



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/174571/>

Version: Published Version

Article:

Mahmood, H., Shaban, M., Rajpoot, N. et al. (2021) Artificial Intelligence-based methods in head and neck cancer diagnosis : an overview. *British Journal of Cancer*, 124 (12). pp. 1934-1940. ISSN: 0007-0920

<https://doi.org/10.1038/s41416-021-01386-x>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



ARTICLE

AI Applied to Cancer

Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview

Hanya Mahmood¹, Muhammad Shaban², Nasir Rajpoot² and Syed A. Khurram³

BACKGROUND: This paper reviews recent literature employing Artificial Intelligence/Machine Learning (AI/ML) methods for diagnostic evaluation of head and neck cancers (HNC) using automated image analysis.

METHODS: Electronic database searches using MEDLINE via OVID, EMBASE and Google Scholar were conducted to retrieve articles using AI/ML for diagnostic evaluation of HNC (2009–2020). No restrictions were placed on the AI/ML method or imaging modality used.

RESULTS: In total, 32 articles were identified. HNC sites included oral cavity ($n = 16$), nasopharynx ($n = 3$), oropharynx ($n = 3$), larynx ($n = 2$), salivary glands ($n = 2$), sinonasal ($n = 1$) and in five studies multiple sites were studied. Imaging modalities included histological ($n = 9$), radiological ($n = 8$), hyperspectral ($n = 6$), endoscopic/clinical ($n = 5$), infrared thermal ($n = 1$) and optical ($n = 1$). Clinicopathologic/genomic data were used in two studies. Traditional ML methods were employed in 22 studies (69%), deep learning (DL) in eight studies (25%) and a combination of these methods in two studies (6%).

CONCLUSIONS: There is an increasing volume of studies exploring the role of AI/ML to aid HNC detection using a range of imaging modalities. These methods can achieve high degrees of accuracy that can exceed the abilities of human judgement in making data predictions. Large-scale multi-centric prospective studies are required to aid deployment into clinical practice.

British Journal of Cancer <https://doi.org/10.1038/s41416-021-01386-x>

BACKGROUND

Head and neck cancers: incidence and diagnosis
Head and neck cancers (HNC) comprise a heterogeneous group of cancers, most commonly squamous cell carcinomas (SCC), that typically arise from the epithelial lining of the oral cavity, sinonasal tract, pharynx, larynx and salivary glands.¹ Most HNC are already at an advanced stage when diagnosed, which significantly reduces the survival rate, even after curative treatment.² Major risk factors include tobacco smoking/chewing,³ excessive alcohol consumption,⁴ areca (betel) nut, paan masala (Gutkha),⁵ gamma and ultraviolet radiation, overexposure to sunlight, a family history of cancer and increasing age.⁶ The role of human papillomavirus (HPV)⁷ and Epstein–Barr virus (EBV) has also been implicated in the development of oropharyngeal and nasopharyngeal SCC.^{3,8} The global incidence of HNC continues to rise^{9,10} with more than half a million cases annually¹¹ and ~12,000 new cases in the UK each year, an increase of 20% in the last decade.¹² Prognosis remains poor, with a 28–67% chance of survival at five years, depending upon the stage at presentation.¹²

Public health screening/awareness programmes, withdrawal of environmental carcinogens and early detection of precancerous lesions remain the focus for primary and secondary prevention.³ However, early detection of some HNC can be difficult due to vague histories and indistinctive diagnostic features. Conventional diagnosis of HNC is based on histopathological evaluation of tissue sections from biopsies or surgical resections, in addition to clinical and

radiological examinations. These methods can be time-consuming and are prone to errors in observation or variations in interpretation,^{13–15} which can result in inconsistencies in cancer grading and prognostication.¹⁶ Consequently, this can cause delays and/or inaccuracies in diagnosis, which can have significant implications on patient management and survival. Indeed, improvements in HNC prediction accuracy and disease outcomes could greatly assist healthcare professionals in the early detection and planning of patient-specific optimal treatments to reduce the disease burden.

Artificial intelligence: machine learning and medical image analysis

Recent technological advancements in Artificial Intelligence (AI) algorithms, computer hardware and big medical imaging datasets have enabled computer scientists and healthcare researchers to collaborate closely to improve consistency in cancer risk stratification over the use of multi-factor analysis, conventional logistic regression and Cox analyses.¹⁷

The recent advent of Machine Learning (ML) has seen a surge of interest with the exponential growth of evidence to support its wide applications in a range of cancers.^{18–24} ML is a branch of AI that uses computational methods to detect patterns, gather insight and make predictions about new data by using historical information that has been ‘learnt’. As the volume of training data increases, ML algorithms can produce more accurate and efficient predictions.²⁵ Deep learning (DL) is a subfield of ML in which

¹Academic Unit of Oral & Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield, Sheffield, UK; ²Department of Computer Science, University of Warwick, Coventry, UK and ³Unit of Oral & Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, Sheffield, UK
Correspondence: Hanya Mahmood (h.mahmood@sheffield.ac.uk)

Received: 28 October 2020 Revised: 11 March 2021 Accepted: 31 March 2021
Published online: 19 April 2021

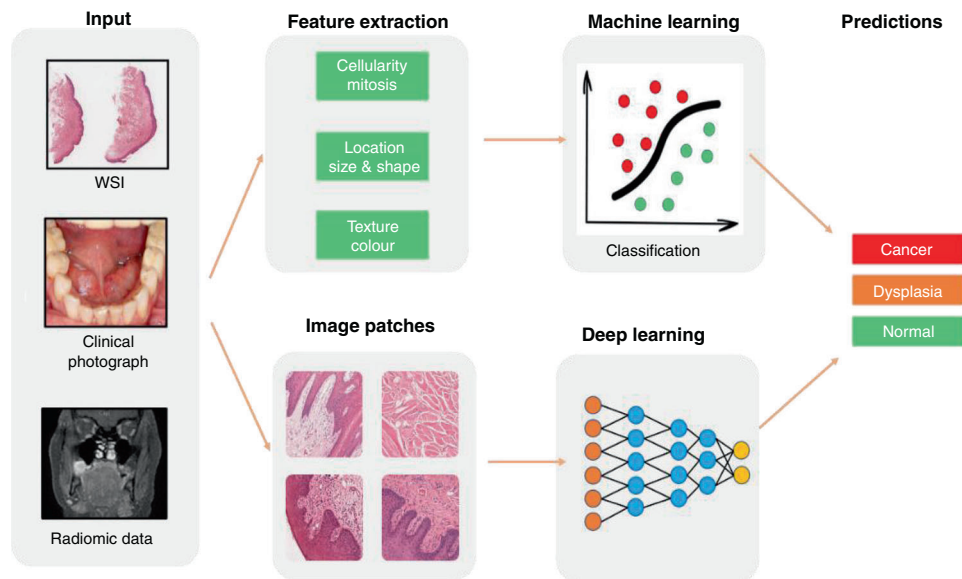


Fig. 1 Feature extraction from primary diagnostic imaging modalities to train ML/DL algorithms to aid outcome prediction. Source: The schematic diagrams were prepared in line with the journal's artwork guidelines; the clinical/histological/radiological images were obtained with appropriate consent from Sheffield Teaching Hospitals NHS Foundation Trust.

algorithms are structured to create artificial neural networks with multiple hidden layers. These methods have also gained significant popularity in recent years due to their achieving relatively high accuracy of prediction.

Much of the recent focus in cancer diagnostics has centred on digital image analysis and processing, which involves extraction of meaningful information from images to enable delineation of features of clinical interest (segmentation) or description of labels (classification).^{26,27} A number of ad hoc (or hand-crafted) feature analysis-based ML approaches have been shown to be successful in different diagnostic applications, by explicitly defining a prior set of features and processing steps²⁸ (Fig. 1). Detection of HNC can be achieved using these ML methods by obtaining clinically important information from primary diagnostic imaging modalities, in which high-dimensional, mineable images can be input to train algorithms. For example, radiomic data can be derived from radiographs, ultrasound (US), computed tomography (CT), magnetic resonance (MR), positron emission tomography (PET) and nuclear medicine imaging methods, such as single-photon emission computed tomography (SPECT). Similarly, histological, cytological and immunohistochemical data can be obtained from high-resolution whole-slide images (WSI) of stained tissue sections from biopsies or surgical resections. Other emerging tools for non-invasive detection of HNC include multispectral narrow-band imaging, Raman spectroscopy, confocal laser endomicroscopy (CLE) and infrared thermal imaging.

This paper seeks to provide an overview of the recently published literature relating to the application of AI/ML methods to aid diagnostic evaluation of HNC. An outline of the different anatomical sites for HNC lesions, the type of diagnostic imaging modality and the AI/ML method used will be presented.

METHODS

Literature search

Electronic database searches using MEDLINE via OVID, EMBASE and Google Scholar were conducted to retrieve articles published in the English language over the last eleven years (2009–2020). This period was chosen due to the evolving application of AI/ML methods in diagnostic cancer research over the last decade.

The search strategy was developed in collaboration with a medical information specialist (Health Sciences Library, University of Sheffield, UK) to ensure keywords were appropriately chosen for optimal identification of articles. A combination of tailored search strings containing database-specific medical subject headings (MeSH) and controlled vocabulary was used (see Supplementary Information), and grey literature screened. Whilst not intended as a formal systematic review, the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and checklist were followed where possible.

Study selection

The selection criterion was jointly developed by the authorship team. The principal inclusion criteria were studies applying AI/ML methods to aid diagnosis of HNC using image analysis, with no restrictions placed on the types of methods or imaging modalities used. Due to the anticipated small number of studies in the field, a broad range of studies such as those using AI/ML to identify or differentiate between benign/pre-malignant/malignant pathology, classify disease subtype, segment cancer regions or predict disease outcome.

Studies using AI/ML to predict cancer susceptibility, metastasis, recurrence, survival or treatment efficacy were not included in this review. Studies focussing solely on the evaluation of oesophageal or thyroid cancers were excluded, unless they were included as part of a larger study that included other HNC lesions. Narrative reviews, letters to editors, commentaries, conference abstracts and animal studies were also excluded. All articles were independently screened by two authors (H.M. and M.S.). The first screen involved the assessment of study title and abstracts and the removal of duplicate articles. The second screen involved comprehensive full-text examination against the predefined criterion. In the case of author discrepancy, two further authors (NMR, SAK) were consulted to make a final decision on article inclusion.

Data capture and synthesis

Relevant data were extracted, tabulated and processed in Microsoft Excel® (Microsoft Corporation, Washington, USA). Data collection included

- Study details (date of publication, authors, study location and aims)
- Study methods (anatomical sites for HNC lesions, diagnostic imaging modality, dataset sizes and application of AI/ML methods to aid cancer diagnosis/outcome)
- AI/ML algorithm performance (reported accuracy measures)

A narrative synthesis with the relevant graphical display is presented. Due to the variations in study outcomes and heterogeneity of data, a meta-analysis for the calculation of adjusted pool estimates was not performed.

RESULTS

The electronic database search retrieved 771 scientific articles. After the removal of duplicates and screening of study titles and abstracts, 698 articles were excluded. Detailed full-text examination of remaining articles excluded a further 41 studies, resulting in 32 articles for inclusion (see Supplementary Table). Among the selected articles, 9 were published between 2009 and 2014, and the remaining 23 articles published between 2015 and 2020. The primary outcomes of interest were anatomical sites of the HNC lesions, diagnostic imaging modalities used for algorithm training/optimisation and the type of AI/ML method.

Anatomical sites for HNC lesions

Figure 2 illustrates the different anatomical sites for the HNC lesions in the selected studies, with the largest proportion (16 studies) involving the oral cavity. Amongst these, nine studies focussed on the assessment of oral squamous cell carcinoma (OSCC) and seven studies focussed on the evaluation of, or differentiation between, oral potentially malignant disorders (OPMD) and OSCC. The remaining studies focussed on assessment of nasopharyngeal SCC ($n=3$),^{29–31} laryngeal SCC ($n=2$),^{32,33} oropharyngeal SCC ($n=3$),^{34–36} parotid gland neoplasms ($n=2$)^{37,38} and differentiation between sinonasal SCC from inverted papilloma ($n=1$).³⁹ In four studies,^{40–42} tissue sections of HNC from various different sites (tongue, floor of mouth, soft palate, mandible, gingivae, alveolar ridge, supraglottis, maxillary sinus, nose, thyroid and parotid gland) were evaluated. One of these studies did not specify the anatomical sites of the HNC lesions.⁴³

Diagnostic imaging modalities for algorithm training/optimisation
Histology WSI of haematoxylin and eosin (H&E)-stained tissue sections was used to train AI/ML algorithms in nine studies (Fig. 3). Radiology image data were used in eight studies and obtained from dynamic contrast-enhanced MRI (DCE-MRI) ($n=3$),^{31,39,43} CT ($n=2$),^{36,37} PET/CT ($n=1$),²⁹ US ($n=1$)³⁸ and plain film intraoral radiographs ($n=1$).⁴⁴ Other imaging modalities included hyperspectral imaging (HSI) ($n=6$), endoscopic/clinical imaging ($n=5$), infrared thermal imaging ($n=1$)⁴⁵ and multimodal optical imaging ($n=1$).⁴⁶ In the remaining two studies, clinicopathologic, genomic and exfoliative cytological data were used to predict outcomes based on traditional statistical analysis methods.^{47,48}

Histology whole-slide imaging

In nine studies, histology WSI was used to develop algorithms for evaluation of OSCC ($n=2$),^{49,50} OPMD ($n=4$),^{51–54} laryngeal SCC ($n=1$),³² oropharyngeal SCC ($n=1$)³⁵ and multiple HNC sites ($n=1$).⁵⁵

These studies used a variety of different ML approaches to delineate specific histological features of interest with downstream statistical analysis to compare differences in spatial architectural patterns for differentiation between benign and malignant lesions. ML tools were developed to assess differences in detection of quantity, geometry, compactness and eccentricity of sub-epithelial connective tissue cells and to identify textural differences between normal and oral submucous fibrosis tissue

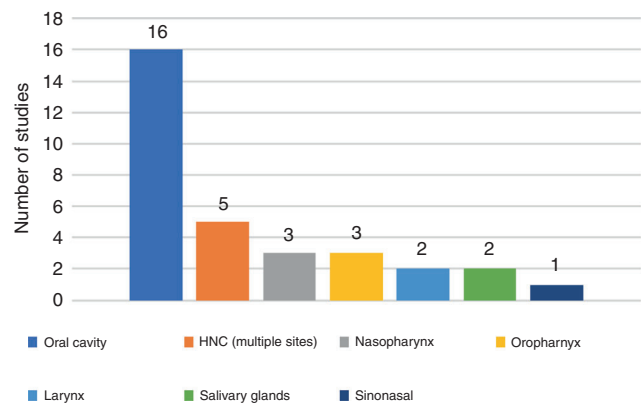


Fig. 2 HNC anatomical site distribution. A bar chart showing the proportion of included studies based on head and neck cancer anatomical subtype.

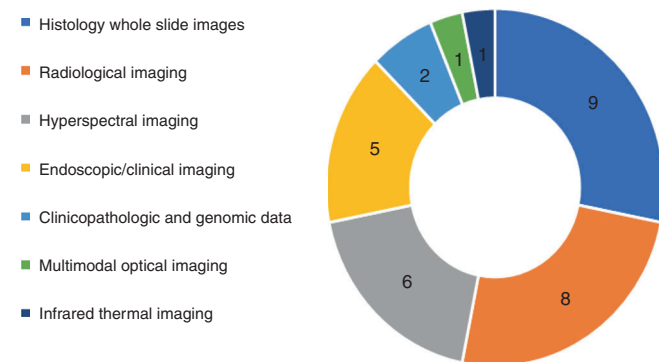


Fig. 3 Type of imaging modality/input data used. A pie chart showing the proportion of identified studies based on the type of imaging modality/input data used to train AI/ML algorithms.

(with and without dysplasia or atrophy) using approaches including Brownian motion.^{51–53}

In one study, unsupervised ML methods were used for the automated identification of tissue compartments in oropharyngeal SCC (OPSCC) tissue microarrays.³⁵ Morphological measurements of cell and nuclei were used for the classification of epithelial and stromal tissue achieving a pixel-level F1 score of 80–81%. A further study showed that stimulated Raman scattering histology integrated with DL algorithms provided the good potential for delivering a rapid intraoperative diagnosis of laryngeal SCC with an accuracy of 90%.³² Findings showed that this method could identify tissue neoplasia at the simulated resection margins that appear grossly normal with the naked eye, highlighting the potential to enhance surgical resection and reduce disease recurrence.

A recent systematic review highlights emerging evidence to support the role of ML methods for histology images as a potentially useful diagnostic aid for the detection of OSCC and some OPMD, but identifies a lack of evidence for other head and neck precancerous or cancerous lesions.⁴⁴ However, the overall quality of evidence in these studies is low, mainly due to the use of small unicentric datasets and a high risk of bias that could have overestimated model accuracy rates.

Radiological imaging

Three studies used radiomic-based feature prediction from MRI to aid assessment of various HNC lesions,⁴³ including nasopharyngeal³¹ and sinonasal SCC.³⁹ Deng et al.⁴³ proposed an automatic segmentation method using traditional ML techniques for

evaluation of HNC lesions demonstrating superior segmentation performance (area overlap measure of 0.76 ± 0.08 and accuracy of $86 \pm 8\%$), which outperformed previous similar studies. Huang et al.³¹ evaluated the performance of two region-based segmentation methods for the evaluation of nasopharyngeal SCC. Ramkumar et al.³⁹ found that MRI-based textural analysis had the potential to differentiate sinonasal SCC from inverted papilloma (accuracy 89.1%) with results comparable to manual assessment by neuroradiologists ($P = 0.0004$).

Three studies used CT-based textural analysis for the evaluation of HNC. Ajmi et al.³⁷ developed an approach using spectral dual-energy CT (DECT) data from multi-energy virtual monochromatic image datasets to capture the energy-dependent changes in tissue attenuation for the classification of common benign parotid gland neoplasms (Warthin tumour and pleomorphic adenoma) with an accuracy of 92%. Whereas Ranjbar et al.³⁶ used CT-based texture analysis to classify the HPV status of oropharyngeal SCC (accuracy 75.7%). In another study, Wu et al.²⁹ developed an automated algorithm for the identification of nasopharyngeal SCC on PET/CT examination with 100% accuracy for detection of hypermetabolic lesions larger than 1 cm in size.

Only one study used textural features derived from an ultrasound (US) using radio-frequency echo signals and image data to enable automatic differentiation between malignant and benign parotid gland lesions (accuracy 91%) based on a supervised classification system.³⁸ In another study, gravitational search-optimised echo-state neural networks were developed for early prediction of OSCC from intraoral X-rays with a detection accuracy of 99.2%.⁵⁶

Endoscopic/clinical imaging

Four studies used clinical data from endoscopic imaging for the detection of oral,⁵⁷ nasopharyngeal,³⁰ oropharyngeal³⁴ and laryngeal cancers.³³ Amongst these, two studies employed DL methods. The first study developed algorithms for early detection of nasopharyngeal malignancies (accuracy of 88.7%)³⁰ providing surgeons with useful biopsy guidance. The second study demonstrated the superior performance of DL compared to textural feature-based ML methods (accuracy 88.3%, sensitivity 86.6% and specificity 90%) in recognition of sub-surface micro-anatomical in vivo cell structures using confocal laser endomicroscopy (CLE) in OSCC patients.⁵⁶

In two studies, traditional ML approaches were used for the detection of oropharyngeal and laryngeal SCC. Mascharak et al.³⁴ used multispectral narrow-band imaging and white-light endoscopy (WLE) to quantify the lymphoepithelial tissues of the oropharynx. The results showed a promising ability to differentiate between oropharyngeal SCC and healthy mucosa based on differences in colour (accuracy 65.9% compared to 52.3% under WLE, $P = 0.0108$), presumably a reflection of surface angiogenic and local inflammatory changes. Whereas Moccia et al.³³ used traditional ML techniques to classify laryngeal tissues (normal vs malignant) in narrow-band endoscopic images by exploiting textural information (classification recall 93–98%). These studies demonstrate a promising step towards an endoscope-based processing system to support the early-stage diagnosis of HNC lesions that may go unnoticed by the human eye.

Song et al.⁵⁸ developed a smartphone-based intraoral dual-modality imaging platform to classify OPMD and malignant lesions based on autofluorescence and white-light images using a convolutional neural network (CNN). The results demonstrated an accuracy of ~86.9%, although the training sample was relatively small (66 normal samples and 64 suspicious). Other limitations, including training the CNN algorithms on tissue from different anatomical regions (for normal, dysplastic and malignant tissues), are likely to exhibit different autofluorescence characteristics due to the varying tissue structural and biochemical compositions.

Other imaging modalities

Six studies used HSI for AI/ML training. Three of those studies focussed on early detection and diagnosis of OSCC. Liu et al.⁵⁹ used HSI to measure the reflectance spectra in the tongue to enable differentiation between normal and cancerous tissue, with a recognition rate of 96.5%. Roblyer et al.⁴⁶ used multispectral wide-field optical imaging—which included white-light reflectance, autofluorescence, narrow-band reflectance and cross-polarised imaging modalities—to distinguish between oral cancer/precancer and non-neoplastic mucosa by evaluating image contrast. Their results showed that autofluorescence imaging at 405-nm excitation wavelength provided the greatest image contrast, and the ratio of red-to-green fluorescence intensity computed from these images provided the best classification of dysplasia/cancer versus non-neoplastic tissue (sensitivity of 100%, specificity of 85%). Although this approach accurately distinguished malignant from benign tissue, the ability to separate precancerous lesions from cancers was found to be limited. In the third study, Quang et al.⁶⁰ also used multimodal optical imaging, in which autofluorescence imaging was used to identify high-risk regions within the oral cavity, followed by high-resolution microendoscopy to confirm or exclude the presence of neoplasia (defined by the authors as diagnoses of moderate dysplasia or worse). Data from 92 sites ($n = 30$) were used to develop algorithms for the automatic identification of OSCC in vivo. Diagnostic performance was evaluated prospectively using images from 114 sites ($n = 70$) and the confirmed histological diagnosis based on either a biopsy or an excised surgical specimen. Amongst the sites that were biopsied ($n = 56$), the classification accuracy for detection of benign and cancerous lesions was 100 and 85%, respectively. Amongst the sites that corresponded to a surgical specimen ($n = 58$), multimodal imaging correctly classified 100% of benign and 61% of neoplastic sites.

Jeyaraj et al.⁴⁵ developed a convolutional neural network (CNN) classifier for OSCC detection using multidimensional HSI (accuracy of 91.4%). Similarly, Halicek et al.⁴¹ also developed a CNN classifier that was trained using HSI of HNSCC surgical specimens, including thyroid cancer, and normal head and neck tissue samples. This model showed the potential to produce near-real-time automatic tissue labelling for intraoperative cancer detection using HSI.

One study explored the viability of digital infrared thermal imaging (DITI) for screening and detection of OSCC.⁶¹ DITI is a non-invasive, non-ionising, radiation hazard-free modality that essentially produces a heat map of an object by capturing its radiated thermal energy. The authors developed a semi-automated screening framework for OSCC by extracting significantly discriminating textural features from facial thermograms for classification into normal, precancerous or malignant categories achieving an accuracy rate of 85.42%.

Another study developed diagnostic algorithms for HNSCC detection using ML constructed by mass spectra obtained from non-cancerous ($n = 15$, 114 mass spectra) and HNSCC ($n = 19$, 141 mass spectra) specimens by probe electrospray ionisation mass spectrometry. The clinical validity of this approach was evaluated to discriminate tumour-specific spectral patterns using intraoperative specimens of HNSCC and normal mucosa with positive and negative-ion modes showing accuracies in HNSCC diagnosis of 90.48% and 95.35%, respectively.

In another study,⁴⁸ exfoliative cytology, histopathology and clinical data of normal subjects ($n = 102$), oral leukoplakia (OLK) patients ($n = 82$) and OSCC patients ($n = 93$) were collected for quantitative risk stratification of OLK lesions. This involved expert-guided data transformation and reconstruction for automatic data processing to reveal a risk index for OSCC prediction (sensitivity: median >0.98 , specificity: median >0.99).

Type of AI/ML method

Figure 4 illustrates the proportion of AI methods used in the selected studies. Traditional ML methods were employed in 22 studies (69%) with common approaches, including Support Vector Machine, Logistic Regression, Random Forest, Decision Tree, *K*-Nearest Neighbour, Bayesian Classifier and Linear Discriminant Analysis. DL-based neural networks were employed in eight studies (25%), and a combination of traditional ML and DL methods were used in two studies (6%).

Figure 5 provides a breakdown of the selected studies by combining the anatomical site of lesion, imaging modality and AI

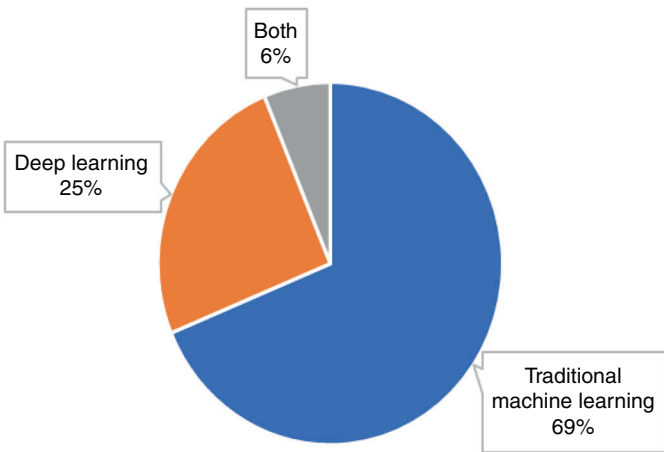


Fig. 4 Types of AI methods used. Proportion of studies using traditional ML methods, modern deep learning methods and a combination of both types of methods.

methods used in the recently published literature. Traditional ML methods were most frequently used for the detection of precancerous or cancerous lesions of the oral cavity (11 studies), and specifically in studies using histology WSI (four studies). DL methods were used for detection of HNSCC lesions of the oral cavity (five studies), nasopharynx (one study), larynx (one study) and various other head and neck sites as specified in a study by Halicek et al.⁴¹

DISCUSSION

This paper provides an insight into the recent application of AI/ML for the evaluation of HNC lesions using digital image analysis. It has shown, primarily, a wide breadth of imaging modalities that have been used to retrieve input data for algorithm training in the last decade. Whilst a detailed statistical analysis of the heterogeneous dataset samples has not been undertaken, most studies have indicated that ML can achieve high degrees of accuracy and precision that can exceed the abilities of standard statistical techniques and human judgement in making predictions about data. This supports seminal claims made by Meehl in 1954,⁶² and more recent meta-analyses,^{63,64} which propose that correctly used mechanical (i.e., algorithmic) methods make more efficient and reliable decisions and predictions about patient outcomes and treatment. However, despite findings highlighted by our paper, very few ML tools have actually been deployed into clinical practice.

Whilst a formal risk of bias analysis has not been conducted for the cited studies, the reported accuracy rates should be interpreted with caution. This is because most studies have used small unicentric datasets that may be biased towards a particular patient demographics. Multi-centric research will inevitably allow a more diverse dataset with the inclusion of patients from different geographical locations, populations and demographics that will enhance algorithm performance by incorporating biological and technical variance.

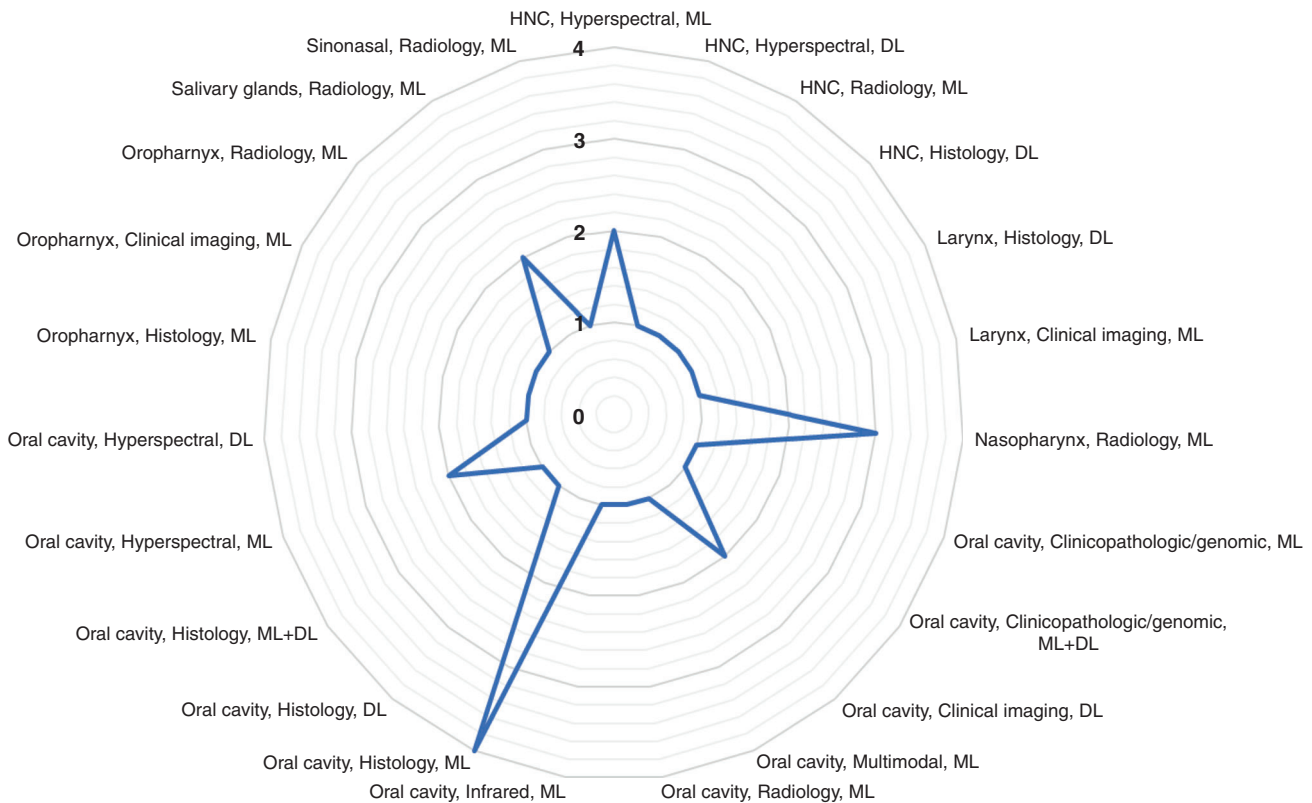


Fig. 5 Overview of recently published literature. A diagram illustrating an overview of the recently published literature based on anatomical site of lesion, imaging modality and AI methods used.

Findings demonstrate the greatest proportion of studies to have evaluated the detection of OPMD and cancerous lesions within the oral cavity (Fig. 2) with histology WSI and radiological imaging being the most frequently used modalities for algorithm training (Fig. 3). This is consistent with the increasing ubiquity of digital slide scanners in pathology laboratories and the emergence of radiomics that has broadened the scope of routine medical imaging in clinical oncology.

With the continued evolution of AI algorithms and computational power, a plethora of computational methods has been developed for fast and reproducible diagnosis and prognosis of HNC, as exemplified in this paper. The emergence of various high-resolution imaging modalities (i.e., multimodal optical, microendoscopic, hyperspectral and infrared thermal) has provided an unprecedented opportunity for quantitative feature extraction by conversion into mineable images at relatively low cost and non-invasively. Having said this, histology WSI remains the most superior imaging modality for data leverage. This is because each image provides multi-gigapixel-level information, thereby resulting in hundreds of thousands of sub-images (image patches) per WSI for analysis and algorithm training.

Early work has been largely based on the development and application of traditional ML methods (Fig. 4); however, in recent years, the use of DL for HNSCC diagnosis and prognostication has evolved. This opens the opportunity to develop state-of-the-art DL techniques that can be combined with traditional approaches to improve detection accuracy of head and neck precancerous and cancerous lesions, as well as predict the course of a precancerous or cancerous lesion learning from retrospective data. Another exciting research avenue would be the development of new data fusion algorithms that combine imaging modalities such as radiologic, histologic and molecular measurements to aid disease detection, classification and outcome prediction.

ACKNOWLEDGEMENTS

We thank the Health Sciences Library at The University of Sheffield for assistance with the electronic database search.

AUTHOR CONTRIBUTIONS

H.M. and S.A.K. conceived and designed the study. H.M. conducted the literature search with support from a medical information specialist (University of Sheffield). H.M. and M.S. screened all articles and collected data that led to the submission. S.A.K. and N.M.R. confirmed the final selection of articles. H.M., N.M.R., S.A.K. and M.S. were involved in analysing results, preparation, editing and approval of the final paper.

ADDITIONAL INFORMATION

Ethics approval and consent to participate Not applicable.

Consent to publish Not applicable.

Data availability All data generated or analysed during this study are included in this published article and its supplementary information files.

Competing interests The authors declare no competing interests.

Funding information H.M. is funded by the National Institute for Health Research and received a small grant from Sheffield Hospitals Charity to support her work. N.M.R. was supported by the PathLAKE digital pathology consortium, which is funded from the Data to Early Diagnosis and Precision Medicine strand of the UK government's Industrial Strategy Challenge Fund (award# 18181), managed and delivered by UK Research and Innovation (UKRI). N.M.R. was also supported by the UK Medical Research Council (award# MR/P015476/1) and the Alan Turing Institute.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41416-021-01386-x>.

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

REFERENCES

1. Pai, S. I. & Westra, W. H. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. *Annu. Rev. Pathol.: Mechanisms Dis.* **4**, 49–70 (2009).
2. Muto, M., Nakane, M., Katada, C., Sano, Y., Ohtsu, A., Esumi, H. et al. Squamous cell carcinoma in situ at oropharyngeal and hypopharyngeal mucosal sites. *Cancer: Interdisciplinary Int. J. Am. Cancer Soc.* **101**, 1375–1381 (2004).
3. Shaw, R. & Beasley, N. Aetiology and risk factors for head and neck cancer: United Kingdom National Multidisciplinary Guidelines. *J. Laryngol. Otol.* **130**, S9–S12 (2016).
4. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Alcohol consumption and ethyl carbamate. *IARC Monogr. Evaluation Carcinogenic Risks Hum.* **96**, 3 (2010).
5. World Health Organization, IARC Working Group on the Evaluation of Carcinogenic Risks to Humans, International Agency for Research on Cancer. *Betel-quid and Areca-nut Chewing and Some Areca-nut-derived Nitrosamines* (IARC, 2004).
6. State of Mouth Cancer UK Report 2020/21. Oral Health Foundation. Available via URL: <https://www.dentalhealth.org/thestateofmouthcancer> (2020).
7. Papillomaviruses, H. *IARC Monographs on the Evaluation of Carcinogenic Risks to Humans*. (IARC, 2011).
8. Shotelersuk, K., Khorprasert, C., Sakdikul, S., Pornthanakasem, W., Voravud, N. & Mutirangura, A. Epstein-Barr virus DNA in serum/plasma as a tumor marker for nasopharyngeal cancer. *Clin. Cancer Res.* **6**, 1046–1051 (2000).
9. Jemal, A., Bray, F., Center, M. M., Ferlay, J., Ward, E. & Forman, D. Global cancer statistics. *CA: Cancer J. Clin.* **61**, 69–90 (2011).
10. Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **70**, 7 (2020).
11. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A. & Jemal, A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: Cancer J. Clin.* **68**, 394–424 (2018).
12. Cancer Research UK. Head and neck cancers incidence statistics 2015–17. Available at URL: <https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers/incidence> (2017).
13. Kujan, O., Khattab, A., Oliver, R. J., Roberts, S. A., Thakker, N. & Sloan, P. Why oral histopathology suffers inter-observer variability on grading oral epithelial dysplasia: an attempt to understand the sources of variation. *Oral. Oncol.* **43**, 224–231 (2007).
14. Vogel, D. W. & Thoeny, H. C. Cross-sectional imaging in cancers of the head and neck: how we review and report. *Cancer Imaging* **16**, 20 (2016).
15. Schlumpf, M. F. & Haerle, S. The current role of imaging in head and neck cancer: a clinician's perspective. *Swiss Medical Weekly* **144**, w14015 (2014).
16. Mehlu, C. S., Larsen, S. R., Kiss, K., Groentved, A. M., Kjaergaard, T., Möller, S. et al. Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. *Laryngoscope* **128**, 2375–2379 (2018).
17. Huang, S., Yang, J., Fong, S. & Zhao, Q. Artificial intelligence in cancer diagnosis and prognosis: opportunities and challenges. *Cancer Lett.* **471**, 61–71 (2020).
18. Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G. et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *J. Am. Med. Assoc.* **318**, 2199–2210 (2017).
19. Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V. & Fotiadis, D. I. Machine learning applications in cancer prognosis and prediction. *Comput. Struct. Biotechnol. J.* **13**, 8–17 (2015).
20. Graham, S. & Rajpoot, N. SAMS-NET: stain-aware multi-scale network for instance-based nuclei segmentation in histology images. In *2018 Institute of Electrical and Electronics Engineers (IEEE) 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 590–594 (IEEE, 2018).
21. Bera, K., Schalper, K. A., Rimm, D. L., Velcheti, V. & Madabhushi, A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. *Nat. Rev. Clin. Oncol.* **16**, 703–715 (2019).
22. Zormpas-Petridis, K., Failmezger, H., Raza, S. E., Roxanis, I., Jamin, Y. & Yuan, Y. Superpixel-based conditional random fields (SuperCRF): incorporating global and local context for enhanced deep learning in melanoma histopathology. *Front. Oncol.* **9**, 1045 (2019).
23. Wang, S., Yang, D. M., Rong, R., Zhan, X., Fujimoto, J., Liu, H. et al. Artificial intelligence in lung cancer pathology image analysis. *Cancers* **11**, 1673 (2019).
24. Sirinukunwattana, K., Raza, S. E., Tsang, Y. W., Snead, D. R., Cree, I. A. & Rajpoot, N. M. Locality sensitive deep learning for detection and classification of nuclei in

- routine colon cancer histology images. *IEEE Trans. Med. Imaging* **35**, 1196–1206 (2016).
25. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521**, 436–444 (2015).
26. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A., Ciampi, F., Ghafoorian, M. et al. A survey on deep learning in medical image analysis. *Med. image Anal.* **42**, 60–88 (2017).
27. Shen, D., Wu, G. & Suk, H. I. Deep learning in medical image analysis. *Annu. Rev. Biomed. Eng.* **19**, 221–248 (2017).
28. Panayides, A. S., Amini, A., Filipovic, N. D., Sharma, A., Tsaftaris, S. A., Young, A. et al. AI and medical imaging informatics: current challenges and future directions. *IEEE J. Biomed. Health Inform.* **24**, 1837–1857 (2020).
29. Wu, B., Khong, P. L. & Chan, T. Automatic detection and classification of nasopharyngeal carcinoma on PET/CT with support vector machine. *Int. J. Computer Assist. Radiol. Surg.* **7**, 635–646 (2012).
30. Li, C., Jing, B., Ke, L., Li, B., Xia, W., He, C. et al. Development and validation of an endoscopic images-based deep learning model for detection with nasopharyngeal malignancies. *Cancer Commun.* **38**, 1–1 (2018).
31. Huang, W., Chan, K. L. & Zhou, J. Region-based nasopharyngeal carcinoma lesion segmentation from MRI using clustering-and classification-based methods with learning. *J. Digital Imaging* **26**, 472–482 (2013).
32. 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* **9**, 2541 (2019).
33. Moccia, S., De Momi, E., Guarnaschelli, M., Savazzi, M., Laborai, A., Guastini, L. et al. Confident texture-based laryngeal tissue classification for early stage diagnosis support. *J. Med. Imaging* **4**, 034502 (2017).
34. Mascharak, S., Baird, B. J. & Holsinger, F. C. Detecting oropharyngeal carcinoma using multispectral, narrow-band imaging and machine learning. *Laryngoscope* **128**, 2514–2520 (2018).
35. Fouad, S., Randell, D., Galton, A., Mehanna, H. & Landini, G. Unsupervised morphological segmentation of tissue compartments in histopathological images. *PLoS ONE* **12**, e0188717 (2017).
36. Ranjbar, S., Ning, S., Zwart, C. M., Wood, C. P., Weindling, S. M., Wu, T. et al. Computed tomography-based texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinoma. *J. computer Assist. Tomogr.* **42**, 299–305 (2018).
37. Al Ajmi, E., Forghani, B., Reinhold, C., Bayat, M. & Forghani, R. Spectral multi-energy CT texture analysis with machine learning for tissue classification: an investigation using classification of benign parotid tumours as a testing paradigm. *Eur. Radiol.* **28**, 2604–2611 (2018).
38. Siebers, S., Zenk, J., Bozzato, A., Klintworth, N., Iro, H. & Ermert, H. Computer aided diagnosis of parotid gland lesions using ultrasonic multi-feature tissue characterization. *Ultrasound Med. Biol.* **36**, 1525–1534 (2010).
39. Ramkumar, S., Ranjbar, S., Ning, S., Lal, D., Zwart, C. M., Wood, C. P. et al. MRI-based texture analysis to differentiate sinonasal squamous cell carcinoma from inverted papilloma. *Am. J. Neuroradiol.* **38**, 1019–1025 (2017).
40. Fei, B., Lu, G., Wang, X., Zhang, H., Little, J. V., Patel, M. R. et al. Label-free reflectance hyperspectral imaging for tumor margin assessment: a pilot study on surgical specimens of cancer patients. *J. Biomed. Opt.* **22**, 086009 (2017).
41. Halicek, M., Lu, G., Little, J. V., Wang, X., Patel, M., Griffith, C. C. et al. Deep convolutional neural networks for classifying head and neck cancer using hyperspectral imaging. *J. Biomed. Opt.* **22**, 060503 (2017).
42. Lu, G., Little, J. V., Wang, X., Zhang, H., Patel, M. R., Griffith, C. C. et al. Detection of head and neck cancer in surgical specimens using quantitative hyperspectral imaging. *Clin. Cancer Res.* **23**, 5426–5436 (2017).
43. 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 & Molecular Imaging* **2017**, 8612519 (2017).
44. Mahmood, H., Shaban, M., Indave, B. I., Santos-Silva, A. R., Rajpoot, N. & Khurram, S. A. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: a systematic review. *Oral. Oncol.* **110**, 104885 (2020).
45. Jeyaraj, P. R. & Nadar, E. R. Computer-assisted medical image classification for early diagnosis of oral cancer employing deep learning algorithm. *J. Cancer Res. Clin. Oncol.* **145**, 829–837 (2019).
46. Roblyer, D. M., Kurachi, C., Stepanek, V., Schwarz, R. A., Williams, M. D., El-Naggar, A. K. et al. Comparison of multispectral wide-field optical imaging modalities to maximize image contrast for objective discrimination of oral neoplasia. *J. Biomed. Opt.* **15**, 066017 (2010).
47. Ashizawa, K., Yoshimura, K., Johno, 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.* **75**, 111–119 (2017).
48. Liu, Y., Li, J., Liu, X., Liu, X., Khawar, W., Zhang, X. et al. Quantitative risk stratification of oral leukoplakia with exfoliative cytology. *PLoS ONE* **10**, e0126760 (2015).
49. Das, D. K., Bose, S., Maiti, A. K., Mitra, B., Mukherjee, G. & Dutta, P. K. Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. *Tissue Cell* **53**, 111–119 (2018).
50. Lu, C., Lewis, J. S., Dupont, W. D., Plummer, W. D., 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.* **30**, 1655–1665 (2017).
51. Baik, J., Ye, Q., Zhang, L., Poh, C., Rosin, M., MacAulay, C. et al. Automated classification of oral premalignant lesions using image cytometry and random forests-based algorithms. *Cell. Oncol.* **37**, 193–202 (2014).
52. Krishnan, M. M., Pal, M., Bomminayuni, S. K., Chakraborty, C., Paul, R. R., Chatterjee, J. et al. Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis—an SVM based approach. *Computers Biol. Med.* **39**, 1096–1104 (2009).
53. Mookiah, M. R., Shah, P., Chakraborty, C. & Ray, A. K. Brownian motion curve-based textural classification and its application in cancer diagnosis. *Anal. Quant. Cytol. Histol.* **33**, 158–168 (2011).
54. Krishnan, M. M., Venkatraghavan, V., Acharya, U. R., Pal, M., Paul, R. R., Min, L. C. et al. Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm. *Micron* **43**, 352–364 (2012).
55. Halicek, M., Shahedi, M., Little, J. V., Chen, A. Y., Myers, L. L., Sumer, B. D. et al. Head and neck cancer detection in digitized whole-slide histology using convolutional neural networks. *Sci. Rep.* **9**, 1–1 (2019).
56. Al-Ma'aitah, M. & AlZubi, A. A. Enhanced computational model for gravitational search optimized echo state neural networks based oral cancer detection. *J. Med. Syst.* **42**, 205 (2018).
57. Aubreville, M., Knipfer, C., Oetter, N., Jaremenko, C., Rodner, E., Denzler, J. et al. Automatic classification of cancerous tissue in laserendoscopy images of the oral cavity using deep learning. *Sci. Rep.* **7**, 1–10 (2017).
58. Song, B., Sunny, S., Uthoff, R. D., Patrick, S., Suresh, A., Kolar, T. et al. Automatic classification of dual-modality, smartphone-based oral dysplasia and malignancy images using deep learning. *Biomed. Opt. Express* **9**, 5318–5329 (2018).
59. Liu, Z., Wang, H. & Li, Q. Tongue tumor detection in medical hyperspectral images. *Sensors* **12**, 162–174 (2012).
60. Quang, T., Tran, E. Q., Schwarz, R. A., Williams, M. D., Vigneswaran, N., Gillenwater, A. M. et al. Prospective evaluation of multimodal optical imaging with automated image analysis to detect oral neoplasia in vivo. *Cancer Prev. Res.* **10**, 563–570 (2017).
61. Chakraborty, M., Mukhopadhyay, S., Dasgupta, A., Patsa, S., Anjum, N. & Ray, J. G. A new approach of oral cancer detection using bilateral texture features in digital infrared thermal images. In *38th Annual International Conference of the Institute of Electrical and Electronics Engineers (IEEE) Engineering in Medicine and Biology Society (EMBC)*. (IEEE, 2016).
62. Meehl, P. E. Clinical versus statistical prediction: a theoretical analysis and a review of the evidence. (University of Minnesota Press, 1954).
63. Egisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S. et al. The meta-analysis of clinical judgment project: fifty-six years of accumulated research on clinical versus statistical prediction. *Counseling Psychologist* **34**, 341–382 (2006).
64. Grove, W. M., Zald, D. H., Lebow, B. S., Snitz, B. E. & Nelson, C. Clinical versus mechanical prediction: a meta-analysis. *Psychol. Assess.* **12**, 19 (2000).



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.