Section in the CLAIM checklist

Data

1. **Data sources:** State the source(s) of data, including publicly available datasets and/or synthetic images. Describe how well the data align with the intended use and target population of the model. (CLAIM 7)

2. **Inclusion and exclusion criteria:** specify inclusion and exclusion criteria, such as location, dates, patient-care setting, demographics (eg, age, sex, race), pertinent follow-up, and results from prior tests. Define how, where, and when potentially eligible participants or studies were identified. Indicate whether a consecutive, random, or convenience series was selected. (CLAIM 8)

3. **Selection of data subsets:** For example, describe whether investigators selected a subset of the images, cropped portions of images, or extracted segments of a report. If this process is automated, describe the tools and parameters used. If performed manually, describe the training of the personnel and criteria used in their instruction. Justify how this manual step would be accommodated in context of the clinical or scientific problem, describing methods of scaling processes, when applicable. (CLAIM 10)

4. **How missing data were handled:** Clearly describe how missing data were handled. For example, describe processes to replace them with approximate, predicted, or proxy values. Discuss biases that imputed data may introduce. (CLAIM 12)

Evaluation

5. **Metrics of model performance:** Describe the metrics used to assess the model's performance and indicate how they address the performance characteristics most important to the clinical or scientific problem. Compare the presented model to previously published models. (CLAIM 28)

6. **Statistical measures of significance and uncertainty:** Indicate the uncertainty of the performance metrics' values, such as with standard deviation and/or confidence intervals. Compute appropriate tests of statistical significance to compare metrics. Specify the statistical software, including version. (CLAIM 29)

7. **Evaluation on internal data:** Document and describe evaluation performed on internal data. (CLAIM 32)

8. **Testing on external data:** Describe the external data used to evaluate the completed algorithm. (CLAIM 33)

Results

9. **Numbers of patients or examinations included and excluded:** Document the numbers of patients, examinations, or images included and excluded based on each of the study's inclusion and exclusion criteria. Include a flowchart or alternative diagram to show selection of the initial patient population and those excluded for any reason (CLAIM 35)

10. **Demographic and clinical characteristics of cases in each partition:** Specify the demographic and clinical characteristics of cases in each partition and dataset. Identify sources of potential bias that may originate from differences in demographic or clinical

characteristics, such as sex distribution, underrepresented racial or ethnic groups, phenotypic variations, or differences in treatment. (CLAIM 36)

11. **Performance metrics and measures of statistical uncertainty:** Report the final model's performance. State the performance metrics on all data partitions and datasets, including any demographic subgroups (CLAIM 37)

12. **Estimates of diagnostic performance and their precision:** For classification tasks, include estimates of diagnostic accuracy and their measures of statistical uncertainty, such as 95% confidence intervals. When the direct calculation of confidence intervals is not possible, report nonparametric estimates from bootstrap samples. State which variables were shown to be predictive of the response variable. If applicable, recognize the presence of class imbalance (uneven distribution across data classes within or between datasets). (how they perform the estimation of CI or statistical uncertainty. If they state which were the predictive variables) (CLAIM 38)

13. **Failure analysis of incorrect results:** Provide information to help understand incorrect results. If the task entails classification into two or more categories, provide a confusion matrix that shows tallies for predicted versus actual categories. Consider presenting examples of incorrectly classified cases to help readers better understand the strengths and limitations of the algorithm. Provide sufficient detail to frame incorrect results in the appropriate medical context (CLAIM 39)

14. **Public availability (code)**