Department of Information and Communication Technology

Machine Learning in Medicine

Report 2

ID - Student Name:
22BI13220 - Nguyen Minh Khoi - Data Science

Lecturer:
Prof. Tran Giang Son

Academic year: 2022 - 2025
Hanoi, March 2025

1 Introduction

This report presents Practical 2 of Machine Learning in Medicine, where a publicly available Kaggle dataset is analyzed to understand its structure and features before developing a regression model. This report presents the implementation of a U-Net based deep learning model to perform segmentation on fetal ultrasound images. The segmented masks are used to identify the fetal head boundaries for subsequent head circumference (HC) measurements. To prevent overfitting and optimize training time, early stopping is implemented during model training.

2 Dataset

The dataset comprises a training set with 999 images and a test set with 335 images. Each 2D ultrasound image has a resolution of 800 × 540 pixels, with pixel sizes ranging from 0.052 mm to 0.326 mm. The pixel size information for each image is provided in two CSV files: `training_set_pixel_size_and_HC.csv` for the training set and `test_set_pixel_size.csv` for the test set. Additionally, the training set includes images with manual head circumference (HC) annotations, created by a trained sonographer. The corresponding HC measurements, recorded in millimeters, are available in the `training_set_pixel_size_and_HC.csv` file. Although the training set contains 999 images, the filenames only go up to 805. This discrepancy is due to some ultrasound images being captured during the same echoscopic examination, resulting in nearly identical appearances. These duplicate images are distinguished by an additional identifier in their filenames, placed between "_" and "HC" (e.g., `010_HC.png` and `010_2HC.png`).

The dataset was split into:

- Training set: 80% of the data
- Validation set: 20% of the data

All images and masks were resized to 128x128 pixels and normalized to the [0, 1] range for model input.

3 Model Architecture

The deep learning model developed for this project is based on the U-Net architecture, which is widely used for biomedical image segmentation tasks due to its ability to capture both local and contextual information. The model consists of an encoder-decoder structure with skip connections that link the corresponding layers in the encoder and decoder paths. In the encoder, the model uses convolutional layers followed by max-pooling operations to progressively reduce the spatial dimensions and extract hierarchical features. Specifically, the encoder includes two convolutional layers with 16 filters in the first block and 32 filters in the second block, each followed by ReLU activations and 'same' padding to preserve spatial resolution. The bottleneck of the network consists of two convolutional layers with 64 filters, which capture the deepest features of the input image.

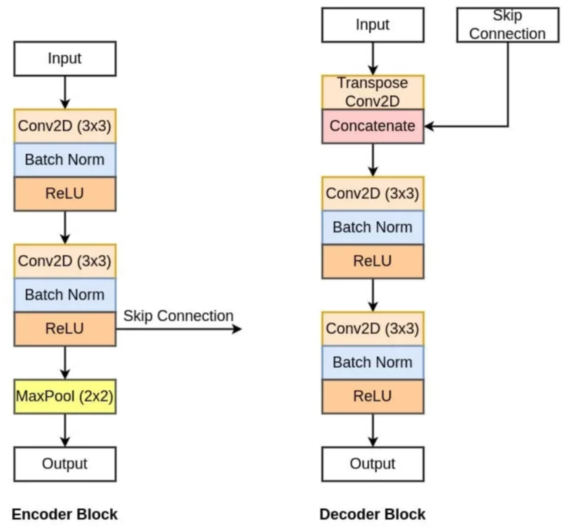

Figure 1: U-net Model

The decoder path reconstructs the spatial resolution of the input using upsampling layers and concatenates the feature maps from the encoder via skip connections. This design helps retain high-resolution features that are crucial for precise segmentation. In the decoder, the model successively applies upsampling, concatenation with encoder features, and convolutional layers with decreasing filter sizes (32 and 16 filters). Finally, the output layer consists of a single convolutional layer with a sigmoid activation function that generates a binary segmentation mask indicating the fetal head region.

The model is compiled with the Adam optimizer, binary crossentropy loss, and accuracy as the evaluation metric. Early stopping is applied during training to prevent overfitting and ensure optimal performance. It monitors the validation loss, and if no improvement is observed for five consecutive epochs, the training process is halted. Additionally, the model restores the weights from the epoch with the lowest validation loss. The network is trained for a maximum of 10 epochs with a batch size of 20, and both the training and validation losses are monitored to evaluate the model's performance throughout the training process.

4 Result

The performance of the U-Net model was evaluated using both training and validation datasets over the course of 10 epochs. The metrics used for evaluation were accuracy and binary cross-entropy loss. The model was trained with early stopping to prevent overfitting and to restore the best weights based on validation loss.

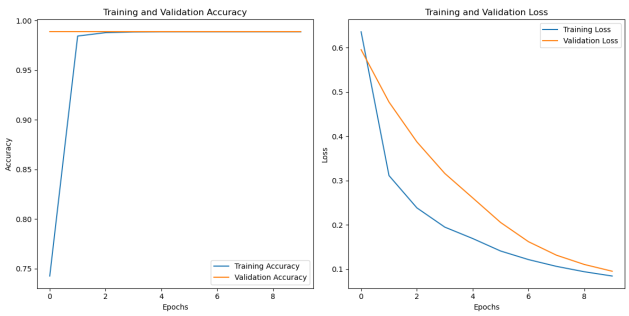

Figure 2: Model Result

Training and Validation Accuracy As illustrated in the left plot, the training accuracy started at approximately 74% and increased rapidly within the first few epochs. By the second epoch, the training accuracy had already surpassed 97%, and it remained nearly perfect at 99% for the remaining epochs. The validation accuracy followed a similar trend, remaining consistently high at around 99% throughout the entire training process. This indicates that the model was able to generalize well to unseen data without significant overfitting.

Training and Validation Loss The loss curves, shown on the right plot, demonstrate a steady decline for both training and validation loss over time. The training loss decreased sharply from around 0.63 in the first epoch to below 0.1 by the final epoch. The validation loss also showed a consistent downward trend, reducing from 0.59 to approximately 0.08. The smooth and parallel descent of both loss curves indicates that the model is learning effectively and is not experiencing overfitting or underfitting.

Summary of Results Final Training Accuracy: 99% Final Validation Accuracy: 99% Final Training Loss: 0.07 Final Validation Loss: 0.08 These results suggest that the U-Net model, with batch normalization and transpose convolutions, was successful in learning accurate segmentation of the fetal head in ultrasound images. The model demonstrates both high accuracy and low loss, reflecting its ability to correctly identify and segment the target regions.