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 **13.67% to 76.67%**, 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	**76.67%**
Lowest Accuracy	**13.67%**
Final Epoch Accuracy	**26.67%**
Final Epoch Loss	**1.1067**

Average Accuracy (30 epochs)	≈ **44.67%**

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

- **Model Stability**
  - • Use a lower learning rate and, if needed, add a simple learning-rate scheduler.
  - • Apply gradient clipping to prevent sudden jumps during training.
  - • Make sure gradients are cleared properly before each backpropagation step.
- **Data & Regularization**
  - • Add basic data augmentation such as flips, rotations, and small brightness changes.
  - • Handle class imbalance using simple oversampling.
  - • Use light regularization such as dropout or basic weight decay.

- **Training Process**
  - Train with multiple episodes per epoch (e.g., 50–100 episodes) and calculate the average accuracy for more stable results.
  - Create a small validation split and monitor validation accuracy and loss.
  - Track per-class accuracy instead of only overall results to see class-wise behavior.
- **Architectural Enhancements**
  - Normalize embeddings and also experiment with cosine distance instead of only Euclidean.
  - Increase the number of k-shot or query samples slightly if the dataset size allows.

---

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.