Supplementary Materials: Evaluating the Checklist for Artificial Intelligence in Medical Imaging (CLAIM)-based Quality of Reports Using Convolutional Neural Network for Odontogenic Cyst and Tumor Detection

Van Nhat Thang Le [1,2,3,4], Jae-Gon Kim [1,2,3], Yeon-Mi Yang [1,2,3] and Dae-Woo Lee [1,2,3,*]

Table 1. Detailed search strategies for each database. Mesh terms, search terms, and combinations of the two were used for each database search.

Database	Detailed search strategies	Studies founded
PubMed	("deep learning"[MeSH Terms] OR deep learning[Text Word] OR convolution neural network[Text Word] OR convolution neural networks[Text Word]) AND ("odontogenic tumors"[MeSH Terms] OR odontogenic tumor[Text Word] OR "odontogenic cysts"[MeSH Terms] OR odontogenic cysts[Text Word])	4
EMBASE	('deep learning' OR 'convolutional neural network') AND ('odontogenic tumor' OR 'odontogenic cyst')	5
SCOPUS	ALL ("deep learning" OR "convolution neural network" OR "convolution neural networks") AND ALL (tumor OR tumors OR cysts OR cyst) AND ALL (odontogenic)	41
Web of Science	ALL FIELDS: ("deep learning" OR "convolution neural network" OR "convolution neural networks") AND ALL FIELDS: (tumor OR tumors OR cysts OR cyst) AND ALL FIELDS: (odontogenic)	5

Ultimately, 55 records were found, 4 from PubMed, 5 from EMBASE, 41 from Scopus, and 5 from the Web of Science. Studies were further selected according to the inclusion criteria listed in the Material and Methods (Figure 1).

Table 2. Evaluating the CLAIM-based quality of CNN reports for odontogenic cyst and tumor detection.

Section/topic	No.	Checklist item - yes if reported	Liu et al.	Kwon et al.	Yang et al.	Ariji et al.	Lee et al.	Poedjiastoeti et al.	No. (%) of reports (n = 6)
TITLE or ABSTRACT									
	1	Identification as a study of AI methodology, specifying the category of technology used (e.g., deep learning)	1	1	1	1	1	1	6 (100%)
ABSTRACT									
Structured summary	2	Structured summary of the study design, methods, results, and conclusions	0	1	1	1	0	1	4 (67%)
INTRODUCTION									
Rationale	3	Scientific and clinical background, including the intended use and clinical role of the AI approach	1	1	1	1	1	1	6 (100%)
Objectives	4	Study objectives and hypotheses	1	1	1	1	1	1	6 (100%)
METHODS									
Study design	5	Prospective or retrospective study	1	1	1	1	1	1	6 (100%)
	6	Study goal, such as model creation, exploratory study, feasibility study, noninferiority trial	0	0	0	0	0	0	0 (0%)
Data	7	Data sources	0	1	1	1	1	1	5 (83%)
	8	Eligibility criteria: how, where, and when potentially eligible participants or studies were identified (e.g., symptoms, results from previous tests, inclusion in registry, patient-care setting, location, and dates)	0	0	1	1	1	0	3 (50%)
	9	Data pre-processing steps	1	1	0	1	1	1	5 (83%)
	10	Selection of data subsets, if applicable	1	1	1	1	1	1	6 (100%)
	11	Definitions of data elements, with references to common data elements	0	0	0	0	0	0	0 (0%)
	12	De-identification methods	0	0	0	0	0	0	0 (0%)
	13	How missing data were handled	0	0	0	0	0	0	0 (0%)

Ground truth	14	Definition of the ground truth reference standard, in sufficient detail to allow replication	0	0	0	0	0	0	0 (0%)
	15	Rationale for choosing the reference standard (if alternatives exist)	0	0	0	0	0	0	0 (0%)
	16	Source of ground truth annotations; qualifications and preparation of annotators	0	1	0	0	1	1	3 (50%)
	17	Annotation tools	0	1	0	1	0	0	2 (33%)
	18	Measurement of inter- and intra-rater variability; methods to mitigate variability and/or resolve discrepancies	0	0	0	0	0	0	0 (0%)
Data partitions	19	Intended sample size and how it was determined	1	1	1	1	1	1	6 (100%)
	20	How data were assigned to partitions; specify proportions	1	1	1	0	1	1	5 (83%)
	21	Level at which partitions are disjoint (e.g., image, study, patient, institution)	0	0	0	0	0	0	0 (0%)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections	1	1	1	1	1	1	6 (100%)
	23	Software libraries, frameworks, and packages	1	0	1	1	1	0	4 (67%)
	24	Initialization of model parameters (e.g., randomization, transfer learning)	1	0	0	0	1	1	3 (50%)
Training	25	Details of training approach, including data augmentation, hyperparameters, and number of models trained	1	1	1	1	1	1	6 (100%)
	26	Method of selecting the final model	0	1	0	0	0	0	1 (17%)
	27	Ensembling techniques, if applicable	1	0	0	0	0	1	2 (33%)
Evaluation	28	Metrics of model performance	0	1	1	1	1	1	5 (83%)
	29	Statistical measures of significance and uncertainty (e.g., confidence intervals)	0	0	1	0	1	1	3 (50%)
	30	Robustness or sensitivity analysis	0	1	0	1	1	1	4 (67%)
	31	Methods of explainability or interpretability (e.g., saliency maps) and how they were validated	0	0	0	1	0	0	1 (17%)
	32	Validation or testing on external data	1	1	1	1	1	1	6 (100%)
RESULTS									

	33	Flow of participants or cases, using a diagram to indicate inclusion and exclusion	0	0	1	0	0	0	1 (17%)
	34	Demographic and clinical characteristics of cases in each partition	0	0	1	1	1	0	3 (50%)
	35	Performance metrics for optimal model(s) on all data partitions	1	1	1	1	1	1	6 (100%)
	36	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)	1	1	1	0	1	1	5 (83%)
	37	Failure analysis of incorrectly classified cases	0	1	1	1	1	1	5 (83%)
DISCUSSION									
	38	Study limitations, including potential bias, statistical uncertainty, and generalizability	0	1	1	1	1	1	5 (83%)
	39	Implications for practice, including the intended use and/or clinical role	1	1	1	1	1	1	6 (100%)
OTHER INFORMATION									
	40	Registration number and name of registry	0	0	0	0	0	0	0 (0%)
	41	Where the full study protocol can be accessed	0	0	0	0	0	0	0 (0%)
	42	Sources of funding and other support; role of the funders	1	1	1	0	1	0	4 (67%)

Data are presented as number (%) of reports featuring the corresponding item.