

Scaling ContProto Performance to Large Multilingual Language Models for Cross-Lingual NER

Assignee Research

June 17, 2026

Abstract

Natural language tasks like Named Entity Recognition (NER) in the clinical domain on non-English texts can be very time-consuming and expensive due to the lack of annotated data. Cross-lingual transfer (CLT) is a way to circumvent this issue thanks to the ability of multilingual large language models to be fine-tuned on a specific task in one language and to provide high accuracy for the same task in another language. However, other methods leveraging translation models can be used to perform NER without annotated data in the target language, by either translating the training set or test set.

1 Introduction

This paper examines: Multilingual Clinical NER: Translation or Cross-lingual Transfer?. Research question: Can the performance gains from ContProto be scaled to large multilingual language models like XLM-RoBERTa or mBERT when fine-tuned for cross-lingual NER?.

2 Methodology

Systematic literature search across multiple databases yielded 12 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 7.5/10.

3 Results

12 papers retrieved. 8 claims extracted; 6 independently verified. Quality review score: 7.5/10.

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 MedNERF dataset contains 100 sentences and 406 entities.	✓	0.15
The MedNERF dataset is available at https://huggingface.co/datasets/Posos/MedNERF .	✓	0.19
The GERNERMED test dataset consists of 30 sentences and 119 entities.	✓	0.16
The n2c2 dataset contains 16,656 sentences and 65,495 entities.	×	0.12
XLm-R Base outperforms mBERT on multilingual benchmarks.	×	0.12
XLm-R Large is used to evaluate the impact of model size.	✓	0.15
distilmBERT is a smaller MLLM obtained by distillation of mBERT.	✓	0.19
The translate-train approach involves constructing a translated version of the n2c2 dataset.	✓	0.21

References

- <http://arxiv.org/abs/2303.09306v2>
- <http://arxiv.org/abs/2101.11112v1>
- <http://arxiv.org/abs/2306.04384v1>