

AN IMPROVED MEASURE OF AUTOMATIC SPEECH RECOGNITION PERFORMANCE IN CLINICAL SETTINGS

Ehud Rivlin

UEG Week Copenhagen 2023

Poster

October 15, 2023

A novel metric called Clinical BERTScore (CBERTScore) was developed to evaluate automatic speech recognition in medical settings by combining AI language understanding with medical domain knowledge, achieving 75.4% agreement with clinician preferences compared to 10.5% for traditional Word Error Rate.

KEY POINTS

- CBERTScore combines BERTScore with a medical Knowledge Graph, weighting medical terms more heavily and tuning the clinical term penalty to match clinician preferences from a dataset of 18 preferences on 150 gastroenterology phrases
- CBERTScore achieved 75.4% agreement with clinician preferences, outperforming BERTScore (66.7%), METEOR (53.4%), BLEU (24.6%), and Word Error Rate (10.5%)
- Traditional Word Error Rate treats all transcription mistakes equally and was the worst at determining which transcripts clinicians find more useful, despite being the most common ASR performance metric
- CBERTScore performed equally well on non-medical text, suggesting broader applicability for evaluating transcripts in general settings
- The authors made the clinician preference dataset publicly available and suggest the metric warrants further validation as a primary tool for training and evaluating speech recognition in clinical settings

VIEW THIS CONTENT ON GUTFLIX

<https://gutflix.eu/search/d/2b65b218-743b-11ee-b271-0242ac140004>



AI-generated summary

This summary was generated by an AI large language model based on the content transcript. It is for informational purposes only and should not be considered a substitute for clinical judgment. Always rely on your professional expertise and the full clinical context when making clinical decisions.