Machine Learning in Otolaryngology-Head and Neck surgery: A Systematic Review Protocol

Chang Woo Lee (changwoo.lee@nhs.net)
Newcastle Upon Tyne Hospitals NHS Foundation Trust https://orcid.org/0000-0001-6552-337X

Angelos Mantelakis
Guy's and Saint Thomas' NHS Foundation Trust

Bhavesh Vijay Tailor
Norfolk and Norwich University Hospitals NHS Foundation Trust

Ankur Khajuria
Imperial College London Department of Surgery and Cancer

Protocol

Keywords: Machine learning, Otolaryngology, Head and neck surgery, Systematic review

Posted Date: August 19th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-56359/v1

Abstract

Background: Machine learning describes a subfield of artificial intelligence which utilises statistical algorithms to identify patterns in large datasets. Based on previous learning, inferences or predictions can be made given novel data. Alongside its promising potential to revolutionise consumer technology, there has been growing interest in the application of machine learning algorithms to medical practice. The aim of this study is to evaluate the applications of machine learning in Otolaryngology-Head and Neck surgery.

Methods: A systematic search of EMBASE, MEDLINE and CENTRAL will be conducted from January 1990 to June 2020. Studies utilising machine learning as a tool for diagnosis, or to predict disease prognosis or post-operative outcomes in the field of Otolaryngology-Head and Neck surgery will be included. The primary outcome of interest is the accuracy of machine learning models for clinical diagnosis, disease prognostication, and in predicting post-operative outcomes. This protocol adheres to the Preferred Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) guidelines.

Discussion: To our knowledge, this will be the first systematic review to assimilate and critically appraise original research on the applications of machine learning across the field of Otolaryngology-Head and Neck surgery. This review has the potential to inform the current state of this technology and guide future study of machine learning approaches within the specialty.

Systematic review registration: PROSPERO CRD42020192493

Background

There has been a rapid increase in volume, variety, and velocity of clinical data available (1). The availability of this “Big Data” represents a potential for better evidence-based medicine allowing the delivery of more efficient healthcare with improved patient outcome (2). However, the absence of techniques for collecting, storing and analysing such large and complex data sets limited the utilisation of this “Big Data” in the past (2, 3). Recently, there has been a growing interest in the application of machine learning algorithms as a method to utilise this “Big Data” in research and healthcare settings (3, 4).

Machine learning describes a subfield of artificial intelligence which utilises statistical algorithms to identify patterns in large datasets. Based on previous learning, inferences or predictions can be made given novel data (3). Machine learning can be classified into two broad categories based on its approach: supervised and unsupervised learning. Supervised learning involves training of an algorithm with a set of existing data made up of inputs that are associated with a known output (i.e. labelled data). Once learned, through optimisation of its algorithms, predicting outcomes from previously unseen data becomes possible (3, 4). With unsupervised learning, there are no outputs to predict, and inputs are unlabelled. The aim is to infer patterns and the structure in the given data set to generate novel associations (4, 5).

Two areas within the medical field have gained particular attention for the application of machine learning: diagnosis and outcome prediction (4). Recently, there has been a drastic increase in the volume of literature describing the application of machine learning across a range of specialities within medicine and surgery (3–6). Otolaryngology-Head and Neck surgery is no exception. For example, studies have reported its use in diagnosing plethora of diseases of the head and neck through evaluation of patients' presenting complaint (for example, voice analysis) (7, 8), or through evaluation of radiological, histological and/or endoscopic images (9–11). Other studies have reported its use in disease prognostication (for example, predicting hearing outcomes in sudden sensorineural hearing loss) (12), or in terms of predicting post-operative outcomes (for example, predicting post-operative complications in head and neck microvascular free tissue transfer) (13).

Aim

In this review, our aim is to evaluate the clinical application of machine learning in the field of Otolaryngology-Head and Neck surgery. Specifically, this will include the following subspecialties: otology and neurotology/lateral skull base surgery; rhinology and anterior skull base surgery; facial plastics; laryngology; head and neck surgery; paediatric otolaryngology.

Objectives

1. To evaluate the accuracy of machine learning models in diagnosing a clinical condition relevant to the field of Otolaryngology-Head and Neck surgery.
2. To evaluate the accuracy of machine learning models in disease prognosis for a condition relevant to the field of Otolaryngology-Head and Neck surgery.
3. To evaluate the accuracy of machine learning models in predicting the post-operative outcomes of Otolaryngology-Head and Neck surgical procedures.

Methods

Protocol and registration

The study has been registered in the Prospective Register of Ongoing Systematic Reviews (PROSPERO) with registration number CRD42020192493. This protocol is being reported in accordance with the Preferred Reporting Items for Systematic Review guidelines and Meta-Analysis Protocols (PRISMA-P 2015) (Additional File 1).

Eligibility criteria

Primary studies with the following characteristics as outlined in PICO (Participants, Interventions, Comparators and Outcomes) format below will be included.

Participants

Adult participants (> 18 years old) with clinical conditions within the specialty of Otolaryngology-Head and Neck surgery will be included.

Interventions and comparators

We will consider studies which utilise machine learning as a tool for diagnosis, to predict disease prognosis or post-operative outcomes in the field of Otolaryngology-Head and Neck surgery. The intervention may be used in isolation, combined with other statistical models, or in comparison with other techniques.

Outcomes

Three primary outcomes will be looked at to evaluate the clinical application of machine learning in the field of Otolaryngology-Head and Neck surgery. The first will be the accuracy of providing a clinical diagnosis. Studies must specify a defined clinical condition for which the model is designed to identify. The second outcome will be the accuracy of predicting the disease outcome (i.e. prognostication). Studies must specify a defined disease outcome for which the model is designed to predict (for example, the rate of tumour growth), with clinical data collected prospectively or retrospectively to validate their model's prediction. The third outcome will be the accuracy of predicting the post-operative outcomes and complications of Otolaryngology-Head and Neck surgical interventions. Similar to above, studies must specify a defined clinical outcome (for example, probability of disease recurrence), with clinical data collected prospectively or retrospectively to validate the model's prediction.

Exclusion criteria

Animal studies, utilisation of machine learning without clinical data, non-English language articles and review articles will be excluded.

Search Strategies

A comprehensive search of MEDLINE (Ovid), EMBASE (Ovid), and the Cochrane Central Register of Controlled Trials (CENTRAL) will be performed. Studies published between 1990 and the date of the search will be considered for review. A search strategy will be developed in collaboration with an academic librarian using a combination of free text and Medical Subject Headings (MeSH) terms relevant to the subject topic. Table 1 shows an example search strategy in MEDLINE.

Table 1
MEDLINE example search strategy

Search number	Search terms
1	(“Deep learning” OR “Artificial Intelligence” OR “Machine learning” OR “Decision trees” OR “Random forests” OR SVM OR “Support vector machine”)
2	exp “Neural networks (computer)” / OR exp “Deep learning”/
3	exp “Artificial Intelligence”/
4	(1 OR 2 OR 3)
5	(Otolaryngology OR Otorhinolaryngology OR Otology OR Neurotology OR Rhinology OR Laryngology OR Phonosurgery OR (surgery AND (“Head and Neck” OR Otoneurological OR Airway OR “Skull Base”))
6	(4 AND 5)

Identification and selection of studies

Retrieved studies will be exported to Endnote X7 library (Clarivate Analytics, USA). Two independent reviewers will screen the studies based on their titles and abstracts adhering to the set eligibility criteria. Full-text articles of the selected studies will be retrieved, and will subsequently undergo a second stage of screening by the same two independent reviewers adhering to the set eligibility criteria. Any discrepancies in either of the stages will be resolved through a consult by a third reviewer.

Data extraction, collection and management

A standardised Microsoft Excel based data extraction form will be developed and piloted. Data will be extracted by two reviewers independently. Any discrepancies identified will be resolved through discussion with a third reviewer. The data extracted will include study characteristics (authors, year of publication, study design), patient demographics (number of participants, sex, mean age), domain of machine learning application (prediction of a diagnosis or disease outcome or treatment outcome), description of algorithm, and primary outcomes – either reported or derived from published data (specificity, sensitivity, positive predictive value and negative predictive values; post-operative function, treatment success, complications and recurrence).

Risk of bias

The risk of bias in the selected studies will be evaluated by two independent reviewers. For randomised controlled trials, the Cochrane Collaboration Risk of Bias tool will be utilised with the following domains: selection, performance, detection, attrition, reporting, and others (14). For non-randomised trials, the Risk of Bias in Non-randomised Studies of Intervention (ROBINS-I) will be utilised (15). Quality Assessment for Diagnostic Accuracy Studies (QUADAS-2) tool, and Quality in Prognosis Studies (QUIPS) tool will be utilised to assess the risk of bias in the performance of machine learning models (16, 17). An overall

grading of unknown, low, medium or high risk of bias will be assigned. Any discrepancies in quality assessment will be resolved through discussion with a third reviewer.

Data analysis

Clinical heterogeneity (population, machine learning approach, machine learning domain – diagnosis or outcome prediction) will be described, and a narrative review will be performed structured around our objectives. Meta-analysis will only be considered if the included studies are sufficiently homogeneous in terms of study design and outcome measures. I^2 statistic will be utilised to quantify statistical heterogeneity, and a random-effects model will be employed for heterogeneous cohorts (I2 > 50%) (18). The Grading of Recommendations Assessment, Development and Evaluation (GRADE) tool will be used to assess the quality of overall evidence (19). Sensitivity analysis will be considered based on the study quality.

Results

Not applicable as this is a systematic review protocol. Added to acknowledge Editorial Office request.

Discussion

Successful utilisation of medical Big data will undoubtedly aid clinicians in their decision-making process, and will lead to better patient outcomes. To our knowledge, this will be the first systematic review to assimilate and critically appraise original research on the applications of machine learning across the field of Otolaryngology-Head and Neck surgery. This review has the potential to inform the current state of this technology and guide future study of machine learning approaches within the specialty.

Conclusion

Not applicable as this is a systematic review protocol. Added to acknowledge Editorial Office request.

Abbreviations

CENTRAL
Cochrane Central Register of Controlled Trials; EMBASE:Excerpta Medica Database; GRADE:Grading of Recommendations Assessment, Development and Evaluation; PICO:Participants, Interventions, Comparators, Outcomes; PRISMA-P:Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols; QUADAS-2:Quality Assessment for Diagnostic Accuracy Studies; QUIPS:Quality in Prognosis Studies; ROBINS-I:Risk of Bias in Non-randomised Studies of Interventions.

Declarations

Acknowledgements

We would like to thank Mr Stephen Phillips (Royal Surrey NHS Foundation Trust) for his assistance in conducting the literature search.

Authors' contributions

All authors contributed to the conception of the protocol and study design, reviewed this report and approved the final manuscript. CWL wrote the protocol.

Funding

No funding was received for this study.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

All authors declare that they have no competing interests.

References

1. Hulsen T, Jamuar SS, Moody AR, Karnes JH, Varga O, Hedensted S, et al. From big data to precision medicine. Front Med. 2019;6(MAR):1–14.
2. Kanevsky J, Corban J, Gaster R, Kanevsky A, Lin S, Gilardino M. Big Data and Machine Learning in Plastic Surgery: A New Frontier in Surgical Innovation. Plast Reconstr Surg. 2016;137(5):890e–897e.
3. Wiens J, Shenoy ES. Machine Learning for Healthcare: On the Verge of a Major Shift in Healthcare Epidemiology. Clin Infect Dis. 2018;66(1):149–53.
4. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19(1):1–18.
5. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920–30.
6. Mantelakis A, Khajuria A. The applications of machine learning in plastic and reconstructive surgery: Protocol of a systematic review. Syst Rev. 2020;9(1):1–4.
7. Fang SH, Tsao Y, Hsiao MJ, Chen JY, Lai YH, Lin FC, et al. Detection of Pathological Voice Using Cepstrum Vectors: A Deep Learning Approach. J Voice. 2019;33(5):634–41.
8. Fujimura S, Kojima T, Okanoue Y, Shoji K, Inoue M, Hori R. Discrimination of "hot potato voice" caused by upper airway obstruction utilizing a support vector machine. Laryngoscope. 2019;129(6):1301–7.

9. Habib AR, Wong E, Sacks R, Singh N. Artificial intelligence to detect tympanic membrane perforations. J Laryngol Otol. 2020;134(4):311–5.
10. Halicek M, Dormer JD, Little JV, Chen AY, Fei B. Tumor detection of the thyroid and salivary glands using hyperspectral imaging and deep learning. Biomed Opt Express. 2020;11(3):1383.
11. 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):1367.
12. Bing D, Ying J, Miao J, et al. Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models. Clin Otolaryngol. 2018;43(3):868–74.
13. 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 [published online ahead of print, 2020 Jan 28]. Laryngoscope. 2020. 10.1002/lary.28508.doi:10.1002/lary.28508.
14. Higgins JPT, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343(7829):1–9.
15. Sterne JA, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:4–10.
16. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol. 2003;3(1):25.
17. Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280–6.
18. Higgins JP, Thompson S, Deeks J, Altman DG. Measuring inconsistency in meta-analyses. BMJ. 2003;327:557–60.
19. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. BMJ. 2004;328:1490.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- PRISMAPchecklist.docx