

Fetal Head Circumference Measurement from Ultrasound Images using Deep Learning

Nghia Hoang

January 2026

1 Introduction

Ultrasound imaging is one of the most prevalent non-invasive methods for evaluating the fetal development of a fetus during pregnancy. While there are a variety of different biometric measurements, fetal head circumference (HC) is one of the most critical biometric parameters that clinicians use to estimate gestational age and to identify potentially abnormal growth of the fetus. Accurate measurement of fetal head circumference therefore plays a major role in the prenatal screening and diagnosis of the fetus.

In the past, fetal HC has primarily been measured manually by sonographers by outlining an ellipse around the fetal head within ultrasound images; this process can be cumbersome and highly variable between different sonographers, causing inconsistencies and making accuracy difficult to achieve. Subsequently, automatic methods using a combination of machine learning and deep learning techniques have been attracting increased interest from researchers conducting medical image analysis.

In our project, we explore estimating fetal head circumference from ultrasound images using the HC18 dataset. We intend to train a single deep learning regression model that takes ultrasound images as input and produces a direct estimate of the fetal head circumference in millimetres (mm). The performance of the proposed model will be evaluated through the Mean Absolute Error (MAE) statistic, which is the standard metric used for this task.

2 Dataset Description

The dataset used in this project is the HC18 fetal ultrasound dataset. It consists of ultrasound images of fetal heads acquired under clinical conditions. The dataset is divided into two subsets:

- A training set containing ultrasound images and corresponding annotation masks.
- A test set containing ultrasound images without ground-truth labels.

All training images will have a corresponding binary annotation mask file that depicts an ellipse fitted around the fetal head/crest. These masks will be processed after the epoch and do not count as supervision in training (but are used in calculating this metric).

For each mask, the largest contour will be extracted and then fitted using the `fitEllipse` function available within OpenCV. The estimation of the HC (Head Circumference) in pixel units is carried out using the Ramanujan formula for the perimeter of an ellipse. The measurements will then be converted from pixels to millimeters using the pixel spacing (millimeters per pixel) information provided in a CSV .

Once limiting invalid or corrupt samples, there will ultimately be 999 training images included in creating the model and validate; the sampled head circumferences fall within a range of approximately 45 to 350 mm and average around 175mm.

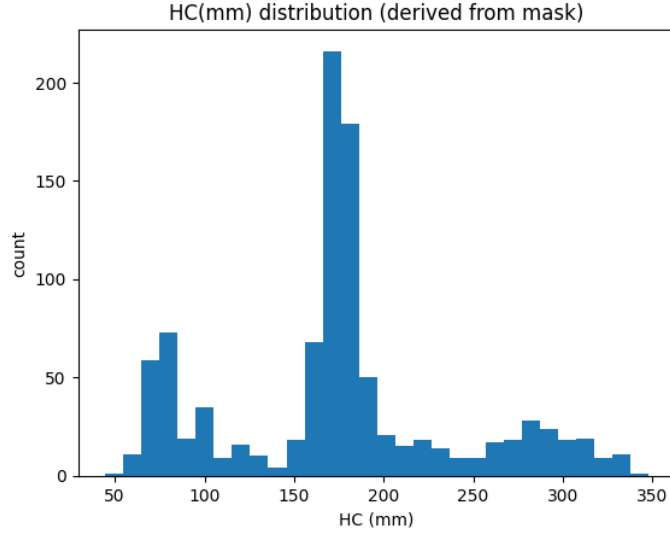


Figure 1: HC(mm) distribution (derived from mask)

3 Methodology

3.1 Preprocessing

To become compatible with CNN architectures, each ultrasound image must have a consistent resolution of 224×224 by having their dimensions resized to pixels. All original grayscale images are converted to three-channel RGB format before being loaded for use in an ImageNet-pretrained model.

In addition to using data augmentation techniques during training (such as

random horizontal flips and small amounts of random rotation) to help produce better generalizations, all images are normalized with respect to the mean and standard deviation of the ImageNet dataset.

3.2 Model Selection

Because this task is to predict a continuous value (numerical) based on some image data; therefore, it is reasonable to use a convolutional neural network, designed specifically for regression, to accomplish this goal. In this research study, the ResNet18 architecture pre-trained on ImageNet is selected as the backbone architecture for the neural network. This neural network architecture provides an acceptable balance between representational capability and the computational efficiency of its computations.

To take the original classification layer off of this network and to create a new layer that will produce a single scalar output (i.e., the predicted fetal head circumference in millimeters), a single fully connected linear layer will replace the original classification layer.

3.3 Training Setup

The training of the model utilizes the Adam optimization algorithm and utilizes a learning rate of 1×10^{-4} that is unspecified. The loss function used during training is the Mean Absolute Error (or also referred to as L1 loss), which is consistent with the evaluation metric.

Data for training and validation purposes is also split 80/20 between training and validation. Training occurs over 20 epochs with the best validation MAE yielding the best-performing model.

4 Experimental Results

Over training, we saw a steady decrease in the validation MAE with each additional epoch, which suggests that stable learning is taking place and that the model is being optimally trained. The best validation performance of the model is shown below:

- Best Validation MAE: 139.64 mm

This result reinforces that the model can learn significant visual features concerning the shape and size of a fetal head using ultrasound images. However, the absolute error level is still much higher than leading methods.

5 Discussion

There is a high MAE because of the unpredictability in estimating fetal head circumference from ultrasound images. Noise, low waveform contrast, various

effort poses, and variability based on how a fetus was acquired all contribute to this unpredictability. The create algorithm used convolutional neural networking for regression purposes only on the actual head circumference labelling; and the segmentation masks were used offline to attain the labels.

There are multiple reasons for the relatively high error rates in the study. These include: 1) the small size of the training sets, 2) the lack of any kind of definition or shape definition during the training sessions, 3) the methods selected would generally classify as a "single-stage" regression analysis method vs. utilizing any multi-task learning approaches. It is possible that multi-task learning or methods that used regression for direct ellipsoidal parameter estimation could effectively establish some advancements in overall accuracy of the measurements.

6 Conclusion

Using the HC18 dataset, a deep learning method was created to generate estimates of fetal head circumference from ultrasound images. A single ResNet18 regression model was trained on predicting values for fetal head circumferences in millimeters. A best validated MAE of 139.64mm was achieved. While this system has not yet achieved state-of-the-art results, it demonstrates that using CNNs may be an appropriate technique for automatically measuring fetal biometric measurements and provides a good baseline from which to identify further potential improvement.