Task 12 – Controlled Optimizer Experiment Conclusion Report

Experimental Setup Description

Controlled experiments were conducted using SGD, Adam, and RMSProp under identical conditions: same dataset splits, same 5×5-kernel CNN architecture, same random seed, same epoch budget, same callbacks, same class-weighting strategy, and identical evaluation protocol. Each optimizer was evaluated in two configurations:

1. Baseline model (no explicit regularization)
2. Regularized model (Batch Normalization + Dropout)

The objective was to compare convergence speed, training stability, and generalization performance.

Convergence Speed

Among the three optimizers, Adam converged the fastest, reaching high training and validation accuracy within the first few epochs in the baseline configuration. RMSProp also showed rapid early convergence but slightly slower than Adam. In contrast, SGD converged more gradually, requiring more epochs to reach comparable validation performance.

Training Stability

SGD produced the most stable training dynamics across both baseline and regularized models. Training and validation losses decreased smoothly, and no abrupt validation collapse was observed. In contrast, Adam and RMSProp exhibited instability when combined with Batch Normalization and Dropout. Although they converged quickly, their validation loss began to increase sharply after a few epochs, triggering learning-rate reductions and early stopping. This behavior indicates sensitivity to regularization and class-weighted gradients rather than implementation errors.

Generalization Performance

In terms of final evaluation metrics on the test set, SGD with regularization achieved the best generalization, yielding the highest Micro and Macro F1-scores (Micro F1 ≈ 0.63, Macro F1 ≈ 0.60) and strong ROC–AUC values.

While Adam and RMSProp achieved good validation accuracy in their baseline runs, their regularized versions generalized poorly, producing highly imbalanced predictions and low F1-scores due to early validation degradation.

Trade-offs Observed

The experiments reveal a clear speed–robustness trade-off. Adam and RMSProp offer fast convergence but are more sensitive to aggressive regularization and class imbalance, making them less robust in this multi-label audio classification task. SGD, although slower, demonstrated greater stability and consistent generalization, particularly in the presence of regularization.

Decision Based on Experimental Evidence

Based on the observed convergence behavior, stability characteristics, and final evaluation metrics, SGD with momentum and regularization was retained as the optimizer for subsequent experiments. This decision was made because it demonstrated consistent and reliable generalization under the given experimental environment, while Adam and RMSProp showed sensitivity to regularization and class imbalance in this specific setup.

Conclusion

Among the evaluated optimizers, Adam converged the fastest, followed by RMSProp, while SGD exhibited slower but more consistent convergence. SGD produced the most stable training and validation curves across both baseline and regularized settings. In contrast, Adam and RMSProp showed validation instability when combined with Batch Normalization and Dropout, leading to early stopping and poor test-set generalization. The final evaluation metrics indicate that SGD achieved the best Micro and Macro F1-scores, demonstrating superior generalization. These results highlight a trade-off between convergence speed and robustness, with SGD being the most reliable optimizer for the given multi-label audio classification task.