

SURVEY

Open Access



Data-centric artificial intelligence in oncology: a systematic review assessing data quality in machine learning models for head and neck cancer

John Adeoye¹, Liuling Hui² and Yu-Xiong Su^{1*}

*Correspondence:
richsu@hku.hk

¹ Division of Oral and Maxillofacial Surgery, Faculty of Dentistry, University of Hong Kong, Hong Kong SAR, China

² Division of Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, University of Hong Kong, Hong Kong SAR, China

Abstract

Machine learning models have been increasingly considered to model head and neck cancer outcomes for improved screening, diagnosis, treatment, and prognostication of the disease. As the concept of data-centric artificial intelligence is still incipient in healthcare systems, little is known about the data quality of the models proposed for clinical utility. This is important as it supports the generalizability of the models and data standardization. Therefore, this study overviews the quality of structured and unstructured data used for machine learning model construction in head and neck cancer. Relevant studies reporting on the use of machine learning models based on structured and unstructured custom datasets between January 2016 and June 2022 were sourced from PubMed, EMBASE, Scopus, and Web of Science electronic databases. Prediction model Risk of Bias Assessment (PROBAST) tool was used to assess the quality of individual studies before comprehensive data quality parameters were assessed according to the type of dataset used for model construction. A total of 159 studies were included in the review; 106 utilized structured datasets while 53 utilized unstructured datasets. Data quality assessments were deliberately performed for 14.2% of structured datasets and 11.3% of unstructured datasets before model construction. Class imbalance and data fairness were the most common limitations in data quality for both types of datasets while outlier detection and lack of representative outcome classes were common in structured and unstructured datasets respectively. Furthermore, this review found that class imbalance reduced the discriminatory performance for models based on structured datasets while higher image resolution and good class overlap resulted in better model performance using unstructured datasets during internal validation. Overall, data quality was infrequently assessed before the construction of ML models in head and neck cancer irrespective of the use of structured or unstructured datasets. To improve model generalizability, the assessments discussed in this study should be introduced during model construction to achieve data-centric intelligent systems for head and neck cancer management.

Keywords: Artificial intelligence, Data-centric AI, Data quality, Head and neck cancer, Machine learning, Review

Introduction

Background

Head and neck cancer is the seventh most common malignancy worldwide [1]. Tumors may affect the lips, oral cavity, nasal cavity, nasopharynx, oropharynx, hypopharynx, larynx, paranasal sinuses, and salivary glands with about 90 to 95% being of the squamous cell carcinoma variant [2, 3]. Predisposing factors for head and neck cancer are diverse including prolonged sunlight exposure; risk habits such as tobacco use, heavy alcohol consumption, and betel nut chewing; infectious agents such as human papillomavirus and Epstein Barr virus; lifestyle factors such as low fruit and vegetable consumption; and a family history of malignancies [3, 4]. Moreover, the prognosis of head and neck cancer is dismal with a five-year survival of about 50% to 60% for most affected sites [5].

To improve the screening, detection, and prognosis of head and neck cancer, artificial intelligence techniques and tools have been proposed to support clinicians in the diagnosis, decision-making, and risk stratification of the disease [6–8]. Specifically, machine learning (ML) models are being increasingly constructed to facilitate disease segmentation from different investigations during diagnosis, treatment planning, and treatment monitoring as well as to predict risk status for screening, prognostic, and treatment complication outcomes with high accuracy [9, 10]. Moreover, disparate reports have showcased that ML models outperform conventional statistical techniques in performing tasks related to head and neck cancer [6, 7, 11, 12].

One of the factors that may hamper the optimal performance of ML models and their generalizability is the quality of data used for model construction [13, 14]. Moreover, better application of the ML models is expected when developed with representative high-quality datasets that can be replicated in different clinical scenarios and settings [13–17]. Several studies and reviews have focused more on the validity estimates and performance of ML models in head and neck cancer without considering whether data quality and standards are sufficient to enable the meaningful realization of the potential of ML and artificial intelligence (AI) in clinical practice [6–12]. To tackle this knowledge gap, this study overviews the quality of structured and unstructured data that have been used to construct machine learning models for head and neck cancer to identify current limitations to their optimal performance and generalizability based on the datasets employed.

Research questions

S/N	Research questions	Motivation
1	What category of datasets was mostly employed to model/predict head and neck cancer outcomes using machine learning?	To highlight the common types of datasets and outcomes available or considered among researchers and clinicians for modeling head and neck cancer outcomes with machine learning algorithms
2	How good are datasets used to construct machine learning models for head and neck cancer outcomes?	To evaluate the quality of datasets that were used to implement machine learning models in head and neck cancer
3	What data quality criteria were often fulfilled or deficient in datasets used to construct machine learning models for head and neck cancer outcomes?	To determine the specific dimensions of data quality that were often met or lacking by datasets during the construction of machine learning models in head and neck cancer. The answer to this question will also assist relevant stakeholders (i.e., clinicians, health informaticians, engineers, developers, and researchers) to determine present limitations to obtaining optimal data quality in head and neck cancer machine learning prediction platforms. Knowledge of these limitations will highlight areas for improvement in future works
4	What is the effect of data quality on the median performance metrics of the machine learning models constructed in head and neck cancer?	To examine whether the dimensions of data quality impact the performances of machine learning models in head and neck cancer

Related research

The utilization and implementation of AI and ML platforms in oncology have increased significantly in recent years [18–21]. Moreover, head and neck cancer represent one of the most commonly modeled malignancies using AI/ML techniques [8, 22, 23]. While this systematic literature review sought to examine the data quality used to construct ML models in head and neck cancer, we highlight previous relevant studies that examined machine learning models for head and neck cancer outcomes to showcase the significance and novelty of our contribution.

Patil et al. [24] reviewed and summarized reports from seven ML models to conclude that the support vector machines (SVM) were mostly used with genomic datasets to predict head and neck cancer prognosis with an accuracy rate between 56.7 to 99.4%. Volpe et al. [25] reviewed 48 studies describing ML models in head and neck cancer from different imaging modalities to highlight their different potential clinical uses during radiation therapy. Bassani et al. [26] also reviewed 13 articles on diagnostic ML models in head and neck cancer and found that models largely had excellent accuracies above 90%; however, this study suggested that models were based on small/restricted datasets and lacked heterogeneity. Giannitto et al. [27] recently over-viewed eight studies that utilized radiomics-based machine learning models for head and neck cancer lymph node metastasis prediction and concluded that the models had a sensitivity and specificity of 65.6 to 92.1% and 74 to 100%. Of note, this systematic review also assessed the methodology of the ML models and suggested that most of them had a biased feature extraction and lacked external validation [27].

Other studies also focused on ML models for specific head and neck cancer subtypes [6, 12, 28]. Alabi et al. [12] from 41 studies found that ML models could aid the diagnosis and prognosis of oral cancer with sensitivity and specificity that ranged from 57 to 100% and specificity from 70 to 100%; although there were concerns about

the explainability of the models. Again, Alabi et al. [28] in another review focusing on deep learning models for oral cancer found that the models had an average AUC of 0.96 and 0.97 for spectral datasets and computed tomography (CT) image datasets respectively. However, ethical concerns were suggested to have limited the application of these deep learning models in oral cancer management [28]. Chiesa-Estomba et al. [29] reviewed eight studies reporting ML models for oral cancer survival and decision-making and suggested that the tools could potentially advance the prediction of these tasks in patient management but highlighted the small number of data available and the use of secondary datasets as limitations to their application. Of note, our group also showed that machine learning models had accuracies that ranged from 64 to 100% for different oral cancer outcomes but that the models were not streamlined enough for clinical application due to lack of external validation, the prevalent use of small single-center datasets, and lack of class imbalance correction [6].

Contributions and novelty

From the studies examined above, it is clear that the majority evaluated AI/ML models for head and neck cancer outcomes largely based on their performance measures and the modeling methodology/technique employed leading to the final model selection (model-centric approach). Moreover, of the few studies that reported on the data quantity, feature selection, or class imbalance correction of the AI/ML models reviewed; none examined the effect of these parameters on the discriminatory or calibration performances of the models. Of note, no review comprehensively highlighted the different dimensions of data quality for assessment or questioned whether the data infrastructures and their quality were sufficient to encourage the meaningful application of the AI/ML models in the management of head and neck cancer or its subtypes. Given this gap, our study sought to assess datasets used for constructing ML models in head and neck cancer using disparate parameters for assessing data quality. Additionally, our contribution examines the relationship between the data quality assessment criteria and the performance of the ML models for head and neck cancer.

Overview of the study

The remainder of this article is structured as broadly as “[Review methodology](#)”, “[Results](#)”, “[Discussion](#)” and “[Conclusion](#)”. “[Review methodology](#)” section details the processes underwent to arrive at the final studies assessed in this systematic literature review. Also, the data abstracted from the individual studies, their risk of bias rating, and the methods for result synthesis are detailed in this section. “[Results](#)” section is further divided into two sub-sections i.e., “[ML models based on structured datasets](#)” and “[ML models based on unstructured datasets](#)” based on the type of datasets utilized for model construction. In each of the “[Results](#)” subsections (structured and unstructured data), we report the general characteristics of the studies, the risk of bias assessment findings, and the findings of the different quality assessment criteria used for data evaluation. “[Discussion](#)” section justifies the findings of the data quality evaluation and presents the limitations

of the review. Finally, the section “**Conclusion**” answers the aims and review questions of this study within the confinements of its limitations and provides suggestions for future works.

Review methodology

Eligibility criteria

Original research articles between January 2016 and June 2022 that reported on the use of machine learning models on custom datasets of head and neck cancer patients were sourced. The rationale for choosing this timeframe was informed by the introduction of a robust methodology and reporting standards such as the Transparent reporting of a multivariate prediction model for individual prognosis or diagnosis (TRIPOD) and the Standards of Reporting Diagnostic Accuracy (STARD) statements in 2015 [30, 31]. The scope of head and neck cancer included malignant tumors (carcinomas) of any histologic variants affecting the lips, oral cavity, nasal cavity, nasopharynx, oropharynx, hypopharynx, larynx, paranasal sinuses, and salivary glands (ICD-10 C00-14, C31 and C32). Machine learning algorithms were limited to conventional learning, deep learning, and deep hybrid learning models that were trained and had at least a form of internal validation using test-train split, cross-validation, or bootstrap resampling. Likewise, structured, and unstructured datasets sourced from health records (electronic or manual), clinical photographs, imaging modalities, spectroscopy techniques, digitized histology slides, and omics technologies were considered for inclusion.

Studies that included other diseases during modeling, explicitly conducted a pilot/preliminary/feasibility study as stated in the report, comprised unstructured data that were not in form of images or spectra data, and utilized nomograms to operationalize models were excluded. Further excluded were treatment planning studies that examined the utilization of ML models for the segmentation of normal structures or organs without emphasis on the tumor areas and studies that were not full research articles. Also, short communications, letters to the editor, case series, and conference proceedings that were not peer-reviewed were not included. Studies that utilized public datasets (such as TCGA or TCIA cohorts) during training were excluded from this study following qualitative evaluation of related articles. The rationale for this exclusion was due to: (i) the preference of this study for custom datasets with clear reporting of the methods used for data collection (ii) abundance of duplicate studies using these cohorts which will introduce bias during result synthesis (iii) most studies based on public databases selected only a proportion of the total cohorts to predict head and neck cancer outcomes which precluded the assessment of the entire database against any of the ML models (iv) ML models using public databases focused on the ML approach and improvement of AI techniques than validation and implementation (v) reduced likelihood of clinical implementation for models trained using these public databases. For duplicate studies of different ML models conducted with the same custom dataset, only the first study was included provided the dataset was not updated with new patients subsequently. Alternatively, the updated study was included while excluding the first set of models constructed.

Data sources and search strategy

Related studies were sourced from four electronic databases—PubMed, Web of Science, EMBASE, and Scopus. Search keywords were first selected based on words pertinent to the research questions and review objectives while also being in line with other literature reviews that assessed the performance of machine learning models in head and neck cancer [6, 7, 9, 12, 26, 28]. Search terms were then implemented in each database with Boolean operators as presented in Additional file 1: Table S1. Retrieved citations were exported to EndNote 20 (Clarivate Analytics, USA) for automated and manual deduplication.

Study selection

This study adopted a two-stage process for the selection of relevant articles. In one stage, two authors independently screened the title and abstract of the citations to identify research articles on ML models constructed for head and neck cancer outcomes. Then, during the second stage, full-length texts of articles considered after screening were assessed strictly based on the eligibility criteria. This was also performed in duplicate and discordant selections were resolved following discussions between the authors. Agreement between authors was the basis for the final study selection in this review. Following article selection, a supplementary manual search was also performed to bolster electronic database searching and ensure that studies missed are included.

Risk of bias assessment and data extraction

Quality rating of the selected studies was performed using the Prediction model Risk of Bias Assessment tool (PROBAST) [32] with domains evaluating the study participants, predictors, outcome, and analytical methods. These four domains were scored as high, low, or unclear using signaling questions recommended for each aspect. The overall risk of bias rating was rated as 'high' if at least one domain was rated as high or low risk of bias if all the domains were rated as 'low'. For studies where three domains were rated as low, and one domain was rated as unclear, the overall risk of bias rating was deemed unclear.

Data abstracted from the individual studies included general study characteristics such as first author names, publication year, study location, income class of study location, number of centers, cancer type, model type, task performed, clinical utility, model outcome, data preprocessing, the use of a reporting guideline, type of validation, and model discrimination estimates (AUC and C-statistic). Data quality parameters were then extracted for all studies using datasets for generating results when available or study metadata if dataset was absent. Data were described as structured if all the features were assigned continuous or categorical labels (including features that were extracted from images or spectra) while unstructured data referred to images or spectra that were utilized without deliberate feature extraction with deep learning or deep hybrid learning techniques. For structured data, the data type and source, multidimensionality, timing of dataset collection, class overlap, label purity, outcome class parity, feature extraction, feature relevance, collinearity, data fairness, data completeness, outlier detection,

and data representativeness were assessed. In addition to some of the aforementioned parameters, image or spectra quality was assessed for studies that utilized unstructured datasets. Definitions and methods of assessing individual data quality parameters are detailed in Additional file 1: Table S2.

Result synthesis and statistical analyses

Qualitative synthesis was adopted for summarizing the data extracted from individual studies. Descriptive statistics were performed and presented in the text and figures as proportions, medians, and interquartile ranges. Statistical differences between two or more categorical variables were assessed using Pearson’s Chi-square test or Fisher’s exact analysis as appropriate. Likewise, median values of continuous variables based on different categories were compared using Mann–Whitney U test and Kruskal–Wallis H test. SPSS v 27 was used for statistical analyses performed with probability values below 5% indicating statistical significance. This review was conducted with according to the Preferred Reporting Items for Systematic reviews and Meta-analysis (PRISMA) guideline [33].

Results

The flowchart depicting the screening and study selection process is presented in Fig. 1. Upon removing duplicates, 2851 citations were screened, and 528 articles were selected for full text evaluation. Three hundred and sixty-nine articles did not fulfil

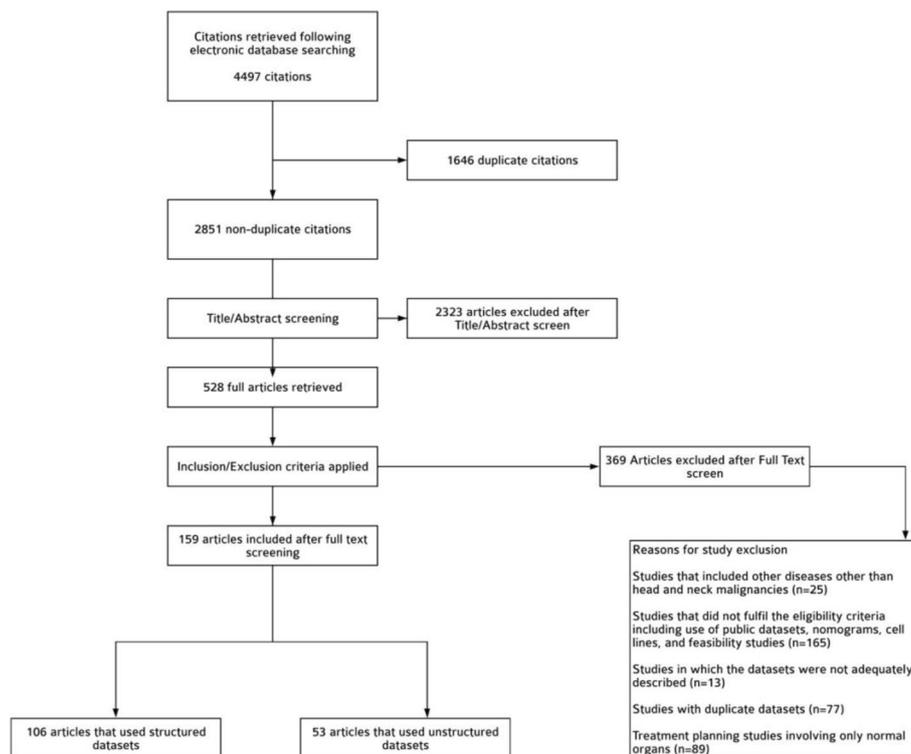


Fig. 1 Flowchart detailing the search strategy and article selection process in the review

the eligibility criteria based on reasons stated in Fig. 1. Overall, 159 articles on ML models for head and neck cancer outcomes were included in this review [34–192] with 106 models based on structured data while 53 articles applied the ML models on unstructured data. A list of all studies included is presented in Additional file 1: Tables S3 and S4.

ML models based on structured datasets

General study characteristics

All 106 studies that utilized structured datasets were published between 2016 and 2022 with more studies in 2021 than other years (Fig. 2). Majority of the studies were conducted in China (26.4%) and the USA (16%) with only 5.7% of the studies conducted in India which represented the only low or lower-middle income country. Based on the subtypes of head and cancer considered, more ML models for structured datasets were used for oral cavity tumors (33%) than nasopharyngeal (18.9%), oropharyngeal (14.2%), or other tumors (6.5%) (Fig. 2). Disparate head and neck cancer subtypes were considered in 27.4% of the studies and no studies on lip cancers were present. Many models based on structured data utilized conventional machine learning (83%) than deep learning (15.1%) or deep hybrid learning (1.9%) algorithmic frameworks. Specifically, many of the models utilized random forest (19.8%), regularized logistic regression (17.8%), and support vector machines (14%) than other models.

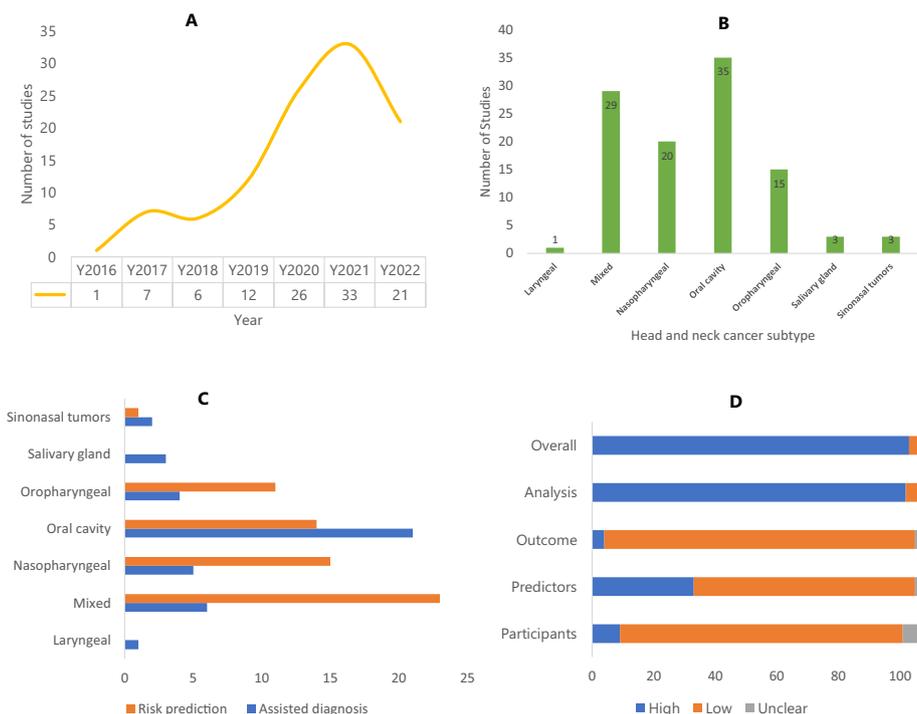


Fig. 2 General characteristics and risk of bias rating for ML studies employing structured datasets. **a** Publication year trend. **b** Plot showing the number of patient datasets by the different head and neck cancer subtypes. **c** Plot showing the clinical utility of the models according to the different cancer subtypes. **d** Risk of bias rating for individual domains of the PROBAST tool

Most ML models for structured datasets were developed to perform classification tasks (96.2%) than segmentation (1.9%), clustering (0.9%), or regression tasks (0.9%). Likewise, the clinical utility of the models was mostly for risk prediction (60.4%) than assisted diagnosis (39.6%). When stratified based on the cancer subtypes, most ML models on structured data for oropharyngeal (73.3%), nasopharyngeal (75%), and combined head and neck cancers (79.3%) were utilized for risk prediction while models for oral cavity cancers (60%), salivary gland tumors (100%), and laryngeal cancer (100%) were mostly for assisted diagnosis ($p=0.002$). Also, cancer outcomes considered using this type of dataset included diagnosis (43.4%), prognosis (33.0%), treatment (16%), and screening (7.5%).

Risk of bias for individual studies

According to the PROBAST tool, only 3 of 106 studies had a low overall risk of bias rating (Fig. 2). Based on the four domains, many head and neck cancer studies fell short mostly regarding their methodology of data analysis and ML modeling (96.2%). Common reasons for high risk of bias rating in this domain was related to the low event-per-variable (EPV) of the dataset used for training or reduced sample size of patients with events during model validation. Likewise, 31.1% had a high bias rating in the 'Predictors' domain of the PROBAST tool mostly because the features used for model construction were determined after the outcome of the studies were known. Only 8.5% and 3.8% of studies had high bias ratings in the 'Participants' and 'Outcome' domains of the PROBAST tool (Fig. 1).

Data quality of ML models based on structured datasets

Of note, only six models (5.7%) followed at least one reporting guideline for data and model description, four of which used the TRIPOD statement. Of 106 ML models for head and neck cancer outcomes based on structured data, only nine have been validated on independent/external cohorts (8.5%). Sample size based on the number of patients in the datasets ranged from 12 to 27,455 with a median (IQR) sample of 165 (68–313) patients. 55.6% of models were based on a cohort size below 200 while 85.9% of studies were based on a cohort size of less than 500 patients. According to the number of settings from which the datasets originated, majority involved single centers (80.2%) while 11.3% were from dual centers. Multi-institutional structured datasets were used in only 9 models (8.5%).

Based on the data types, most were based on radiomic features (44.3%), and clinical features obtained from health records (40.6%). Molecular, pathological, and spectral datasets were used in 17%, 6.6%, and 5.7% respectively. Feature extraction was performed in 66% of the ML models, and for studies that used images and reported their resolution ($n=20$), relatively similar proportion of models used images below 224×224 (45%) as well as 512×512 and above (40%). Multidimensional data comprising different feature types were used in 17.9% of the ML models. Likewise, pattern of data collection was mostly retrospective (73.6%) than prospective (26.4%). Data preprocessing was performed and reported for most models (95.3%) while the quality of structured datasets was deliberately assessed in few studies (14.2%).

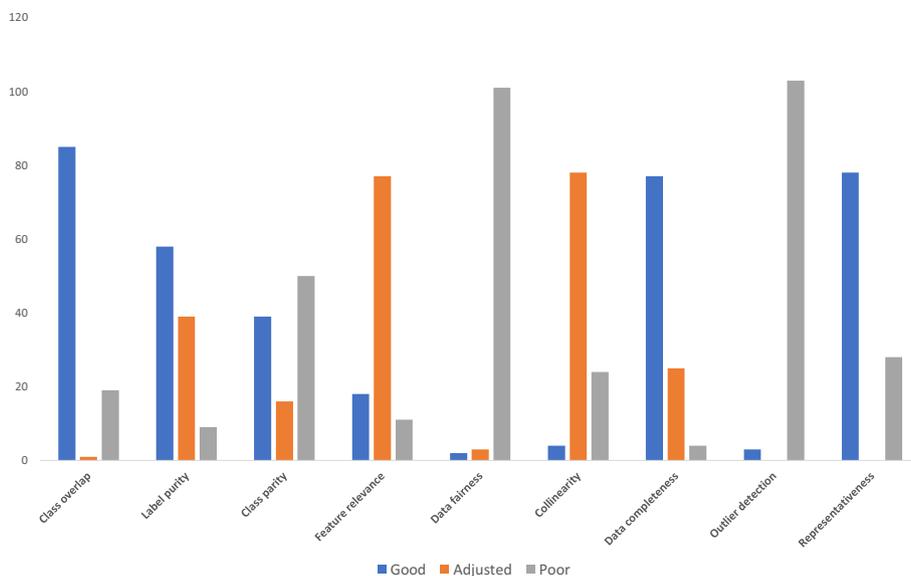


Fig. 3 Distribution of structured datasets according to the data quality assessment parameters

Based on the class overlap for all studies that performed classification, most studies were satisfactory in the separation of the outcome class for all possible categories (80.2%; Fig. 3). Likewise, noisy/inconsistent instances were likely absent in 54.7% of datasets with 36.8% adjusting for noise using dimensionality reduction techniques such as principal component analysis. However, class imbalance in the outcome category were observed among more datasets (47.1%) than limitations in class overlapping or label purity in classification tasks. Feature relevance in many datasets was often adjusted using feature selection methods (72.6%) while 17% of studies that presented feature importance measures used $\geq 50\%$ of features that were related to the outcomes of interest. Similarly, the predominant use of dimensionality reduction or feature selection techniques in structured head and neck cancer datasets also meant that redundant and correlated features were dropped and adjusted for most datasets (73.6%) during training (Fig. 3).

Of note, most of the datasets fell short regarding data fairness as 95.3% of them had sample sizes that did not allow adequate representation of those with events across different levels of the predictors. Datasets were often complete (72.6%) owing to the predominant use of feature extraction techniques to generate structured data; however, missing data handling techniques were used in 23.6% of datasets most of which involved variable exclusion. Outlier detection was performed for datasets of three models (2.8%) and most datasets (73.6%) had outcome classes which were representative of all possible scenarios according to the aims of the studies (Fig. 3).

Relating ML model performance and data quality parameters

The metrics of model discrimination in all studies ranged from 0.60 to 1.00 with a median value (IQR) of 0.85 (0.77–0.91). Table 1 shows the distribution of the median model discrimination metrics across various aspects of data quality assessment parameters. Notably, models with good balance in the outcome class had significantly higher

Table 1 Crosstabulation of the data quality assessment parameters and the median model discrimination metrics obtained in individual studies with structured datasets

Assessment parameter		Median discrimination metrics (IQR)	p-value ^a
Class overlap	Adjusted/Good	0.84 (0.77–0.91)	0.331
	Poor	0.87 (0.81–0.92)	
Label purity	Adjusted/Good	0.85 (0.77–0.91)	0.759
	Poor	0.85 (0.78–0.90)	
Class parity	Adjusted/Good	0.88 (0.80–0.94)	0.014
	Poor	0.81 (0.75–0.89)	
Feature extraction	Yes	0.86 (0.77–0.94)	0.260
	No	0.84 (0.78–0.90)	
Feature relevance	Adjusted/Good	0.86 (0.77–0.91)	0.204
	Poor	0.79 (0.76–0.88)	
Data fairness	Adjusted/Good	0.77 (0.70–0.84)	0.079
	Poor	0.86 (0.78–0.91)	
Collinearity	Adjusted/Good	0.86 (0.78–0.92)	0.227
	Poor	0.81 (0.77–0.89)	
Data completeness	Adjusted/Good	0.85 (0.77–0.91)	0.612
	Poor	0.87 (0.78–0.95)	
Outlier detection	Adjusted/Good	0.90 (0.77–0.99)	0.431
	Poor	0.85 (0.77–0.91)	
Data representativeness	Adjusted/Good	0.84 (0.77–0.90)	0.059
	Poor	0.90 (0.79–0.96)	

^a Mann Whitney U test used for analysis

p-value in bold is statistically significant

median discrimination than those that did not adjust for class imbalance (0.88 vs 0.81, $p = 0.014$). Although, not statistically significant, this review found a higher median model discrimination for datasets with less collinearity (0.86 vs 0.81) and datasets where outliers were detected and adjusted (0.90 vs 0.85).

Datasets with high EPV had lower median performances than those with low EPV (0.77 vs 0.86). Likewise, data with outcome classes that were not representative of possible outcomes in clinical settings generated models with higher median performance than those with representative multiple classes (0.90 vs 0.84). As many models have not been externally validated, these may suggest model overfitting among models with low EPV and unrepresentative outcome classes, and a likelihood of reduced performance may be expected during generalizability assessment.

ML models based on unstructured datasets

General study characteristics

The 53 ML models based on unstructured datasets were published between 2017 and 2022 (Fig. 4). Most models used datasets from Chinese populations (41.5%) and only 9.4% of studies were developed in low or lower-middle income countries (four from India and one from Pakistan). Similar to the distribution of head and neck cancer subtypes for structured datasets, oral cavity tumors were more commonly modelled (28.3%) followed by nasopharyngeal cancer (22.6%), salivary gland tumors (11.3%), and laryngeal cancer (9.4%). Only one study was found for lip cancer and sinonasal tumors and

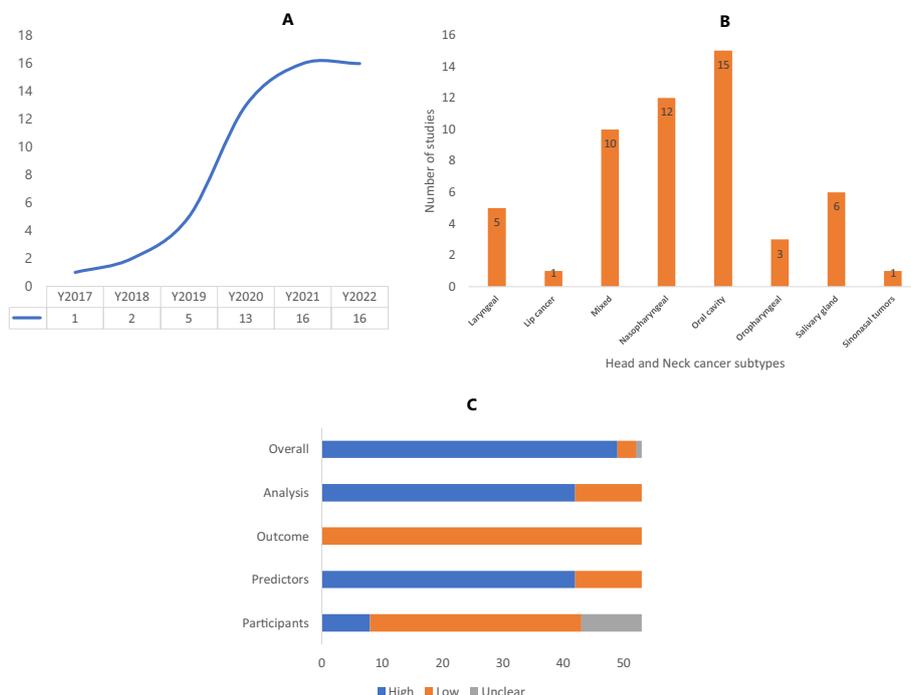


Fig. 4 General characteristics and risk of bias rating for ML studies employing unstructured datasets. **a** Publication year trend **b** Plot showing the number of patient datasets by the different head and neck cancer subtypes **c** Risk of bias rating for individual domains of the PROBAST tool

combined head and neck cancer subtypes were considered in 18.9% of models (Fig. 4). All of the models were of deep learning type based on convolutional neural networks (CNN). U-Net was the most common CNN framework used (15.1%) followed by Inception V3 (7.5%).

Regardless of the commonality of a segmentation CNN architecture, most of the deep learning models were developed to perform classification (81.1%) than segmentation tasks (18.9%). Unlike for structured datasets, majority of the ML models for unstructured datasets were constructed for potential clinical utility in assisted diagnosis (90.6%) than risk prediction (9.4%). Of the models intended for risk prediction, four were related to disease prognosis while one was related to the risk of treatment complications.

Risk of bias for individual studies

Distribution of risk of bias rating is shown in Fig. 4. Only three studies (5.7%) had low overall risk of bias rating while one study had an unclear overall bias based on the lack of description for the patient cohort used. All others had a high risk of bias. Regarding the PROBAST tool domains, many studies had high risk of bias regarding the analysis (79.2%) and use of predictors (79.2%) mostly due to a reduced EPV in the training dataset, a lack of independent samples for validation, and the determination of predictors while the patient outcomes were known.

Data quality of ML models based on unstructured datasets

Only two studies used reporting guidelines (TRIPOD and STROBE) for data and model description. Further, only three models (5.7%) with unstructured datasets have been externally validated. For studies that reported the sample size based on patients ($n=47$), cohort size ranged from 12 to 24,667 with a median (IQR) of 219 (102–502) patients. Unstructured data obtained for these patients in form of images was from 72 to 840,000 with a median (IQR) of 2726 (1291–8776). Most datasets were from single centers (81.1%) with 9.4% and 9.4% from two centers and multiple centers respectively.

Radiomics datasets were mostly utilized for ML model construction (54.7%) majority of which were either CT/PET images (30.2%) or MRI images (26.4%). Endoscopy images were used in 18.9% of the models while histopathology slide images and clinical photographs were used in 13.2% and 9.4% of models. Raman spectra were used in two studies. Also, unstructured datasets were often collected in a retrospective (88.7%) than prospective (11.3%) manner. Multidimensional dataset (comprising different feature categories) was used in a single study (1.9%). Data preprocessing before model training was reported to be performed for 83% of datasets with quality assessment determined for only 11.3% of datasets. Image resolution was mostly found to be between 224×224 and 512×512 pixels (46.3%) with 22.9% of datasets comprising input images of 512×512 and above.

For studies that performed classification tasks ($n=40$), class overlap was not present in most datasets (93%) (Fig. 5). However, the sample size per patient and images/spectra were not sufficient in many models (69.8%). Some studies adjusted the sample size by incorporating different data augmentation techniques (9.4%) while 11.8% of datasets were deemed sufficient for model training and validation. Class parity were slightly more insufficient (39.6%) than sufficient without imbalance correction (37.7%). Imbalance correction methods were introduced in 22.6% of the datasets.

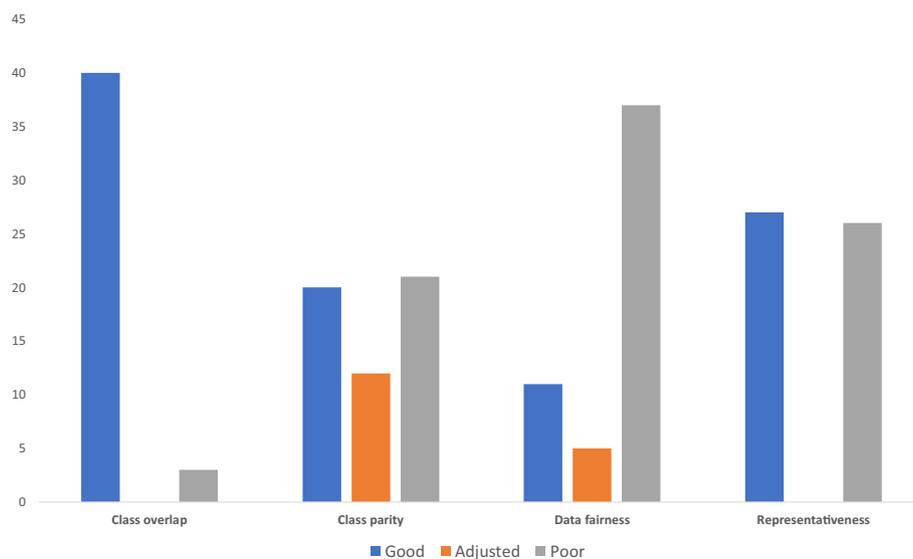


Fig. 5 Distribution of unstructured datasets according to the data quality assessment parameters

Table 2 Crosstabulation of the data quality assessment parameters and the median model discrimination metrics obtained in individual studies with unstructured datasets

Assessment parameter		Median discrimination metrics (IQR)	p-value
Image resolution	< 224 × 224	0.86 (0.81–0.96)	0.029^a
	224 × 224 to 512 × 512	0.96 (0.94–0.99)	
	≥ 512 × 512	0.94 (0.90–0.96)	
Class overlap	Adjusted/Good	0.95 (0.87–0.98)	0.030^b
	Poor	0.80 (0.76–0.90)	
Class parity	Adjusted/Good	0.94 (0.86–0.98)	0.125 ^b
	Poor	0.93 (0.84–0.97)	
Data fairness	Adjusted/Good	0.96 (0.92–0.98)	0.881 ^b
	Poor	0.92 (0.84–0.96)	
Data representativeness	Adjusted/Good	0.92 (0.84–0.96)	0.217 ^b
	Poor	0.96 (0.88–0.98)	

^a Kruskal-Wallis H test^b Mann Whitney U test used for analysis

P-value in bold is statistically significant

Also, a similar proportion of datasets had outcome classes which were representative of possible clinical outcomes (50.9% vs 49.1%) (Fig. 5).

Relating ML model performance and data quality parameters

The model discrimination metrics in all studies with unstructured data ranged from 0.72 to 1.00 with a median value (IQR) of 0.85–0.98. According to the distribution of data assessment parameters, significant differences were observed in the median model discrimination metrics based on the image quality requirements and class overlap (Table 2). Median performance of ML models for image datasets with resolutions above 224 × 224 was higher than those models based on lower resolutions. Likewise, ML models based on datasets without class overlap had better median performance than the few with class overlap (0.95 vs 0.80).

Discussion

This study overviewed structured and unstructured datasets employed for ML model construction in head and neck cancer and assessed their quality to unravel their limitations and improve future performance and the generalizability of the models.

Overall, the majority of the ML models reviewed used structured than unstructured datasets for their development. This increased utilization of structured over unstructured datasets may be attributed to the common choice of conventional ML architectures such as support vector machines, k-nearest neighbors, and ensemble learners that requires highly organized data for their implementation in this study [193]. Moreover, this finding is supported by the frequent application of feature extraction techniques based on size, shape, and texture to abstract radiomics features from raw medical images for training with these conventional classifiers [194, 195]. Likewise, electronic health records are one of the most abundant data sources in healthcare which are often

obtained for ML applications as structured than unstructured datasets and may contribute to this finding [196]. However, this finding suggests reduced application of AI architectures for automated handling of unstructured image and spectra datasets (such as deep learning) in head and neck cancer [23, 197]. Notably, irrespective of the data type many models had a high risk of bias that was often attributed to different issues in data sampling and resampling (i.e., cross-validation, bootstrapping) during model training. The lack of systematic ML modeling according to a standard guideline like the TRIPOD statement [31], the lack of external validation for most models, and the reduced application of multidimensional features represent other factors that hamper their clinical utility in head and neck cancer management [13, 198]. Nonetheless, these issues are consistent with other reports as limitations of AI models that are presently being constructed [27, 199, 200]. This study also noticed that while data preprocessing seemed to be commonly performed in many studies, deliberate assessment of data quality was lacking. This may be due to a lack of awareness of the different dimensions involved in data quality evaluation according to the types of datasets available for model construction and the lack of robust data-centric AI reports detailing the procedures involved in assessing data quality as part of data preprocessing [14]. Incorporating data quality dimensions into methodological and reporting guidelines/checklists may also assist developers and investigators imbibe this practice during ML model development [13].

This review observed that lack of outlier detection, lack of data fairness, and imbalance in outcome classes represented the most common limitations for structured datasets being utilized with ML models proposed for clinical utility in head and neck cancer. Of note, this study even observed a substantial decline in the median discriminatory performance of ML models with imbalanced outcome datasets that were not adjusted using correction techniques. As such, the current models proposed may suffer from errors during generalizability, especially among minority and/or unconventional cases that are infrequently encountered. Furthermore, as most of the models based on structured datasets were developed for risk prediction, if the ML models were applied in patient cohorts or regions in which events were prevalent (dataset shift), most of these cases may be missed resulting in poor performance of the models in real clinical scenarios [199]. Since the majority of the structured datasets were employed for classification, our findings on the effect of class parity/balance on model performance corroborate those of Budach et al. [14] that observed a moderate effect of class imbalance on classification tasks using machine learning. However, the effects of completeness, feature accuracy, and target accuracy on the performance of ML classifiers on structured datasets (depicted as data completeness and label purity in our study) were not observed. This may be due to the differences in the assessment of model performance as Budach et al. [14] utilized the F1-score to assess model performance and our study utilized the AUC as it was the performance metric commonly reported in the included studies. Also, this disparity may be explained by the lower proportion of studies that had poor ratings in data completeness and label purity following quality evaluation in our study.

For unstructured datasets, this review showed that lack of data fairness, class imbalance, and the use of outcome classes that were not representative of the entire clinical spectrum of the head and neck cancer subtypes of interest were the most common limitations for ML model construction in head and neck oncology. However, we observed

that these common limitations did not affect the discriminatory performance of the models. Of note, it was found that using overlapping outcome classes in unstructured datasets resulted in a lower discriminatory model performance of the model. This suggests that it is optimal for target classes of unstructured datasets to be used for multi-nomial/multilabel classification rather than binarizing the outcomes which also groups the input features. For example, unstructured data from patients with benign head and neck conditions should not be combined with those of normal patients as a “non-malignant class” and input features from patients with premalignant conditions should not be collated with those of patients with head and neck cancer as a “malignant class”. Likewise, this study observed that the image resolution of the input data may significantly affect the discriminatory performance of the model based on a cutoff resolution of 224×224 pixels. This finding supports the observations of Sabottke et al. [201] and Thambawita et al. [202] on the increase in the performance of neural networks when high-resolution medical images were used for model training.

Interestingly, in both structured and unstructured datasets, data fairness represented a common limitation in obtaining quality datasets for ML model development but did not significantly affect the discriminatory performance of the models. This finding may be attributed to publication bias and small-study effect since models with high AUC and C-statistics (>0.9) were likely to be published irrespective of whether their sample sizes for training and validation were adequate [203, 204]. Future research should ensure that the quantity of data for training and validation are sufficient, and this could be supplemented using data augmentation techniques [205, 206]. Also, journals may employ/enforce the use of robust checklists pertinent to clinical model development to mitigate publication bias and allow authors justify their methodology [30–32, 200].

Limitations

While this study uniquely overviews the data quality of ML models for head and neck cancer, it is not without limitations. First, a few large public databases with open-source datasets were excluded from the review. However, this was justified due to the potential of these datasets to introduce bias into the study due to duplication, unclear data collection methods, and the utilization of the databases for the improvement of ML techniques rather than their implementation/performance in modeling head and neck cancer outcomes. Second, some studies did not present the raw datasets used for ML modeling and the dimensions of data quality were assessed using the metadata. However, only data quality evaluation criteria that could be reasonably determined using this information were selected in this study. Last, the majority of the studies included and assessed using the data quality dimensions utilized ML for classification tasks than segmentation, clustering, or regression, especially for structured datasets. Nonetheless, this represents the real-world application of ML models to serve as decision-support tools in head and neck cancer management.

Conclusions and recommendations

Overall, this review found that structured datasets were mostly used to predict head and neck cancer outcomes using machine learning than unstructured datasets which is largely due to the increased use of traditional machine learning architectures that

requires highly organized datasets. Data quality was infrequently assessed before the construction of ML models in head and neck cancer for both structured or unstructured datasets. Irrespective of the datasets employed for model construction, class imbalance and lack of data fairness represented the most common limitations. Also, in structured datasets, lack of outlier detection was common while in unstructured datasets many datasets target classes were not representative of the clinical spectrum of the disease conditions or health status. Class imbalance had the most effect on the internal validation performance of the models for structured datasets while image resolution and class overlap had the most effect on the performance of ML models for unstructured datasets.

Based on the findings of this review and the need to bolster the implementation of ML models in contemporary head and neck cancer management this study recommends that: (i) Studies deliberately assess the quality of structured/unstructured data used for model construction. This could be done using automated comprehensive tools (such as in [207]) or manually based on the individual data parameters assessed in this review (especially data fairness, outlier detection, class imbalance detection, influence of image resolution, and outcome class representativeness) [13, 14] (ii) Relevant stakeholders develop a standard for assessing and reporting data quality parameters for studies based on ML models (iii) Studies strive to adhere to reporting guidelines such as TRIPOD, PROBAST, or STARD to ensure standardization of modeling approach (iv) Available models be retrained and externally validated using data that is of sufficient quality and quantity to fully evaluate their generalizability (v) Avenues for model updating be provided during model development to facilitate retraining when sufficient quality of data is available in the future to ensure data fairness before implementation. Alternatively, online learning techniques may be adopted.

Abbreviations

AI	Artificial intelligence
EPV	Event-per-variable
PROBAST	Prediction model Risk of Bias Assessment
ML	Machine learning
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-analysis
PROBAST	Prediction model risk of bias assessment tool
STARD	Standards of Reporting Diagnostic Accuracy
TRIPOD	Transparent reporting of a multivariate prediction model for individual prognosis or diagnosis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40537-023-00703-w>.

Additional file 1. Search keywords.

Acknowledgements

Not applicable.

Author contributions

JA conceptualized the study, performed the literature search, synthesized the data, and wrote the original draft of the manuscript. LH performed the literature search and assisted in manuscript review and editing. Y-XS assisted in data interpretation, provided supervision and resources, and assisted in manuscript review and editing. All authors approved the final manuscript version.

Funding

No external funding.

Availability of data and materials

No new data was generated in this study.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 21 July 2022 Accepted: 23 February 2023

Published online: 04 March 2023

References

- Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209–49.
- Johnson DE, Burtneß B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. *Nat Rev Dis Primers*. 2020;6(1):92.
- Pfister DG, Ang K-K, Brizel DM, Burtneß BA, Cmelak AJ, Colevas AD, et al. Head and neck cancers. *J Natl Compr Canc Netw*. 2011;9(6):596–650.
- Adeoye J, Thomson P. Strategies to improve diagnosis and risk assessment for oral cancer patients. *Faculty Dental J*. 2020;11(3):122–7.
- Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. *Oral Oncol*. 2009;45(4–5):309–16.
- Adeoye J, Tan JY, Choi S-W, Thomson P. Prediction models applying machine learning to oral cavity cancer outcomes: A systematic review. *Int J Med Inform*. 2021;154: 104557.
- Volpe S, Pepa M, Zaffaroni M, Bellerba F, Santamaria R, Marvaso G, et al. Machine Learning for Head and Neck Cancer: A Safe Bet?—A Clinically Oriented Systematic Review for the Radiation Oncologist. *Front Oncol*. 2021;11:89.
- Mahmood H, Shaban M, Rajpoot N, Khurram SA. Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. *Br J Cancer*. 2021;124(12):1934–40.
- Chinnery T, Arifin A, Tay KY, Leung A, Nichols AC, Palma DA, et al. Utilizing artificial intelligence for head and neck cancer outcomes prediction from imaging. *Can Assoc Radiol J*. 2020;72(1):73–85.
- Abdel Razek AAK, Khaled R, Helmy E, Naglah A, AbdelKhalek A, El-Baz A. Artificial intelligence and deep learning of head and neck cancer. *Magnetic Resonance Imaging Clinics*. 2022;30(1):81–94.
- García-Pola M, Pons-Fuster E, Suárez-Fernández C, Seoane-Romero J, Romero-Méndez A, López-Jornet P. Role of artificial intelligence in the early diagnosis of oral cancer: a scoping review. *Cancers*. 2021;13(18):4600.
- Alabi RO, Youssef O, Pirinen M, Elmusrati M, Mäkitie AA, Leivo I, et al. Machine learning in oral squamous cell carcinoma: Current status, clinical concerns and prospects for future—A systematic review. *Artif Intell Med*. 2021;115: 102060.
- de Hond AAH, Leeuwenberg AM, Hoofst L, Kant IMJ, Nijman SWJ, van Os HJA, et al. Guidelines and quality criteria for artificial intelligence-based prediction models in healthcare: a scoping review. *Digital Med*. 2022;5(1):2.
- Budach L, Feuerpeil M, Ihde N, Nathansen A, Noack N, Patzlaff H, et al. The Effects of Data Quality on Machine Learning Performance. *arXiv preprint arXiv:220714529*. 2022.
- Luca AR, Ursuleanu TF, Gheorghe L, Grigorovici R, Iancu S, Hlusneac M, et al. Impact of quality, type and volume of data used by deep learning models in the analysis of medical images. *Inform Med Unlocked*. 2022;29: 100911.
- Zhou L, Pan S, Wang J, Vasilakos AV. Machine learning on big data: Opportunities and challenges. *Neurocomputing*. 2017;237:350–61.
- Verma AA, Murray J, Greiner R, Cohen JP, Shojania KG, Ghassemi M, et al. Implementing machine learning in medicine. *CMAJ*. 2021;193(34):E1351–7.
- Bertsimas D, Wiberg H. Machine learning in oncology: methods, applications, and challenges. *JCO Clin Cancer Inform*. 2020;4:885–94.
- Tran KA, Kondrashova O, Bradley A, Williams ED, Pearson JV, Waddell N. Deep learning in cancer diagnosis, prognosis and treatment selection. *Genome Med*. 2021;13(1):152.
- Kourou K, Exarchos KP, Papaloukas C, Sakaloglou P, Exarchos T, Fotiadis DI. Applied machine learning in cancer research: A systematic review for patient diagnosis, classification and prognosis. *Comput Struct Biotechnol J*. 2021;19:5546–55.
- Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J*. 2015;13:8–17.
- Abdel Razek AAK, Khaled R, Helmy E, Naglah A, AbdelKhalek A, El-Baz A. Artificial intelligence and deep learning of head and neck cancer. *Magn Reson Imaging Clin N Am*. 2022;30(1):81–94.
- Adeoye J, Akinshipo A, Koohi-Moghadam M, Thomson P, Su Y-X. Construction of machine learning-based models for cancer outcomes in low and lower-middle income countries: A scoping review. *Front Oncol*. 2022;12:89.
- Patil S, Habib Awan K, Arakeri G, Jayampath Seneviratne C, Muddur N, Malik S, et al. Machine learning and its potential applications to the genomic study of head and neck cancer—A systematic review. *J Oral Pathol Med*. 2019;48(9):773–9.
- Volpe S, Pepa M, Zaffaroni M, Bellerba F, Santamaria R, Marvaso G, et al. Machine learning for head and neck cancer: a safe bet?—a clinically oriented systematic review for the radiation oncologist. *Front Oncol*. 2021;11: 772663.

26. Bassani S, Santonicco N, Eccher A, Scarpa A, Vianini M, Brunelli M, et al. Artificial intelligence in head and neck cancer diagnosis. *J Pathol Inform.* 2022;13:100153.
27. Giannitto C, Mercante G, Ammirabile A, Cerri L, De Giorgi T, Lofino L, et al. Radiomics-based machine learning for the diagnosis of lymph node metastases in patients with head and neck cancer: Systematic review. *Head Neck.* 2023;45(2):482–91.
28. Alabi RO, Bello IO, Youssef O, Elmusrati M, Mäkitie AA, Almangush A. Utilizing deep machine learning for prognostication of oral squamous cell carcinoma—a systematic review. *Front Oral Health.* 2021;2:78.
29. Chiesa-Estomba CM, Graña M, Medela A, Sistiaga-Suarez JA, Lechien JR, Calvo-Henriquez C, et al. Machine learning algorithms as a computer-assisted decision tool for oral cancer prognosis and management decisions: a systematic review. *ORL J Otorhinolaryngol Relat Spec.* 2022;84(4):278–88.
30. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
31. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Cancer.* 2015;112(2):251–9.
32. Moons KG, Wolff RF, Riley RD, Whiting PF, Westwood M, Collins GS, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration. *Ann Intern Med.* 2019;170(1):W1–33.
33. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372: n71.
34. Adeoye J, Hui L, Koochi-Moghadam M, Tan JY, Choi SW, Thomson P. Comparison of time-to-event machine learning models in predicting oral cavity cancer prognosis. *Int J Med Inform.* 2022;157: 104635.
35. Adeoye J, Koochi-Moghadam M, Lo AWI, Tsang RK, Chow VLY, Zheng LW, et al. Deep learning predicts the malignant-transformation-free survival of oral potentially malignant disorders. *Cancers (Basel).* 2021;13(23):78.
36. Agarwal P, Yadav A, Mathur P, Pal V, Chakrabarty A. BID-Net: an automated system for bone invasion detection occurring at stage T4 in oral squamous carcinoma using deep learning. *Comput Intell Neurosci.* 2022;2022:4357088.
37. Akcay M, Etiz D, Celik O, Ozen A. Evaluation of prognosis in nasopharyngeal cancer using machine learning. *Technol Cancer Res Treat.* 2020;19:1533033820909829.
38. Alabi RO, Elmusrati M, Sawazaki-Calone I, Kowalski LP, Haglund C, Coletta RD, et al. Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *Int J Med Inform.* 2020;136: 104068.
39. Alhazmi A, Alhazmi Y, Makrami A, Masmali A, Salawi N, Masmali K, et al. Application of artificial intelligence and machine learning for prediction of oral cancer risk. *J Oral Pathol Med.* 2021;50(5):444–50.
40. Alkhadar H, Macluskey M, White S, Ellis I, Gardner A. Comparison of machine learning algorithms for the prediction of five-year survival in oral squamous cell carcinoma. *J Oral Pathol Med.* 2021;50(4):378–84.
41. Arijji Y, Kise Y, Fukuda M, Kuwada C, Arijji E. Segmentation of metastatic cervical lymph nodes from CT images of oral cancers using deep-learning technology. *Dentomaxillofac Radiol.* 2022;51(4):20210515.
42. Ashizawa K, Yoshimura K, John H, Inoue T, Katoh R, Funayama S, et al. Construction of mass spectra database and diagnosis algorithm for head and neck squamous cell carcinoma. *Oral Oncol.* 2017;75:111–9.
43. Aubreville M, Knipfer C, Oetter N, Jaremenko C, Rodner E, Denzler J, et al. Automatic Classification of Cancerous Tissue in Laserendoscopic Images of the Oral Cavity using Deep Learning. *Sci Rep.* 2017;7(1):11979.
44. Azam MA, Sampieri C, Ioppi A, Africano S, Vallin A, Mocellin D, et al. Deep learning applied to white light and narrow band imaging video laryngoscopy: toward real-time laryngeal cancer detection. *Laryngoscope.* 2021;89:56.
45. Banavar G, Ogundijo O, Toma R, Rajagopal S, Lim YK, Tang K, et al. The salivary metatranscriptome as an accurate diagnostic indicator of oral cancer. *Genomic Med.* 2021;6(1):78.
46. Bhowmik A, Ghosh B, Pal M, Paul RR, Chatterjee J, Chakraborty S. Portable, handheld, and affordable blood perfusion imager for screening of subsurface cancer in resource-limited settings. *Proc Natl Acad Sci U S A.* 2022;119:2.
47. Bielak L, Wiedenmann N, Berlin A, Nicolay NH, Gunashekar DD, Hägele L, et al. Convolutional neural networks for head and neck tumor segmentation on 7-channel multiparametric MRI: a leave-one-out analysis. *Radiat Oncol.* 2020;15(1):181.
48. Bogowicz M, Tanadini-Lang S, Guckenberger M, Riesterer O. Combined CT radiomics of primary tumor and metastatic lymph nodes improves prediction of locoregional control in head and neck cancer. *Sci Rep.* 2019;9:89.
49. Bos P, van den Brekel MWM, Gouw ZAR, Al-Mamgani A, Taghavi M, Waktola S, et al. Improved outcome prediction of oropharyngeal cancer by combining clinical and MRI features in machine learning models. *Eur J Radiol.* 2021;139: 109701.
50. Braz DC, Neto MP, Shimizu FM, Sá AC, Lima RS, Gobbi AL, et al. Using machine learning and an electronic tongue for discriminating saliva samples from oral cavity cancer patients and healthy individuals. *Talanta.* 2022;243: 123327.
51. Brouwer de Koning SG, Baltussen EJM, Karakullukcu MB, Dashtbozorg B, Smit LA, Dirven R, et al. Toward complete oral cavity cancer resection using a handheld diffuse reflectance spectroscopy probe. *J Biomed Opt.* 2018;23(12):1–8.
52. Cai MW, Wang JZ, Yang Q, Guo Y, Zhang Z, Ying HM, et al. Combining Images and T-staging information to improve the automatic segmentation of nasopharyngeal carcinoma tumors in MR Images. *IEEE ACCESS.* 2021;9:21323–31.
53. Carnielli CM, Macedo CCS, De Rossi T, Granato DC, Rivera C, Domingues RR, et al. Combining discovery and targeted proteomics reveals a prognostic signature in oral cancer. *Nat Commun.* 2018;9(1):3598.
54. Cegla P, Currie G, Wroblewska JP, Cholewinski W, Kazmierska J, Marszalek A, et al. Influence of Semiquantitative [F-18]FDG PET and Hematological Parameters on Survival in HNSCC Patients Using Neural Network Analysis. *Pharmaceuticals.* 2022;15(2):8.
55. Chang CC, Huang TH, Shueng PW, Chen SH, Chen CC, Lu CJ, et al. Developing a stacked ensemble-based classification scheme to predict second primary cancers in head and neck cancer survivors. *Int J Environ Res Public Health.* 2021;18:23.

56. Chang YJ, Huang TY, Liu YJ, Chung HW, Juan CJ. Classification of parotid gland tumors by using multimodal MRI and deep learning. *NMR Biomed*. 2021;34(1): e4408.
57. Chen C, Qin Y, Chen H, Cheng J, He B, Wan Y, et al. Machine learning to differentiate small round cell malignant tumors and non-small round cell malignant tumors of the nasal and paranasal sinuses using apparent diffusion coefficient values. *Eur Radiol*. 2022;32(6):3819–29.
58. Chen H, Qi YX, Yin Y, Li TX, Liu XQ, Li XL, et al. MMFNet: A multi-modality MRI fusion network for segmentation of nasopharyngeal carcinoma. *Neurocomputing*. 2020;394:27–40.
59. Chen L, Dohopolski M, Zhou Z, Wang K, Wang R, Sher D, et al. Attention guided lymph node malignancy prediction in head and neck cancer. *Int J Radiat Oncol Biol Phys*. 2021;110(4):1171–9.
60. Chen MY, Chen JW, Wu LW, Huang KC, Chen JY, Wu WS, et al. Carcinogenesis of male oral submucous fibrosis alters salivary microbiomes. *J Dent Res*. 2021;100(4):397–405.
61. Chen X, Li Y, Li X, Cao X, Xiang Y, Xia W, et al. An interpretable machine learning prognostic system for locoregionally advanced nasopharyngeal carcinoma based on tumor burden features. *Oral Oncol*. 2021;118: 105335.
62. Cho SI, Sun S, Mun JH, Kim C, Kim SY, Cho S, et al. Dermatologist-level classification of malignant lip diseases using a deep convolutional neural network. *Br J Dermatol*. 2020;182(6):1388–94.
63. Choi N, Kim Z, Song BH, Park W, Chung MJ, Cho BH, et al. Prediction of risk factors for pharyngo-cutaneous fistula after total laryngectomy using artificial intelligence. *Oral Oncol*. 2021;119: 105357.
64. Chu YW, Chen F, Tang Y, Chen T, Yu YX, Jin HL, et al. Diagnosis of nasopharyngeal carcinoma from serum samples using hyperspectral imaging combined with a chemometric method. *Opt Express*. 2018;26(22):28661–71.
65. Chuang WY, Chang SH, Yu WH, Yang CK, Yeh CJ, Ueng SH, et al. Successful Identification of Nasopharyngeal Carcinoma in Nasopharyngeal Biopsies Using Deep Learning. *Cancers (Basel)*. 2020;12(2):8.
66. Corino VDA, Bologna M, Calareso G, Resteghini C, Sdao S, Orlandi E, et al. Refining Tumor Treatment in Sinonasal Cancer Using Delta Radiomics of Multi-Parametric MRI after the First Cycle of Induction Chemotherapy. *J Imaging*. 2022;8:2.
67. Cui C, Wang S, Zhou J, Dong A, Xie F, Li H, et al. Machine learning analysis of image data based on detailed MR image reports for nasopharyngeal carcinoma prognosis. *Biomed Res Int*. 2020;2020:8068913.
68. D'Urso P, Farneti A, Marucci L, Marzi S, Piludu F, Vidiri A, et al. Predictors of outcome after (chemo)radiotherapy for node-positive oropharyngeal cancer: the role of functional MRI. *Cancers*. 2022;14(10):2477.
69. da Costa NL, de Sa AM, de Sa RN, Bandeira CM, Oliveira Alves MG, Mendes MA, et al. Finding the combination of multiple biomarkers to diagnose oral squamous cell carcinoma - a data mining approach. *Comput Biol Med*. 2022;143: 105296.
70. Das N, Hussain E, Mahanta LB. Automated classification of cells into multiple classes in epithelial tissue of oral squamous cell carcinoma using transfer learning and convolutional neural network. *Neural Netw*. 2020;128:47–60.
71. De Martin A, Lütge M, Stanossek Y, Engetschwiler C, Cupovic J, Brown K, et al. Distinct microbial communities colonize tonsillar squamous cell carcinoma. *Oncoimmunology*. 2021;10(1):1945202.
72. Deng W, Luo L, Lin X, Fang T, Liu D, Dan G, et al. Head and neck cancer tumor segmentation using support vector machine in dynamic contrast-enhanced MRI. *Contrast Media Mol Imaging*. 2017;2017:8612519.
73. Deng Y, Li C, Lv X, Xia W, Shen L, Jing B, et al. The contrast-enhanced MRI can be substituted by unenhanced MRI in identifying and automatically segmenting primary nasopharyngeal carcinoma with the aid of deep learning models: an exploratory study in large-scale population of endemic area. *Comput Methods Programs Biomed*. 2022;217: 106702.
74. Devakumar D, Sunny G, Sasidharan BK, Bowen SR, Nadaraj A, Jeyseelan L, et al. Framework for machine learning of CT and PET radiomics to predict local failure after radiotherapy in locally advanced head and neck cancers. *J Med Phys*. 2021;46(3):181–8.
75. Diao S, Hou J, Yu H, Zhao X, Sun Y, Lambo RL, et al. Computer-aided pathologic diagnosis of nasopharyngeal carcinoma based on deep learning. *Am J Pathol*. 2020;190(8):1691–700.
76. Dinia A, Ammari S, Filtes J, Classe M, Moya-Plana A, Bidault F, et al. Events prediction after treatment in HPV-driven oropharyngeal carcinoma using machine learning. *Eur J Cancer*. 2022;171:106–13.
77. Dittberner A, Rodner E, Ortman W, Stadler J, Schmidt C, Petersen I, et al. Automated analysis of confocal laser endomicroscopy images to detect head and neck cancer. *Head Neck*. 2016;38(Suppl 1):E1419–26.
78. Dohopolski M, Wang K, Morgan H, Sher D, Wang J. Use of deep learning to predict the need for aggressive nutritional supplementation during head and neck radiotherapy. *Radiother Oncol*. 2022;171:129–38.
79. Dong F, Tao C, Wu J, Su Y, Wang Y, Wang Y, et al. Detection of cervical lymph node metastasis from oral cavity cancer using a non-radiating, noninvasive digital infrared thermal imaging system. *Sci Rep*. 2018;8(1):7219.
80. Du R, Lee VH, Yuan H, Lam KO, Pang HH, Chen Y, et al. Radiomics model to predict early progression of nonmetastatic nasopharyngeal carcinoma after intensity modulation radiation therapy: a multicenter study. *Radiol Artif Intell*. 2019;1(4): e180075.
81. Farrokhian N, Holcomb AJ, Dimon E, Karadaghy O, Ward C, Whiteford E, et al. Development and validation of machine learning models for predicting occult nodal metastasis in early-stage oral cavity squamous cell carcinoma. *JAMA Netw Open*. 2022;5(4): e227226.
82. Fei Y, Zhang F, Zu C, Hong M, Peng X, Xiao J, et al. MRF-RFS: A modified random forest recursive feature selection algorithm for nasopharyngeal carcinoma segmentation. *Methods Inf Med*. 2020;59(4–05):151–61.
83. Florez E, Thomas TV, Howard CM, Khosravi HR, Lirette ST, Fatemi A. Machine learning based on ct radiomic features predicts residual tumor in head and neck cancer patients treated with chemoradiotherapy. *Biomed Sci Instrum*. 2021;57(2):199–211.
84. Folkert MR, Setton J, Apte AP, Grkovski M, Young RJ, Schöder H, et al. Predictive modeling of outcomes following definitive chemoradiotherapy for oropharyngeal cancer based on FDG-PET image characteristics. *Phys Med Biol*. 2017;62(13):5327–43.
85. Fontaine P, Acosta O, Castelli J, De Crevoisier R, Müller H, Depeursinge A. The importance of feature aggregation in radiomics: a head and neck cancer study. *Sci Rep*. 2020;10(1):19679.

86. Formeister EJ, Baum R, Knott PD, Seth R, Ha P, Ryan W, et al. Machine learning for predicting complications in head and neck microvascular free tissue transfer. *Laryngoscope*. 2020;130(12):E843–9.
87. Fouad S, Landini G, Robinson M, Song TH, Mehanna H. Human papilloma virus detection in oropharyngeal carcinomas with in situ hybridisation using hand crafted morphological features and deep central attention residual networks. *Comput Med Imaging Graph*. 2021;88: 101853.
88. Fu Q, Chen Y, Li Z, Jing Q, Hu C, Liu H, et al. A deep learning algorithm for detection of oral cavity squamous cell carcinoma from photographic images: a retrospective study. *EClinicalMedicine*. 2020;27: 100558.
89. Fujima N, Andreu-Arasa VC, Meibom SK, Mercier GA, Salama AR, Truong MT, et al. Deep learning analysis using FDG-PET to predict treatment outcome in patients with oral cavity squamous cell carcinoma. *Eur Radiol*. 2020;30(11):6322–30.
90. Fujima N, Andreu-Arasa VC, Meibom SK, Mercier GA, Truong MT, Hirata K, et al. Prediction of the local treatment outcome in patients with oropharyngeal squamous cell carcinoma using deep learning analysis of pretreatment FDG-PET images. *BMC Cancer*. 2021;21(1):900.
91. Gangil T, Shahabuddin AB, Rao BD, Palanisamy K, Chakrabarti B, Sharan K. Predicting clinical outcomes of radiotherapy for head and neck squamous cell carcinoma patients using machine learning algorithms. *J Big Data*. 2022;9:1.
92. González-García I, Pierre V, Dubois VFS, Morsli N, Spencer S, Baverel PG, et al. Early predictions of response and survival from a tumor dynamics model in patients with recurrent, metastatic head and neck squamous cell carcinoma treated with immunotherapy. *CPT Pharmacometrics Syst Pharmacol*. 2021;10(3):230–40.
93. Grillone GA, Wang Z, Krisciunas GP, Tsai AC, Kannabiran VR, Pistey RW, et al. The color of cancer: Margin guidance for oral cancer resection using elastic scattering spectroscopy. *Laryngoscope*. 2017;127(4):S1–9.
94. Groendahl AR, Skjei Knudtsen I, Huynh BN, Mulstad M, Moe YM, Knuth F, et al. A comparison of methods for fully automatic segmentation of tumors and involved nodes in PET/CT of head and neck cancers. *Phys Med Biol*. 2021;66(6): 065012.
95. Gunduz E, Alçin OF, Kizilay A, Yildirim IO. Deep learning model developed by multiparametric MRI in differential diagnosis of parotid gland tumors. *Eur Arch Otorhinolaryngol*. 2022;89:7.
96. Guo Y, Jiang W, Lakshminarayanan P, Han P, Cheng Z, Bowers M, et al. Spatial radiation dose influence on xerostomia recovery and its comparison to acute incidence in patients with head and neck cancer. *Adv Radiat Oncol*. 2020;5(2):221–30.
97. Halicek M, Dormer JD, Little JV, Chen AY, Myers L, Sumer BD, et al. Hyperspectral imaging of head and neck squamous cell carcinoma for cancer margin detection in surgical specimens from 102 patients using deep learning. *Cancers (Basel)*. 2019;11(9):340.
98. He S, Lian C, Thorstad W, Gay H, Zhao Y, Ruan S, et al. A novel systematic approach for cancer treatment prognosis and its applications in oropharyngeal cancer with microRNA biomarkers. *Bioinformatics*. 2021;5:98.
99. He YR, Cheng YD, Huang ZG, Xu W, Hu R, Cheng LY, et al. A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis. *Ann Transl Med*. 2021;9:24.
100. Hellwig K, Ellmann S, Eckstein M, Wiesmueller M, Rutzner S, Semrau S, et al. Predictive value of multiparametric MRI for response to single-cycle induction chemo-immunotherapy in locally advanced head and neck squamous cell carcinoma. *Front Oncol*. 2021;11: 734872.
101. Heo J, Lim JH, Lee HR, Jang JY, Shin YS, Kim D, et al. Deep learning model for tongue cancer diagnosis using endoscopic images. *Sci Rep*. 2022;12(1):6281.
102. Hsu CW, Chen YT, Hsieh YJ, Chang KP, Hsueh PC, Chen TW, et al. Integrated analyses utilizing metabolomics and transcriptomics reveal perturbation of the polyamine pathway in oral cavity squamous cell carcinoma. *Anal Chim Acta*. 2019;1050:113–22.
103. Huang B, Chen Z, Wu PM, Ye Y, Feng ST, Wong CO, et al. Fully Automated Delineation of Gross Tumor Volume for Head and Neck Cancer on PET-CT Using Deep Learning: A Dual-Center Study. *Contrast Media Mol Imaging*. 2018;2018:8923028.
104. Humbert-Vidan L, Patel V, Oksuz I, King AP, Guerrero UT. Comparison of machine learning methods for prediction of osteoradionecrosis incidence in patients with head and neck cancer. *Br J Radiol*. 2021;94(1120):20200026.
105. Inaba A, Hori K, Yoda Y, Ikematsu H, Takano H, Matsuzaki H, et al. Artificial intelligence system for detecting superficial laryngopharyngeal cancer with high efficiency of deep learning. *Head Neck*. 2020;42(9):2581–92.
106. Intarak S, Chongpison Y, Vimolnoch M, Oonsiri S, Kitpanit S, Prayongrat A, et al. Tumor Prognostic Prediction of Nasopharyngeal Carcinoma Using CT-Based Radiomics in Non-Chinese Patients. *Front Oncol*. 2022;12: 775248.
107. Jeng MJ, Sharma M, Sharma L, Chao TY, Huang SF, Chang LB, et al. Raman spectroscopy analysis for optical diagnosis of oral cancer detection. *J Clin Med*. 2019;8:9.
108. Jiang R, You R, Pei XQ, Zou X, Zhang MX, Wang TM, et al. Development of a ten-signature classifier using a support vector machine integrated approach to subdivide the M1 stage into M1a and M1b stages of nasopharyngeal carcinoma with synchronous metastases to better predict patients' survival. *Oncotarget*. 2016;7(3):3645–57.
109. Kang L, Niu YL, Huang R, Lin SF, Tang QL, Chen AL, et al. Predictive value of a combined model based on pre-treatment and mid-treatment MRI-radiomics for disease progression or death in locally advanced nasopharyngeal carcinoma. *Front Oncol*. 2021;11:89.
110. Kann BH, Aneja S, Loganadane GV, Kelly JR, Smith SM, Decker RH, et al. Pretreatment identification of head and neck cancer nodal metastasis and extranodal extension using deep learning neural networks. *Sci Rep*. 2018;8(1):14036.
111. Khodrog OA, Cui F, Xu N, Han Q, Liu J, Gong T, et al. Prediction of squamous cell carcinoma cases from squamous cell hyperplasia in throat lesions using CT radiomics model. *Saudi Med J*. 2021;42(3):284–92.
112. Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep*. 2019;9(1):6994.
113. Koenen L, Arens P, Olze H, Dommerich S. Classifying and predicting surgical complications after laryngectomy: a novel approach to diagnosing and treating patients. *Ent-Ear Nose Throat J*. 2021;7:67.

114. Kono M, Ishihara R, Kato Y, Miyake M, Shoji A, Inoue T, et al. Diagnosis of pharyngeal cancer on endoscopic video images by Mask region-based convolutional neural network. *Dig Endosc.* 2021;33(4):569–76.
115. Kubo K, Kawahara D, Murakami Y, Takeuchi Y, Katsuta T, Imano N, et al. Development of a radiomics and machine learning model for predicting occult cervical lymph node metastasis in patients with tongue cancer. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2022;134(1):93–101.
116. Lam SK, Zhang J, Zhang YP, Li B, Ni RY, Zhou T, et al. A Multi-Center Study of CT-Based Neck Nodal Radiomics for Predicting an Adaptive Radiotherapy Trigger of Ill-Fitted Thermoplastic Masks in Patients with Nasopharyngeal Carcinoma. *Life (Basel).* 2022;12:22.
117. Lau K, Wilkinson J, Moorhy R. A web-based prediction score for head and neck cancer referrals. *Clin Otolaryngol.* 2018;43(4):1043–9.
118. Lee JY, Shi T, Petyuk VA, Schepmoes AA, Fillmore TL, Wang YT, et al. Detection of head and neck cancer based on longitudinal changes in serum protein abundance. *Cancer Epidemiol Biomarkers Prev.* 2020;29(8):1665–72.
119. Leger S, Zwanenburg A, Pilz K, Lohaus F, Linde A, Zöphel K, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Sci Rep.* 2017;7(1):13206.
120. Leng X, Fang P, Lin H, Qin C, Tan X, Liang Y, et al. Application of a machine learning method to whole brain white matter injury after radiotherapy for nasopharyngeal carcinoma. *Cancer Imaging.* 2019;19(1):19.
121. Li S, Deng YQ, Hua HL, Li SL, Chen XX, Xie BJ, et al. Deep learning for locally advanced nasopharyngeal carcinoma prognostication based on pre- and post-treatment MRI. *Comput Methods Programs Biomed.* 2022;219: 106785.
122. Li S, Xiao J, He L, Peng X, Yuan X. The tumor target segmentation of nasopharyngeal cancer in CT images based on deep learning methods. *Technol Cancer Res Treat.* 2019;18:1533033819884561.
123. Li Z, Li Z, Chen Q, Zhang J, Dunham ME, McWhorter AJ, et al. Machine-learning-assisted spontaneous Raman spectroscopy classification and feature extraction for the diagnosis of human laryngeal cancer. *Comput Biol Med.* 2022;146: 105617.
124. Liao H, Chen X, Lu S, Jin G, Pei W, Li Y, et al. MRI-based back propagation neural network model as a powerful tool for predicting the response to induction chemotherapy in locoregionally advanced nasopharyngeal carcinoma. *J Magn Reson Imaging.* 2021;56:3.
125. Lin H, Chen H, Weng L, Shao J, Lin J. Automatic detection of oral cancer in smartphone-based images using deep learning for early diagnosis. *J Biomed Opt.* 2021;26:8.
126. Liu GS, Yang A, Kim D, Hojel A, Voevodsky D, Wang J, et al. Deep learning classification of inverted papilloma malignant transformation using 3D convolutional neural networks and magnetic resonance imaging. *Int Forum Allergy Rhinol.* 2022;8:56.
127. Liu J, Mao Y, Li ZJ, Zhang DK, Zhang ZC, Hao SN, et al. Use of texture analysis based on contrast-enhanced MRI to predict treatment response to chemoradiotherapy in nasopharyngeal carcinoma. *J Magn Reson Imaging.* 2016;44(2):445–55.
128. Liu X, Pan Y, Zhang X, Sha Y, Wang S, Li H, et al. A Deep Learning Model for Classification of Parotid Neoplasms Based on Multimodal Magnetic Resonance Image Sequences. *Laryngoscope.* 2022;7:e43.
129. Liu Y, Li Y, Fu Y, Liu T, Liu X, Zhang X, et al. Quantitative prediction of oral cancer risk in patients with oral leukoplakia. *Oncotarget.* 2017;8(28):46057–64.
130. Lu C, Lewis JS Jr, Dupont WD, Plummer WD Jr, Janowczyk A, Madabhushi A. An oral cavity squamous cell carcinoma quantitative histomorphometric-based image classifier of nuclear morphology can risk stratify patients for disease-specific survival. *Mod Pathol.* 2017;30(12):1655–65.
131. Marschner SN, Lombardo E, Minibek L, Holzgreve A, Kaiser L, Albert NL, et al. Risk Stratification Using (18)F-FDG PET/CT and Artificial Neural Networks in Head and Neck Cancer Patients Undergoing Radiotherapy. *Diagnostics (Basel).* 2021;11:9.
132. Matsuo H, Nishio M, Kanda T, Kojita Y, Kono AK, Hori M, et al. Diagnostic accuracy of deep-learning with anomaly detection for a small amount of imbalanced data: discriminating malignant parotid tumors in MRI. *Sci Rep.* 2020;10(1):19388.
133. McRae MP, Modak SS, Simmons GW, Trocheset DA, Kerr AR, Thornhill MH, et al. Point-of-care oral cytology tool for the screening and assessment of potentially malignant oral lesions. *Cancer Cytopathol.* 2020;128(3):207–20.
134. Men K, Chen X, Zhu J, Yang B, Zhang Y, Yi J, et al. Continual improvement of nasopharyngeal carcinoma segmentation with less labeling effort. *Phys Med.* 2020;80:347–51.
135. Men K, Geng H, Zhong H, Fan Y, Lin A, Xiao Y. A deep learning model for predicting xerostomia due to radiation therapy for head and neck squamous cell carcinoma in the RTOG 0522 Clinical Trial. *Int J Radiat Oncol Biol Phys.* 2019;105(2):440–7.
136. Mentel S, Gallo K, Wagendorf O, Preissner R, Nahles S, Heiland M, et al. Prediction of oral squamous cell carcinoma based on machine learning of breath samples: a prospective controlled study. *BMC Oral Health.* 2021;21(1):500.
137. Mermod M, Jourdan EF, Gupta R, Bongiovanni M, Tolstonog G, Simon C, et al. Development and validation of a multivariable prediction model for the identification of occult lymph node metastasis in oral squamous cell carcinoma. *Head Neck.* 2020;42(8):1811–20.
138. Min Park Y, Yol Lim J, Woo Koh Y, Kim SH, Chang CE. Prediction of treatment outcome using MRI radiomics and machine learning in oropharyngeal cancer patients after surgical treatment. *Oral Oncol.* 2021;122: 105559.
139. Mohammed MA, Abd Ghani MK, Arunkumar N, Mostafa SA, Abdullah MK, Burhanuddin MA. Trainable model for segmenting and identifying Nasopharyngeal carcinoma. *Comput Electr Eng.* 2018;71:372–87.
140. Moor JW, Paleri V, Edwards J. Patient classification of two-week wait referrals for suspected head and neck cancer: a machine learning approach. *J Laryngol Otol.* 2019;133(10):875–8.
141. Musulin J, Štifanić D, Zulijani A, Čabov T, Dekanić A, Car Z. An Enhanced Histopathology Analysis: An AI-Based System for Multiclass Grading of Oral Squamous Cell Carcinoma and Segmenting of Epithelial and Stromal Tissue. *Cancers (Basel).* 2021;13(8):8.

142. Nawandhar A, Kumar N, Veena R, Yamujala L. Stratified squamous epithelial biopsy image classifier using machine learning and neighborhood feature selection. *Biomedical Signal Processing Control*. 2020;55:9.
143. Pan X, Zhang T, Yang Q, Yang D, Rwigema JC, Qi XS. Survival prediction for oral tongue cancer patients via probabilistic genetic information optimized neural network models. *Br J Radiol*. 2020;93(1112):20190825.
144. Panigrahi S, Bhuyan R, Kumar K, Nayak J, Swarnkar T. Multistage classification of oral histopathological images using improved residual network. *Math Biosci Eng*. 2022;19(2):1909–25.
145. Paul A, Srivastava S, Roy R, Anand A, Gaurav K, Husain N, et al. Malignancy prediction among tissues from Oral SCC patients including neck invasions: a (1)H HRMAS NMR based metabolomic study. *Metabolomics*. 2020;16(3):38.
146. Prezioso E, Izzo S, Giampaolo F, Piccialli F, Dell'Aversana Orabona G, Cuocolo R, et al. Predictive Medicine for Salivary gland tumours identification through Deep Learning. *IEEE J Biomed Health Inform*. 2021;87:4.
147. Qi Y, Li J, Chen H, Guo Y, Yin Y, Gong G, et al. Computer-aided diagnosis and regional segmentation of nasopharyngeal carcinoma based on multi-modality medical images. *Int J Comput Assist Radiol Surg*. 2021;16(6):871–82.
148. Raghavan Nair JK, Vallières M, Mascarella MA, El Sabbagh N, Duchatellier CF, Zeitouni A, et al. Magnetic resonance imaging texture analysis predicts recurrence in patients with nasopharyngeal carcinoma. *Can Assoc Radiol J*. 2019;70(4):394–402.
149. Rahman TY, Mahanta LB, Das AK, Sarma JD. Automated oral squamous cell carcinoma identification using shape, texture and color features of whole image strips. *Tissue Cell*. 2020;63: 101322.
150. Ren J, Eriksen JG, Nijkamp J, Korreman SS. Comparing different CT, PET and MRI multi-modality image combinations for deep learning-based head and neck tumor segmentation. *Acta Oncol*. 2021;60(11):1399–406.
151. Ren J, Jing X, Wang J, Ren X, Xu Y, Yang Q, et al. Automatic recognition of laryngoscopic images using a deep-learning technique. *Laryngoscope*. 2020;130(11):E686–93.
152. Ren J, Qi M, Yuan Y, Duan S, Tao X. Machine learning-based MRI texture analysis to predict the histologic grade of oral squamous cell carcinoma. *AJR Am J Roentgenol*. 2020;215(5):1184–90.
153. Rodríguez Outeiral R, Bos P, Al-Mamgani A, Jasperse B, Simões R, van der Heide UA. Oropharyngeal primary tumor segmentation for radiotherapy planning on magnetic resonance imaging using deep learning. *Phys Imaging Radiat Oncol*. 2021;19:39–44.
154. Shaban M, Khurram SA, Fraz MM, Alsubaie N, Masood I, Mushtaq S, et al. A novel digital score for abundance of tumour infiltrating lymphocytes predicts disease free survival in oral squamous cell carcinoma. *Sci Rep*. 2019;9(1):13341.
155. Shan J, Jiang R, Chen X, Zhong Y, Zhang W, Xie L, et al. Machine learning predicts lymph node metastasis in early-stage oral tongue squamous cell carcinoma. *J Oral Maxillofac Surg*. 2020;78(12):2208–18.
156. Shao S, Mao N, Liu W, Cui J, Xue X, Cheng J, et al. Epithelial salivary gland tumors: utility of radiomics analysis based on diffusion-weighted imaging for differentiation of benign from malignant tumors. *J Xray Sci Technol*. 2020;28(4):799–808.
157. Shimpi N, Glurich I, Rostami R, Hegde H, Olson B, Acharya A. Development and validation of a non-invasive, chairside oral cavity cancer risk assessment prototype using machine learning approach. *J Pers Med*. 2022;12(4):8.
158. Shu C, Yan H, Zheng W, Lin K, James A, Selvarajan S, et al. Deep learning-guided fiberoptic raman spectroscopy enables real-time in vivo diagnosis and assessment of nasopharyngeal carcinoma and post-treatment efficacy during endoscopy. *Anal Chem*. 2021;93(31):10898–906.
159. Song LL, Chen SJ, Chen W, Shi Z, Wang XD, Song LN, et al. Radiomic model for differentiating parotid pleomorphic adenoma from parotid adenolymphoma based on MRI images. *BMC Med Imaging*. 2021;21(1):54.
160. Song X, Yang X, Narayanan R, Shankar V, Ethiraj S, Wang X, et al. Oral squamous cell carcinoma diagnosed from saliva metabolic profiling. *Proc Natl Acad Sci U S A*. 2020;117(28):16167–73.
161. Suh CH, Lee KH, Choi YJ, Chung SR, Baek JH, Lee JH, et al. Oropharyngeal squamous cell carcinoma: radiomic machine-learning classifiers from multiparametric MR images for determination of HPV infection status. *Sci Rep*. 2020;10(1):17525.
162. Sun TG, Mao L, Chai ZK, Shen XM, Sun ZJ. Predicting the proliferation of tongue cancer with artificial intelligence in contrast-enhanced CT. *Front Oncol*. 2022;12: 841262.
163. Tekchandani H, Verma S, Londhe ND, Jain RR, Tiwari A. Computer aided diagnosis system for cervical lymph nodes in CT images using deep learning. *Biomedical Signal Processing Control*. 2022;71:7.
164. Teng F, Fan W, Luo Y, Xu S, Gong H, Ge R, et al. A Risk prediction model by LASSO for radiation-induced xerostomia in patients with nasopharyngeal carcinoma treated with comprehensive salivary gland-sparing helical tomotherapy technique. *Front Oncol*. 2021;11: 633556.
165. Tighe D, Fabris F, Freitas A. Machine learning methods applied to audit of surgical margins after curative surgery for head and neck cancer. *Br J Oral Maxillofac Surg*. 2021;59(2):209–16.
166. Tomita H, Yamashiro T, Heianna J, Nakasone T, Kimura Y, Mimura H, et al. Nodal-based radiomics analysis for identifying cervical lymph node metastasis at levels I and II in patients with oral squamous cell carcinoma using contrast-enhanced computed tomography. *Eur Radiol*. 2021;31(10):7440–9.
167. Tomita H, Yamashiro T, Iida G, Tsubakimoto M, Mimura H, Murayama S. Unenhanced CT texture analysis with machine learning for differentiating between nasopharyngeal cancer and nasopharyngeal malignant lymphoma. *Nagoya J Med Sci*. 2021;83(1):135–49.
168. Tosado J, Zdilar L, Elhalawani H, Elgohari B, Vock DM, Marai GE, et al. Clustering of largely right-censored oropharyngeal head and neck cancer patients for discriminative groupings to improve outcome prediction. *Sci Rep*. 2020;10(1):3811.
169. Tseng YJ, Wang HY, Lin TW, Lu JJ, Hsieh CH, Liao CT. Development of a machine learning model for survival risk stratification of patients with advanced oral cancer. *JAMA Netw Open*. 2020;3(8): e2011768.
170. van der Heijden M, Essers PBM, Verhagen CVM, Willems SM, Sanders J, de Roest RH, et al. Epithelial-to-mesenchymal transition is a prognostic marker for patient outcome in advanced stage HNSCC patients treated with chemoradiotherapy. *Radiother Oncol*. 2020;147:186–94.
171. Wada T, Yokota H, Horikoshi T, Starkey J, Hattori S, Hashiba J, et al. Diagnostic performance and inter-operator variability of apparent diffusion coefficient analysis for differentiating pleomorphic adenoma and carcinoma ex pleomorphic

- adenoma: comparing one-point measurement and whole-tumor measurement including radiomics approach. *Jpn J Radiol.* 2020;38(3):207–14.
172. Wang J, Liu R, Zhao Y, Nantavithya C, Elhalawani H, Zhu H, et al. A predictive model of radiation-related fibrosis based on the radiomic features of magnetic resonance imaging and computed tomography. *Transl Cancer Res.* 2020;9(8):4726–38.
 173. Wang T, Hu JY, Huang QT, Wang WW, Zhang XY, Zhang LW, et al. Development of a normal tissue complication probability (NTCP) model using an artificial neural network for radiation-induced necrosis after carbon ion re-irradiation in locally recurrent carcinoma. *Ann Transl Med.* 2021;89:6.
 174. Wang X, Yang J, Wei C, Zhou G, Wu L, Gao Q, et al. A personalized computational model predicts cancer risk level of oral potentially malignant disorders and its web application for promotion of non-invasive screening. *J Oral Pathol Med.* 2020;49(5):417–26.
 175. Wang Y, Xie W, Huang S, Feng M, Ke X, Zhong Z, et al. The diagnostic value of ultrasound-based deep learning in differentiating parotid gland tumors. *J Oncol.* 2022;2022:8192999.
 176. Warin K, Limprasert W, Suebnukarn S, Jinaporntham S, Jantana P. Automatic classification and detection of oral cancer in photographic images using deep learning algorithms. *J Oral Pathol Med.* 2021;50(9):911–8.
 177. Wilde DC, Castro PD, Bera K, Lai S, Madabhushi A, Corredor G, et al. Oropharyngeal cancer outcomes correlate with p16 status, multinucleation and immune infiltration. *Mod Pathol.* 2022;7:9.
 178. Wong LM, King AD, Ai QYH, Lam WKJ, Poon DMC, Ma BBY, et al. Convolutional neural network for discriminating nasopharyngeal carcinoma and benign hyperplasia on MRI. *Eur Radiol.* 2021;31(6):3856–63.
 179. Wu J, Gensheimer MF, Zhang N, Han F, Liang R, Qian Y, et al. Integrating tumor and nodal imaging characteristics at baseline and mid-treatment computed tomography scans to predict distant metastasis in oropharyngeal cancer treated with concurrent chemoradiotherapy. *Int J Radiat Oncol Biol Phys.* 2019;104(4):942–52.
 180. Wu W, Ye J, Wang Q, Luo J, Xu S. CT-based radiomics signature for the preoperative discrimination between head and neck squamous cell carcinoma grades. *Front Oncol.* 2019;9:17.
 181. Xiong H, Lin P, Yu JG, Ye J, Xiao L, Tao Y, et al. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine.* 2019;48:92–9.
 182. Xu J, Wang J, Bian X, Zhu JQ, Tie CW, Liu X, et al. Deep Learning for nasopharyngeal Carcinoma Identification Using Both White Light and Narrow-Band Imaging Endoscopy. *Laryngoscope.* 2022;132(5):999–1007.
 183. Yan Y, Liu Y, Tao J, Li Z, Qu X, Guo J, et al. Preoperative prediction of malignant transformation of sinonasal inverted papilloma using MR Radiomics. *Front Oncol.* 2022;12: 870544.
 184. Yang SY, Li SH, Liu JL, Sun XQ, Cen YY, Ren RY, et al. Histopathology-Based Diagnosis of Oral Squamous Cell Carcinoma Using Deep Learning. *J Dental Res.* 2022;67:9.
 185. Yoshizawa K, Ando H, Kimura Y, Kawashiri S, Yokomichi H, Moroi A, et al. Automatic discrimination of Yamamoto-Kohama classification by machine learning approach for invasive pattern of oral squamous cell carcinoma using digital microscopic images: a retrospective study. *Oral Surg Oral Med Oral Pathol Oral Radiol.* 2022;133(4):441–52.
 186. Yu H, Ma SJ, Farrugia M, Iovoli AJ, Wooten KE, Gupta V, et al. Machine learning incorporating host factors for predicting survival in head and neck squamous cell carcinoma patients. *Cancers (Basel).* 2021;13:18.
 187. Yuan W, Cheng L, Yang J, Yin B, Fan X, Yang J, et al. Noninvasive oral cancer screening based on local residual adaptation network using optical coherence tomography. *Med Biol Eng Comput.* 2022;60(5):1363–75.
 188. Zhang L, Wu Y, Zheng B, Su L, Chen Y, Ma S, et al. Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy. *Theranostics.* 2019;9(9):2541–54.
 189. Zhong J, Frood R, Brown P, Nelstrop H, Prestwich R, McDermott G, et al. Machine learning-based FDG PET-CT radiomics for outcome prediction in larynx and hypopharynx squamous cell carcinoma. *Clin Radiol.* 2021;76(1):78.e9–e17.
 190. Zhong YW, Jiang Y, Dong S, Wu WJ, Wang LX, Zhang J, et al. Tumor radiomics signature for artificial neural network-assisted detection of neck metastasis in patient with tongue cancer. *J Neuroimaging.* 2022;49(2):213–8.
 191. Zhou X, Hao Y, Peng X, Li B, Han Q, Ren B, et al. The clinical potential of oral microbiota as a screening tool for oral squamous cell carcinomas. *Front Cell Infect Microbiol.* 2021;11: 728933.
 192. Zlotogorski-Hurvitz A, Dekel BZ, Malonek D, Yahalom R, Vered M. FTIR-based spectrum of salivary exosomes coupled with computational-aided discriminating analysis in the diagnosis of oral cancer. *J Cancer Res Clin Oncol.* 2019;145(3):685–94.
 193. Sarker IH. Machine learning: algorithms, real-world applications and research directions. *SN Computer Sci.* 2021;2(3):160.
 194. Koçak B, Durmaz E, Ateş E, Kılıçkesmez Ö. Radiomics with artificial intelligence: a practical guide for beginners. *Diagn Interv Radiol.* 2019;25(6):485–95.
 195. van Dijk LV, Fuller CD. Artificial intelligence and radiomics in head and neck cancer care: opportunities, mechanics, and challenges. *Am Soc Clin Oncol Educ Book.* 2021;41:e225–35.
 196. Casey JA, Schwartz BS, Stewart WF, Adler NE. Using electronic health records for population health research: a review of methods and applications. *Annu Rev Public Health.* 2016;37:61–81.
 197. Alzubaidi L, Zhang J, Humaidi AJ, Al-Dujaili A, Duan Y, Al-Shamma O, et al. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *J Big Data.* 2021;8(1):53.
 198. Adeoye J, Akinshipo A, Thomson P, Su Y-X. Artificial intelligence-based prediction for cancer-related outcomes in Africa: Status and potential refinements. *J Global Health.* 2022;12:7.
 199. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med.* 2019;17(1):195.
 200. Cabitza F, Campagner A. The need to separate the wheat from the chaff in medical informatics: Introducing a comprehensive checklist for the (self)-assessment of medical AI studies. *Int J Med Informatics.* 2021;153: 104510.
 201. Sabottke CF, Spieler BM. The effect of image resolution on deep learning in radiography. *Radiol Artif Intell.* 2020;2(1): e190015.

202. Thambawita V, Strümke I, Hicks SA, Halvorsen P, Parasa S, Riegler MA. Impact of Image Resolution on Deep Learning Performance in Endoscopy Image Classification: An Experimental Study Using a Large Dataset of Endoscopic Images. *Diagnostics (Basel)*. 2021;11:12.
203. Nair AS. Publication bias - Importance of studies with negative results! *Indian J Anaesth*. 2019;63(6):505–7.
204. Murad MH, Chu H, Lin L, Wang Z. The effect of publication bias magnitude and direction on the certainty in evidence. *BMJ Evidence-Based Medicine*. 2018;23(3):84.
205. Motamed S, Rogalla P, Khalvati F. Data augmentation using Generative Adversarial Networks (GANs) for GAN-based detection of Pneumonia and COVID-19 in chest X-ray images. *Inform Med*. 2021;27: 100779.
206. Shorten C, Khoshgoftaar TM. A survey on Image Data Augmentation for Deep Learning. *J Big Data*. 2019;6(1):60.
207. Gupta N, Patel H, Afzal S, Panwar N, Mittal RS, Guttula S, et al. Data Quality Toolkit: Automatic assessment of data quality and remediation for machine learning datasets. *arXiv preprint arXiv:210805935*. 2021.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- ▶ Convenient online submission
- ▶ Rigorous peer review
- ▶ Open access: articles freely available online
- ▶ High visibility within the field
- ▶ Retaining the copyright to your article

Submit your next manuscript at ▶ [springeropen.com](https://www.springeropen.com)
