Filled Checklist: STARD

1. Title or abstract: Identification as a study of diagnostic accuracy

Yes, this is a study of diagnostic accuracy. The title and abstract are not explicitly mentioned in the provided context, but the content suggests that it is a research article evaluating the diagnostic performance of a deep learning model for radiology report retrieval.

2. Abstract: Structured summary

The abstract is not provided in the given text.

3. Introduction: Scientific and clinical background

There is no introduction or scientific and clinical background section in the provided text. The text appears to be a research article with various sections (e.g., abstract, methods, results, discussion), but it does not include an introductory section that explains the context and significance of the study.

4. Introduction: Study objectives and hypotheses

The study objectives and hypotheses are not explicitly stated in the provided text.

5. Methods: Study design

The study design is not explicitly stated in the provided text.

6. Methods: Participants: Eligibility criteria

There is no mention of participants or eligibility criteria in the provided text.

7. Methods: Participants: Sources and methods of participant selection

The study does not provide specific details about participants or their selection process. It only mentions authors affiliated with various departments and institutions. Therefore, I cannot provide a concise answer to this question as requested.

8. Methods: Participants: Data collection

The participants and data collection methods are not specified in the provided text.

9. Methods: Test methods: Index test details

The study used a deep learning model to adapt transformer-based language models to radiology and compared its performance with another pre-trained model called GTE-Large-En-V1.5. The study also evaluated the model's ability to retrieve relevant radiology reports using a benchmark dataset.

10. Methods: Test methods: Reference standard details

The reference standard is not explicitly mentioned in the provided text.

11. Methods: Test methods: Rationale for choosing reference standard

The reference standard used in this study is not explicitly stated as a separate entity, but rather it appears to be the "sentence-transformers/all-mpnet-base-v2" model (a pre-trained language model) which was fine-tuned on the provided text data.

12. Methods: Test methods: Blinding

The study was not blinded as there is no mention of blinding in the provided text.

13. Methods: Statistical methods: Handling of indeterminate results

The text does not mention handling of indeterminate results in the statistical methods section.

14. Methods: Statistical methods: Assessment of bias

The text does not mention any statistical methods or assessment of bias in the study's methodology.

15. Methods: Statistical methods: Sample size calculation

Not applicable, as the sample size calculation is not mentioned in the provided text.

16. Methods: Statistical methods: Methods for estimating diagnostic accuracy

The study used sentence-transformers/all-mpnet-base-v2 and Alibaba-NLP/gte-large-en-v1.5 as models to estimate diagnostic accuracy. However, the specific statistical methods used for estimating diagnostic accuracy are not mentioned in the provided text.

17. Results: Participants: Flow diagram

The provided text does not contain a flow diagram or any information about participants. The text appears to be an article discussing a study on developing a model for radiology report retrieval using natural language processing techniques.

18. Results: Participants: Baseline demographic and clinical characteristics

There is no information about participants, baseline demographic and clinical characteristics in the provided text.

19. Results: Participants: Distribution of severity/alternative diagnoses

The provided text does not mention any specific results, participants, or distribution of severity/alternative diagnoses. It appears to be a research article discussing the development and evaluation of a deep learning model for radiology report retrieval, but it does not include the requested information.

20. Results: Test results: Cross tabulation of index test vs reference standard

The article presents a study on developing an artificial intelligence model for radiology report retrieval using sentence embeddings. The results are presented in the form of cross-tabulations comparing the index test (the AI model) with the reference standard. However, based solely on the

provided text, it is not possible to provide detailed information about the specific results or outcomes of the study, such as sensitivity, specificity, accuracy, or other performance metrics. The text does mention that the authors compared their model's performance with a pre-trained model called "gte-large-en-v1.5" and found that their model performed better in certain aspects, but it does not provide detailed statistics. It appears that the study aimed to evaluate the performance of different sentence embedding models for radiology report retrieval, using metrics such as average precision and area under the ROC curve.

21. Results: Test results: Time interval and interventions between tests

There is no mention of test results, time intervals, or interventions between tests in the provided text.

22. Results: Estimates: Diagnostic accuracy estimates with confidence intervals

RadBERT achieved better performance than GTE-Large and MS-MARCO DistilBERT in radiology report retrieval tasks. However, I must note that the context is limited, and more specific information about the results of RadBERT compared to other models is not provided.

23. Results: Estimates: Adverse events from index test or reference standard

There is no mention of adverse events from the index test or reference standard in the provided text.

24. Discussion: Study limitations

The study has several limitations mentioned in the discussion section of the paper. Specifically: * The model's utility and limitations need to be further explored, and this will be done in a separate future study. Additionally, some potential limitations are implied by the references cited, but not explicitly stated as limitations of the current study.

25. Discussion: Implications for practice

RadBERT and other language models have been adapted to radiology tasks, but they may be biased due to coarse race and ethnicity labels. Mitigating bias in radiology machine learning is crucial for accurate diagnoses. The study discusses a new model, RadBERT, which adapts transformer-based language models to radiology. However, the implications of this research for practice are not explicitly discussed in the provided text.

26. Other Information: Registration number and name of registry

There is no mention of a registration number or the name of a registry in the provided text.

27. Other Information: Where the full study protocol can be accessed

The full study protocol cannot be accessed as per the provided text.

28. Other Information: Sources of funding and role of funders

The study was supported by the Department of Radiology and Biomedical Imaging at the University of California, San Francisco.