Scalability of Vision Transformers and CNNs in Few-Shot Medical Image Classification

Assignee Research

June 7, 2026

Abstract

This report synthesises findings from 15 peer-reviewed papers addressing the following research question: How does the scalability of Vision Transformer depth and width compare to CNN architectures in few-shot medical image classification tasks, measured by AUC-ROC across multiple ECG datasets. 10 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 3.7/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Cross-dimensional transfer learning in medical image segmentation with deep learning. Research question: How does the scalability of Vision Transformer depth and width compare to CNN architectures in few-shot medical image classification tasks, measured by AUC-ROC across multiple ECG datasets?.

2 Methodology

Systematic literature search across multiple databases yielded 15 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

15 papers retrieved. 10 claims extracted; 0 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 ECOVNet architecture uses an EfficientNet with weights pre-trained on ImageNet.	×	0.05
ECOVNet was proposed by Chowdhury et al. (2021) for COVID-19 detection.	×	0.04
Messaoudi et al. (2021) proposed an asymmetric 3D U-Net architecture for brain tumor segmentation that includes Efficien	×	0.04
In the Messaoudi et al. (2021) architecture, the first layers of the encoder reduce the depth dimension to fit the 2D Ef	×	0.04
Omnia-Net was applied to the CAMUS dataset for 2D echocardiography and achieved a 1st place ranking.	×	0.05
DS-Net was applied to the CHAOS dataset for 3D abdominal imaging and achieved a 3rd place ranking.	×	0.06
DX-Net was applied to the BraTS dataset for 3D multi-modal brain tumors using Dimensional Transfer Learning.	×	0.14
The BEASM-semi model achieved a score of 0.920 with values 2.2 and 6.0 in the O1a vs O2 comparison.	×	0.02
The U-Net 2 model achieved a score of 0.954 with values 1.7 and 6.0 in the second reported metric column.	×	0.04
The SHN model achieved a score of 0.951 in the second reported metric column.	×	0.04

References

- `http://arxiv.org/abs/2307.15872v1`

- http://arxiv.org/abs/2101.11986v3
- http://arxiv.org/abs/2111.10480v6