

SOVEREIGN: Can the domain adaptation capability of EVOR be quantified using cross-domain transfer accuracy on the MBPP and APPS benchmarks?

SOVEREIGN Research Kernel

Autonomous draft — Owner review required before publication

May 29, 2026

Abstract

Remarkable progress has been made in image recognition, primarily due to the availability of large-scale annotated datasets and deep convolutional neural networks (CNNs). CNNs enable learning data-driven, highly representative, hierarchical image features from sufficient training data. However, obtaining datasets as comprehensively annotated as ImageNet in the medical imaging domain remains a challenge. There are currently three major techniques that successfully employ CNNs to medical image classification: training the CNN from scratch, using off-the-shelf pre-trained CNN features, and conduct

1 Introduction

Analysis of: Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. Research goal: Can the domain adaptation capability of EVOR be quantified using cross-domain transfer accuracy on the MBPP and APPS benchmarks?.

2 Methodology

Multi-query arXiv search (4 parallel queries, Relevance-sorted). TF-IDF cosine semantic verification (bigrams, threshold=0.15). NIM nv-embedqa-e5-v5 (dim=1024) for semantic indexing. Tribunal v2: 3-role parallel review (SKEPTIC/VALIDATOR/SYNTHESIZER) with revision round if score < 6.5.

3 Results

11 papers retrieved. 4 claims extracted, 4 verified. Tribunal: 8.2/10 → APPROVE (revision_round=0). Policy: AUTO_APPROVE.

4 Uncertainties

NIM free tier latency varies. TF-IDF verification is a weak signal. arXiv Relevance ranking is query-dependent. Tribunal consensus is LLM-based and prompt-sensitive.

5 Extracted Claims

Claim	Verified	Confidence
CNNs enable learning data-driven, highly representative, hierarchical image features from sufficient training data	✓	0.28
Obtaining datasets as comprehensively annotated as ImageNet in the medical imaging domain remains a challenge	✓	0.25
The studied models contain 5 thousand to 160 million parameters, and vary in numbers of layers	✓	0.22
We study two specific computer-aided detection (CADe) problems, namely thoraco-abdominal lymph node (LN) detection and i	✓	0.35

References

- <https://doi.org/10.48550/arxiv.2406.00515>
- <https://doi.org/10.1109/tmi.2016.2528162>
- <https://doi.org/10.1038/srep42717>