

Few-shot classification of ultrasound breast cancer images using meta-learning algorithms

1. Introduction

This report summarizes the results obtained from training a baseline deep learning model following the methodology inspired by the referenced Springer paper on few-shot breast ultrasound lesion classification. The aim was to reproduce a basic experimental setup, observe training behavior, and identify potential improvements to stabilize and enhance performance.

2. Experimental Setup

A 30-epoch training run was conducted using the notebook provided. The model was trained on the available dataset without advanced augmentation, learning-rate scheduling, or meta-learning techniques. Only training loss and training accuracy were recorded per epoch.

3. Training Results

Training performance showed significant fluctuations across epochs, indicating unstable optimization. Accuracy ranged from **6.67%** to **66.67%**, with no consistent improvement trend. Loss values also oscillated, including several sharp spikes.

Table 1 — Training Metrics (Epoch 1–30)

Metric	Value
Best Training Accuracy	66.67%
Lowest Accuracy	6.67%
Final Epoch Accuracy	26.67%
Final Epoch Loss	1.1067

Average Accuracy (30 epochs)	$\approx 42.67\%$
------------------------------	-------------------------------------

These results suggest inconsistent learning and potential issues with model stability, data distribution, or training configuration.

4. Analysis

The instability observed across the 30 epochs indicates that the model is not converging reliably. Large accuracy swings and loss spikes are commonly associated with:

- Inappropriate or excessively high learning rate
- Limited data or lack of augmentation
- Class imbalance affecting batch composition
- Missing regularization strategies
- Lack of validation monitoring and early stopping
- Possible issues inside the training loop (e.g., gradient accumulation, mode switching)

Compared to findings in the referenced paper, which emphasize the effectiveness of transfer learning and meta-learning approaches (e.g., ResNet backbones and prototypical networks), the baseline model in this experiment lacks several stabilizing components.

5. Recommendations for Improvement

To achieve meaningful and stable performance, the following improvements are recommended:

Model Stability

- Use a lower learning rate and introduce a learning-rate scheduler.
- Apply gradient clipping to prevent exploding gradients.

- Ensure proper training loop structure with clean gradients each step.

Data & Regularization

- Introduce targeted data augmentation (rotations, flips, brightness variation).
- Address class imbalance through class-weighted loss or oversampling.
- Add dropout or weight decay if not already used.

Training Process

- Create a separate validation split and track validation loss/accuracy.
- Add early stopping and checkpointing based on validation performance.
- Monitor per-class metrics rather than only overall accuracy.

Architectural Enhancements

- Use transfer learning with a pretrained ResNet backbone, as supported by literature.
 - Consider adopting meta-learning techniques such as Prototypical Networks for few-shot settings.
-

6. Conclusion

The current baseline experiment demonstrates unstable learning and insufficient predictive performance, indicating that the model and training setup require improvements. Incorporating validation monitoring, more robust optimization strategies, appropriate data augmentation, and stronger feature extractors (e.g., ResNet) is expected to significantly enhance stability and accuracy. These refinements align with best practices and with the methodological strengths highlighted in the referenced research.