and error distribution.

#### A. U-Net Segmentation Results

Fig. 4 shows the training and validation loss curves for all three loss functions. The MSE and BCE losses converge quickly but fail to produce meaningful segmentations, as shown by their IoU scores of 0. This is because these losses do not directly optimize for overlap between predicted and ground truth regions.

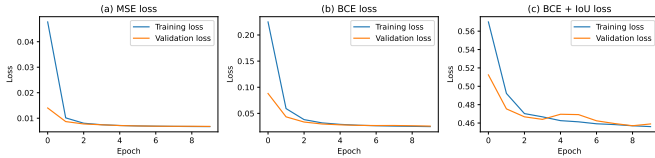


Fig. 4. Training loss curves for U-Net with (a) MSE, (b) BCE, and (c) BCE+IoU loss functions.

Only the BCE+IoU loss produces valid segmentation masks that can be used for HC measurement. Fig. 5 shows sample predictions comparing all three loss functions. The BCE+IoU model learns to segment the fetal head region, though with room for improvement.

Fig. 6 shows the scatter plot of predicted vs true HC values for the BCE+IoU model.

#### V. DISCUSSION

The choice of loss function significantly impacts U-Net segmentation quality. MSE and BCE losses optimize pixel-wise accuracy but can converge to trivial solutions (e.g., predicting all zeros) when class imbalance is severe. The IoU loss directly optimizes for region overlap, forcing the model to learn meaningful segmentation masks.

The relatively high MAE (62.16mm) of the BCE+IoU model suggests that while the model learns to identify the head region, the predicted masks are not precise enough for accurate circumference calculation. This could be improved by:

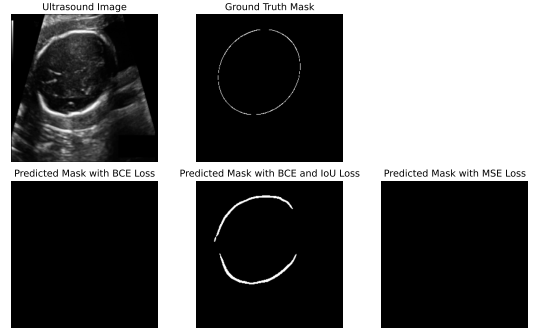


Fig. 5. Sample predictions: ultrasound image, ground truth mask, and predicted masks from BCE, BCE+IoU, and MSE loss models.

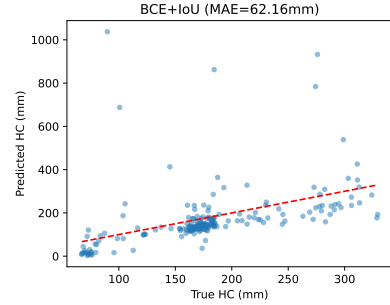


Fig. 6. Predicted vs true HC for U-Net with BCE+IoU loss.

- Training for more epochs
- Using data augmentation
- Increasing model capacity
- Adding Dice loss for better boundary delineation

#### VI. CONCLUSION

This study demonstrates that classical image processing with ellipse fitting provides highly accurate HC measurement (0.71mm MAE) when annotations are available. Random Forest regression offers a machine learning alternative with 4.18mm MAE. For deep learning segmentation, the choice of loss function is critical—BCE+IoU loss is necessary to learn meaningful segmentations, while pure MSE or BCE losses fail to produce valid masks. Future work should focus on improving U-Net segmentation accuracy through longer training, data augmentation, and architecture refinements.