How does the scaling of synthetic pretraining data volume influence the convergence rate and final validation

Assignee Research

June 10, 2026

Abstract

Evidential deep learning, built upon belief theory and subjective logic, offers a principled and computationally efficient way to turn a deterministic neural network uncertainty-aware. The resultant evidential models can quantify fine-grained uncertainty using the learned evidence. To ensure theoretically sound evidential models, the evidence needs to be non-negative, which requires special activation functions for model training and inference. This constraint often leads to inferior predictive performance compared to standard softmax models, making it challenging to extend them to many large-

1 Introduction

This paper examines: Learn to Accumulate Evidence from All Training Samples: Theory and Practice. Research question: How does the scaling of synthetic pretraining data volume influence the convergence rate and final validation loss of tabular foundation models compared to real-data baselines?.

2 Methodology

Systematic literature search across multiple databases yielded 13 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 3.7/10.

3 Results

13 papers retrieved. 16 claims extracted; 1 independently verified. Quality review score: 3.7/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

5 Extracted Claims

Claim	Verified	Confidence
The study utilizes the MNIST, Cifar10, and Cifar100 datasets for standard supervised classification tasks.	×	0.03
The study utilizes the mini-ImageNet dataset for few-shot classification tasks.	×	0.02
The LeNet model is employed for experiments on the MNIST dataset.	×	0.02
The ResNet18 model is employed for experiments on the Cifar10 and Cifar100 datasets.	×	0.03
The ResNet12 model is employed for experiments on the mini-ImageNet dataset.	×	0.02
Strong regularization in existing evidential models pushes many training samples into or close to zero-evidence regions,	✓	0.19
The proposed RED model achieves a test accuracy of 99.10% $\pm$ 0.02 on MNIST using the exp activation function.	×	0.02
The proposed RED model achieves a test accuracy of 95.24% $\pm$ 0.06 on Cifar10 using the exp activation function.	×	0.02
The proposed RED model achieves a test accuracy of 76.43% $\pm$ 0.21 on Cifar100 using the exp activation function.	×	0.03
Using the ReLU activation function, the RED model yields 0 samples with evidence E $>$ 1.0 out of 50,000 tested samples.	×	0.06
Using the exp activation function, the RED model yields 24,846 samples with evidence E $>$ 1.0 out of 50,000 tested sample	×	0.05
With 10% of training data, the RED model using the exp activation function achieves 97.27% accuracy.	×	0.03
With 100% of training data, the RED model using the exp activation function achieves 76.43% accuracy.	×	0.03
When trained with Log loss (23) and the exp activation function, the RED model achieves an accuracy of 76.43% $\pm$ 0.2.	×	0.01
The RED model exhibits an In-Distribution Vacuity of 0.2729 using the exp activation function.	×	0.02
The RED model exhibits an Out-of-Distribution Vacuity of 0.7552 on the SVHN dataset using the exp activation function.	×	0.02

References

- http://arxiv.org/abs/2512.03307v1
- http://arxiv.org/abs/2306.11113v2
- http://arxiv.org/abs/2601.04110v2