Fetal Head Circumference Measurement from Ultrasound (HC18)
ResNet-18 Regression with MAE Evaluation

Nguyen The Cuong
22BA13058
Machine Learning in Medicine 2026
University of Science and Technology of Hanoi (USTH)
Email: cuongnt.22ba13058@usth.edu.vn

***Abstract*—Head circumference (HC) is an important fetal biometric used in prenatal ultrasound to assess fetal growth and estimate gestational age. This report studies automatic HC measurement using the HC18 dataset of 2D fetal head ultrasound images. We build a single deep learning regression model based on ResNet-18 to predict head circumference in millimeters. The model uses both the ultrasound image and the provided pixel size (mm/pixel) to handle scale differences across acquisitions. We evaluate the approach using mean absolute error (MAE) and perform basic hyperparameter experiments to improve performance.**

***Index Terms*—Ultrasound, fetal biometry, head circumference, HC18, regression, ResNet-18, MAE.**

I. Introduction

Fetal head circumference (HC) is a common measurement in pregnancy ultrasound because it helps doctors monitor fetal growth and make clinical decisions. Normally, HC is measured by choosing the correct head view and fitting an ellipse along the skull boundary. However, this manual process takes time and the results may differ between operators.

In this lab, we build one machine learning/deep learning model to predict fetal head circumference directly from ultrasound images using a regression approach. We use the HC18 dataset and evaluate the model with mean absolute error (MAE) in millimeters. Our method is based on a ResNet-18 network adapted for grayscale ultrasound images. In addition, we include the pixel size (mm/pixel) as an input so the model can handle different image scales and produce HC values in real units.

II. Dataset (HC18)

We use the HC18 dataset, which includes 2D grayscale ultrasound images captured on the standard fetal head plane. For each image, the dataset provides the pixel size in mm/pixel, which reflects the physical scale of the scan. In the training set, ground-truth head circumference (HC) values in millimeters are also available, together with an annotation mask that represents an ellipse along the head boundary.

A. Data organization

The dataset is organized as:

- `training_set`: images and ellipse masks
- `test_set`: images (no HC labels)
- `training_set_pixel_size_and_HC.csv`: filename, pixel size (mm/pixel), head circumference (mm)
- `test_set_pixel_size.csv`: filename, pixel size (mm/pixel)

B. Exploration

We compute descriptive statistics from the training CSV:

- Number of training images: 999
- Number of test images: 333
- HC(mm): min=44.3000, max=346.4000, mean=174.3831, std=65.2821
- Pixel size(mm/pixel): min=0.049415, max=0.393280, mean=0.139846, std=0.053005

C. Qualitative examples

Figure 1 shows sample ultrasound images with the provided ellipse annotation overlaid. Figure 2 shows the distribution of HC values in the training set.

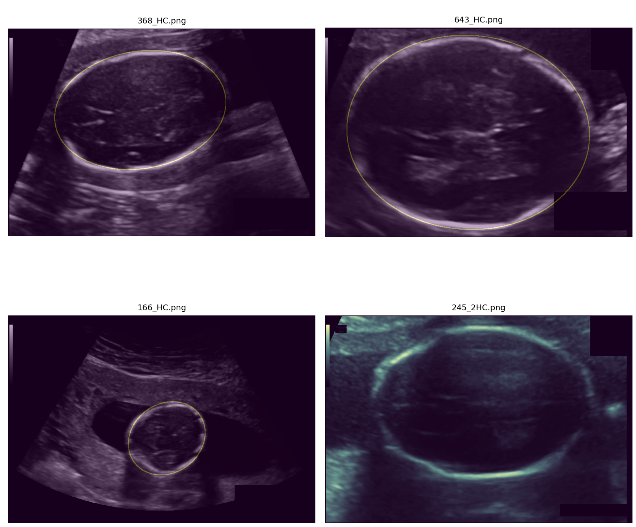

Fig. 1: Example training images with ellipse annotation overlay.

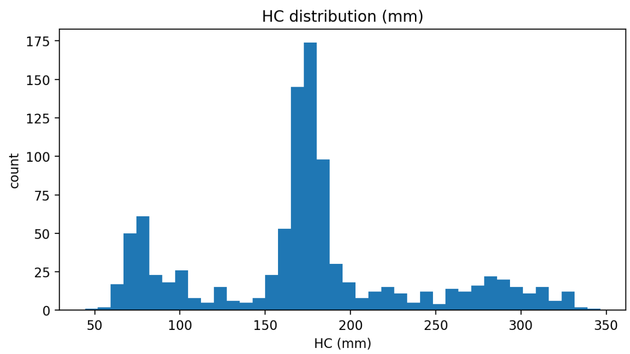

Fig. 2: Distribution of head circumference (mm) in the training set.

III. Methods

A. Problem formulation

Given an ultrasound image I and its pixel size s (mm/pixel), the goal is to predict head circumference $\hat{y}$ in millimeters. We train using mean absolute error (MAE), equivalent to L1 loss:

$$\mathcal{L} = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - y_i| \, .$$

B. Preprocessing

Ultrasound frames often include black borders and non-informative areas outside the scanning region. We apply:

- **Auto-cropping:** remove near-black borders by cropping the non-black region
- **Resizing:** resize to a fixed input resolution (e.g., 256×256)
- **Normalization:** scale intensities to $[0, 1]$

C. Model architecture (ResNet-18 regression)

We use ResNet-18 as the main backbone and adjust the first convolution layer so it can take one-channel (grayscale) ultrasound images. After the backbone, global average pooling is used to extract a compact feature vector that summarizes the image. Because HC is measured in millimeters and depends on image scale, we also include the pixel size s as additional input. We concatenate s with the image feature vector and feed the combined representation into a small MLP regression head to predict HC in mm.

D. Training details

We split the training set into train/validation subsets using an 80/20 split with a fixed random seed. We train using Adam and monitor validation MAE.

- Input size: **2**56
- Batch size: **3**2
- Learning rate: **0**.0003
- Weight decay: **0**.0001
- Epochs: **2**0
- Augmentation (optional): random flip and small rotations

E. Evaluation protocol

We report MAE on the internal validation split. We plot both training MAE and validation MAE across epochs to diagnose underfitting/overfitting.

IV. Results

A. Training and validation MAE

Figure 3 shows the MAE curves for training and validation sets across epochs. The best validation MAE achieved by our ResNet-18 regression model is 11.315 mm at epoch 18.

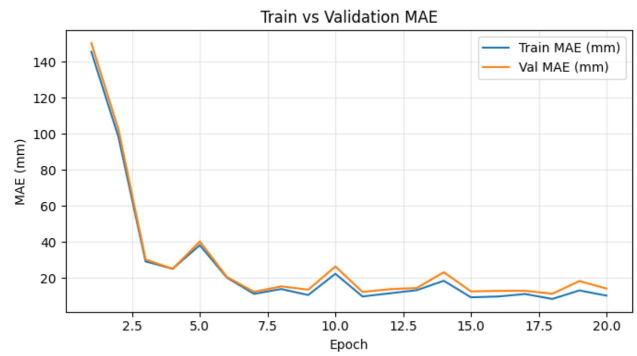

Fig. 3: Train vs. validation MAE (mm) across epochs.

B. Hyperparameter experiments

Table I summarizes a small hyperparameter study. We vary learning rate, input size, and weight decay and report validation MAE.

Run	Input size	LR	Weight decay	Val MAE (mm)
Baseline	256	3e-4	1e-4	11.0577
A1	224	3e-4	1e-4	11.3004
A2	256	1e-4	1e-4	11.8807
A3	256	3e-4	0	12.3811

TABLE I: Validation MAE for different hyperparameter settings.

C. Comparison to the benchmark/leaderboard

The HC18 benchmark often evaluates methods that first predict the head boundary and then compute HC using segmentation and ellipse fitting. In our work, we predict HC directly with a regression model, so the results are not perfectly comparable to those benchmark methods. However, MAE in millimeters is still a useful and clear metric. Therefore, we report the MAE on our validation split as a practical estimate of model performance.

V. Discussion

From our experiments, we find that accuracy depends a lot on image quality and how clearly the skull boundary can be seen. Some images contain strong speckle noise, low contrast around the skull edge, or only part of the head, which makes prediction more difficult. In addition, different acquisition settings and zoom levels can change the scale of the image.

One important design choice in our model is to include the pixel size (mm/pixel) as an input feature. If pixel size is not provided, the model must guess the real-world scale only from the image appearance, and this can be unreliable when zoom changes. By concatenating pixel size with CNN features, the model can better predict HC in real units (mm).

A limitation of our regression approach is that it is less interpretable than segmentation-based pipelines. With segmentation and ellipse fitting, we can visualize the predicted boundary, which is easier to check and may generalize better in some situations. In the future, performance could be improved by using stronger but anatomically safe augmentations, applying more robust cropping of the ultrasound fan region, and using multi-task learning to predict both HC and a coarse boundary representation.

VI. Conclusion and Future Work

In this report, we presented a deep learning regression method for predicting fetal head circumference from 2D ultrasound images using the HC18 dataset. We implemented a ResNet-18-based regressor that takes both the ultrasound image and the pixel size (mm/pixel) as input, and we trained the model with MAE loss. The best validation MAE achieved in our experiments is **11.315 mm**. Future work may focus on segmentation-based measurement pipelines, improved preprocessing of the scanning region, and ensembling methods to further reduce error.