Weekly Meeting with Dr. Hannah

Hasan Shaikh

Quantitative Imaging Research and
Artificial Intelligence Lab (QIRAIL)

Paper Studied

scientific reports

Explore content · About the journal · Publish with us

nature › scientific reports › articles › article

Article | Open access | Published: 24 October 2023

Image based prognosis in head and neck cancer using convolutional neural networks: a case study in reproducibility and optimization

Pedro Mateus, Leroy Volmer, Leonard Wee, Hugo J. W. L. Aerts, Frank Hoebers, Andre Dekker & Inigo Bermejo

Scientific Reports 13, Article number: 18176 (2023) | Cite this article

2162 Accesses | 5 Citations | 1 Altmetric | Metrics

Abstract

In the past decade, there has been a sharp increase in publications describing applications of convolutional neural networks (CNNs) in medical image analysis. However, recent reviews have warned of the lack of reproducibility of most such studies, which has impeded closer examination of the models and, in turn, their implementation in healthcare. On the other hand, the performance of these models is highly dependent on decisions on architecture and image pre-processing. In this work, we assess the reproducibility of three studies that use CNNs for head and neck cancer outcome prediction by attempting to reproduce the published results. In addition, we propose a new network structure and assess the impact of image pre-processing and model selection criteria on performance. We used two publicly available datasets: one with 298 patients for training and validation and another with 137 patients from a different institute for testing. All three studies failed to report elements required to reproduce their results thoroughly, mainly the image pre-processing steps and the random seed. Our model either outperforms or achieves similar performance to the existing models with considerably fewer parameters. We also observed that the pre-processing efforts significantly impact the model's performance and that some model selection criteria may lead to suboptimal models. Although there have been improvements in the reproducibility of deep learning models, our work suggests that wider implementation of reporting standards is required to avoid a reproducibility crisis.

Key Learning from Paper Studied

- **Need for Reporting Standards:** Reproducibility issues due to missing preprocessing details, random seeds, unclear environment configs, and lack of inclusion criteria. Recommends frameworks like CLAIM and TRIPOD.
- **Transparency:** Provides Dockerized pipeline, configs, scripts, and sample cases — exemplary for reproducibility.
- **Concept Shift:** Differences in scanning protocols/patient demographics impact model performance — more diverse data is needed.
- **Clinical Data Integration:** Mixed results but improved performance for locoregional failure prediction.
- **Impact of Preprocessing:** Tumor-centered cropping and HU windowing (e.g., 125/350 HU) improved AUC.

NITK Manuscript Editing

Reviewer Comments:

- Lack of clarity in loss function explanation.
- Basic preprocessing is underspecified.
- Missing experimental details like learning rate schedule, fine-tuning.
- Visuals for performance per fold needed.
- Overfitting concerns: training/validation curves requested.

Technical Upskilling

🐳 **Docker for Machine Learning Development**:

- Objective: Learn containerization for scalable ML pipeline deployment and reproducibility.

SERB Proposal Studied

- **"Developing a Clinically Relevant Fully Automated and Scalable AI-Based System for Tumor and Organ-at-Risk Segmentation Using CT Scan Alone."**