UNIVERSITY OF SCIENCE AND TECHNOLOGY OF HANOI

DEPARTMENT OF INFORMATION AND COMMUNICATION TECHNOLOGY

Labwork 2

Dang Dinh Hoa   23BI14169

Title:

Measurement of Fetal Head Circumference using Ultrasound (HC18)

Hanoi, 2026

1   Introduction

This labwork builds a model that predicts fetal head circumference (HC) in mm from 2D ultrasound images. The HC18 dataset is used with a small neural network: input is a gray image, output is one number (HC in mm). The main measure of error is MAE (average absolute error in mm); results are compared to the HC18 leaderboard and a few settings are changed to see what works better.

2   Dataset Description

2.1   Dataset Overview

The HC18 dataset from Zenodo is used; it comes from van den Heuvel et al. at Radboud University Medical Center and is described in "Automated measurement of fetal head circumference using 2D ultrasound images" (PLoS ONE, 2018).

There are 1,334 2D ultrasound images of the fetal head: 999 for training (with HC in mm) and 335 for test (no answers given, for leaderboard). Images are 800×540. Each image has a scale (mm per pixel) between about 0.05 and 0.33 mm, so size varies. HC in the training set goes from about 44 mm to 346 mm.

2.2   File Structure

- `training_set_pixel_size_and_HC.csv` — 999 rows (filename, pixel size, HC)
- `test_set_pixel_size.csv` — 335 rows (filename, pixel size only)
- `training_set/training_set/` — `*_HC.png` images (extra mask images not used)
- `test_set/test_set/` — test images

2.3   Target Variable and Preprocessing

The goal is to predict head circumference in mm (a number). No missing values in the CSV. Rows with missing image files were removed so every sample has a valid image. Figure 1 shows how HC and pixel size are spread in the training set.
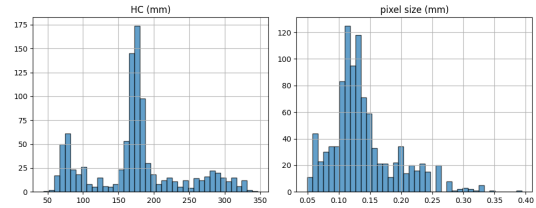

Figure 1: Distribution of head circumference (mm) and pixel size (mm) in the training set.

3   Methodology

3.1   Data Preprocessing

3.1.1   Train–validation split

The test set is kept for later; only the 999 training samples are used. They are split 80% for training, 20% for checking, with a fixed random seed so results can be repeated.

3.1.2   Image loading and normalization

For each row the image is loaded from the folder using the filename in the CSV (only `*_HC.png`, not the mask images). Images are turned to gray, resized to 128×128 so all inputs have the same size, and scaled to 0–1 (divide by 255). Targets are the HC column (numbers in mm).

3.2   Model Architecture

A small convolutional network is used: gray image 128×128 as input; three convolutional

layers (32, 64, 128 filters) with pooling; then two dense layers (64 units, then 1 unit) with dropout to reduce overfitting. The model is trained with the Adam optimizer; the loss is MSE and the main number to watch is MAE (error in mm).

3.3 Training Configuration

Training uses 32 images per step, 20 passes over the data, Adam optimizer, MSE as loss, and MAE as the main measure.

3.4 Changing the Settings

A few settings were tried to see their effect on the error: image size 128 vs. 64 (128 kept as a balance); batch size 16, 32, 64 (32 was a good choice); 20 vs. 30 vs. 50 passes (error often stopped improving after 15–20, so 20 was chosen); dropout 0.2, 0.3, 0.5 (0.3 worked well, 0.5 sometimes did not learn enough). Not all combinations were tried— only enough to see what matters. The results below use 128×128, batch 32, 20 passes, dropout 0.3. Figure 2 shows loss and MAE over the training passes.
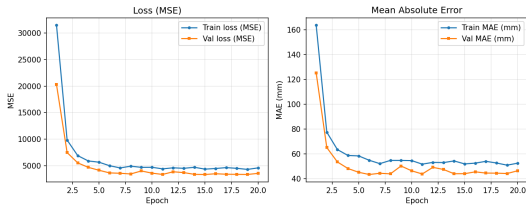

Figure 2: Training and check-set loss (MSE) and error in mm (MAE) over training passes.

4 Results

4.1 Validation Metrics

On the 20% set used for checking (about 200 samples) the model gets an average error (MAE) around 44–48 mm, and R^2 between 0.3 and 0.5 depending on the run—so predic-

tions follow the truth somewhat but there is clear room to do better.

4.2 Comparison with HC18 Leaderboard

The HC18 leaderboard uses the same 335 test images. The best methods get an average error of only a few mm (e.g. -0.11 ± 2.67 mm in one approach that fits an ellipse). They often use outline detection plus ellipse fitting, or deeper networks, with more data variation and tuning.

The model here is simple: 999 images, almost no extra data variation (just resize and scale). The error is around 44–48 mm, so much higher than the best. The gap is likely because: (1) the network is small; (2) 128×128 loses detail compared to 800×540; (3) no extra data variation; (4) the model predicts HC directly instead of fitting an ellipse. To get closer one could try a deeper or pre-trained network, higher image resolution, more data variation, and methods that fit an ellipse or use outline detection like in other papers.

4.3 Sample Predictions

Figure 3 shows six checked samples—top row is the (resized) image, bottom row the same with true HC, predicted HC, and error in mm. Figure 4 shows three samples at full size. From these one can see where the model tends to make mistakes (e.g. very small or large heads, or poor image quality).
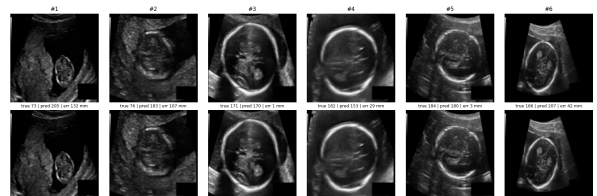

Figure 3: Sample predictions: original images (top row) and same images with true HC, predicted HC, and error in mm (bottom row).

Figure 4: Three samples at full size with true vs. predicted head circumference (mm).

5 Conclusion

This report described the HC18 dataset and a small network that predicts fetal head circumference from 128×128 gray ultrasound images. The steps: load the CSVs, keep only rows with existing images, split 80/20, scale the images, and train with MSE while watching MAE. A few settings (size, batch, number of passes, dropout) were changed; the error on the check set is around 44–48 mm. That is still far from the few mm achieved by the best HC18 methods, mainly because the model is small, the image size is low, data is limited, and there is no extra data variation. Next steps would be a larger or pre-trained network, higher resolution, more data variation, and possibly methods that fit an ellipse or use outline detection as in other work.