Artificial Intelligence based Assessment of Oral Precancer to Aid Early Detection of Oral Cancer

Hanya Mahmood

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Supervisors: Professors Syed Ali Khurram, Nasir Rajpoot & Daniel Lambert

Academic Unit of Oral & Maxillofacial Surgery
Faculty of Medicine, Dentistry and Health
School of Clinical Dentistry, University of Sheffield

November 2023

Dedication

بِسۡمِ ٱللَّهِ ٱلرَّحۡمَٰنِ ٱلرَّحِيمِ

This thesis is dedicated to my parents, Rehana and Zaigham Mahmood.

My father, especially, has inspired me to continue a path of life-long learning; to discover, seek knowledge and acquire wisdom.

I am infinitely grateful for their unrelenting love, support and encouragement.

I hope I have made you both proud.

Acknowledgements

First and foremost, I thank Professor Syed Ali Khurram, my primary supervisor, for his exceptional supervision, training, and guidance throughout my PhD. Professor Khurram has given me invaluable opportunities to excel as a researcher and advance my clinical, academic and professional skills and learning. He has taught me the importance of pushing boundaries, and I cannot thank him enough for his persistent encouragement, kindness, and patience.

I also wish to thank Professor Nasir Rajpoot, for whom I have great admiration and respect. Professor Rajpoot has provided expert knowledge, direction and technical support in key aspects of my research, particularly in the application of computer vision.

I also extend my gratitude to a number of other individuals and teams who have made valuable contributions to this work: 1) NIHR Clinical Research Network and Research Design Service, for assistance with developing the initial fellowship proposal; 2) Sheffield Hospitals Charity, for funding preliminary research that led to this work; 3) Patient and Public Involvement groups (Sheffield Hospitals Patient Advisory Panel, Independent Cancer Patients' Voice Group, Yorkshire and Humber Consumer Research Panel, Swallows Head and Neck Cancer Group) for continuous feedback and engagement with my work; 4) Mr Mike Bradburn, from Sheffield Clinical Trials Research Unit, for ongoing statistical support; 5) Dr Adam Shephard, from Warwick University, for computational support; 6) Professor Zoe Marshman, my academic mentor, for pastoral care; 7) Professor Daniel Lambert, for support towards my professional development and 8) Dr Simon Atkins, my Educational Supervisor, for giving me the opportunity to maintain surgical exposure during this fellowship.

I must also acknowledge the NIHR for funding this research as part of a Doctoral Training Fellowship and giving me the opportunity to undertake this important research.

Finally, I thank my siblings and my husband, Arif Zafar, my biggest pillar of strength and positivity, without whom this journey would not have been possible. I am eternally grateful for your extraordinary love, compassion and kindness.

List of Outputs

Published articles related to this work:

- Hankinson P*, **Mahmood H*,** Walsh H, Speight PM, Khurram SA. Demystifying oral epithelial dysplasia: a histological guide. Pathology. 2023 Nov 16. **Joint first authorship.*
- Sathyamoorthy H*, **Mahmood H*,** Zubir AZ, Hankinson P, Khurram SA. Prognostic importance of mitosis quantification and PHH3 expression in oral epithelial dysplasia. Virchows Archiv. 2023 Oct 26:1-3. **Joint first authorship.*
- **Mahmood H,** Shephard A, Hankinson P, Bradburn M, Araujo AL, Santos-Silva AR, Lopes MA, Vargas PA, McCombe KD, Craig SG, James J, Brooks J, Nankivell P, Mehanna H, Rajpoot N, Khurram SA. Development and validation of a multivariable model for prediction of malignant transformation and recurrence of oral epithelial dysplasia. British Journal of Cancer. 2023 Sep 27:1-9.
- Bashir RM, Shephard AJ, **Mahmood H,** Azarmehr N, Raza SE, Khurram SA, Rajpoot NM. A digital score of peri-epithelial lymphocytic activity predicts malignant transformation in oral epithelial dysplasia. The Journal of Pathology. 2023 Feb 22.
- **Mahmood H,** Bradburn M, Rajpoot N, Islam NM, Kujan O, Khurram SA. Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models. Modern Pathology. 2022 Sep 1;35(9):1151-9.
- **Mahmood H,** Shaban M, Rajpoot N, Khurram SA. Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. British Journal of Cancer. 2021 Apr 19:1-7.
- **Mahmood H,** Shaban M, Indave BI, Santos-Silva AR, Rajpoot N, Khurram SA. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review. Oral Oncology. 2020 Nov 1;110:104885.
- Camalan S, **Mahmood H,** Binol H, Araújo ALD, Santos-Silva AR, Vargas PA, Lopes MA, Khurram SA, Gurcan MN. Convolutional Neural Network-Based Clinical Predictors of Oral Dysplasia: Class Activation Map Analysis of Deep Learning Results. Cancers. 2021; 13(6):1291.

Published abstracts related to this work:

- Shephard AJ, **Mahmood H,** Raza SE, Araujo AL, Santos-Silva AR, Lopes MA, Vargas PA, McCombe K, Craig S, James J, Brooks J. Transformer-based Model for Oral Epithelial Dysplasia Segmentation. arXiv preprint arXiv:2311.05452. 2023 Nov 9.
- Shephard AJ, Bashir RM, **Mahmood H,** Jahanifar M, Minhas F, Raza SE, McCombe KD, Craig SG, James J, Brooks J, Nankivell P. A Fully Automated and Explainable Algorithm for the

Prediction of Malignant Transformation in Oral Epithelial Dysplasia. arXiv preprint arXiv:2307.03757. 2023 Jul 6.

- Bashir RM, Shephard A, **Mahmood H,** Azarmehr N, Raza SE, Khurram A, Rajpoot N. A digital score of peri-epithelial lymphocytic activity predicts malignant transformation in oral epithelial dysplasia. medRxiv. 2023:2023-02.
- Azarmehr N, Shephard A, **Mahmood H,** Rajpoot N, Khurram SA. A Neural Architecture Search Based Framework for Segmentation of Epithelium, Nuclei and Oral Epithelial Dysplasia Grading. InMedical Image Understanding and Analysis: 26th Annual Conference, MIUA 2022, Cambridge, UK, July 27–29, 2022, Proceedings 2022 Jul 25 (pp. 357-370). Cham: Springer International Publishing.
- Shephard A, Azarmehr N, Bashir RM, Raza SE, **Mahmood H,** Khurram SA, Rajpoot N. A Fully Automated Multi-Scale Pipeline for Oral Epithelial Dysplasia Grading and Outcome Prediction. InMedical Imaging with Deep Learning 2022 Apr 22.
- Azarmehr N, Shephard A, **Mahmood H,** Rajpoot N, Khurram SA. Automated Oral Epithelial Dysplasia Grading Using Neural Networks and Feature Analysis. InMedical Imaging with Deep Learning 2022 Apr 22.
- Shephard AJ, Graham S, Bashir RM, Jahanifar M, **Mahmood H,** Khurram SA, Rajpoot NM. Simultaneous Nuclear Instance and Layer Segmentation in Oral Epithelial Dysplasia. arXiv preprint arXiv:2108.13904. 2021 Aug 31.
- **Mahmood H,** Shaban M, Indave I, Santos-Silva A, Rajpoot N, Khurram SA. Artificial intelligence to aid the diagnosis of head and neck precancerous and cancerous lesions: a systematic review. InVIRCHOWS ARCHIV 2020 Dec 1 (Vol. 477, pp. S37-S37).
- Bashir RS, **Mahmood H,** Shaban M, Raza SE, Fraz MM, Khurram SA, Rajpoot NM. Automated grade classification of oral epithelial dysplasia using morphometric analysis of histology images. InMedical Imaging 2020: Digital Pathology 2020 Mar 16 (Vol. 11320, p. 1132011).

Presentations to learned societies:

- **Mahmood H,** Rajpoot N, Khurram SA. 'Digital analysis of oral dysplasia to aid early cancer detection' (Oral Presentation). British Society for Oral and Dental Research (BSODR) Meeting, 6-8th September 2023, London.
- **Mahmood H,** Rajpoot N, Khurram SA. 'Development and validation of a multivariable model for prediction of malignant transformation and recurrence of oral epithelial dysplasia' (Poster presentation). Academy of Medical Sciences Clinical Academics in Training Annual Conference, 7th June 2023, Cambridge.
- **Mahmood H,** Rajpoot N, Khurram SA. 'Histological models for outcome prediction of oral epithelial dysplasia' (Oral Presentation). Association of British Academic Oral and Maxillofacial Surgeons (ABAOMS) Scientific Meeting, 24-25th November 2022, Glasgow.

- **Mahmood H,** Rajpoot N, Khurram SA. 'Quantitative histological analysis in oral epithelial dysplasia' (Oral Presentation) Pan European Region of the International Association of Dental, Oral, and Craniofacial Research (PER-IADR) Congress, 15-17th September 2022, Marseille
- **Mahmood H,** Rajpoot N, Khurram SA. 'Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models' (Oral Presentation). British Society of Oral & Maxillofacial Pathology (BSOMP) Scientific Meeting, 28-29th April 2022

Invited Talks

- 'Digital analysis of oral dysplasia for early cancer detection'. ABAOMS Conference, 9th November 2023, London.
- 'Machine learning based analysis of oral epithelial dysplasia for prediction of malignant transformation'. Tissue Image Analysis Centre Research Seminar Series, 20th March 2023
- 'AI in cancer: Fad or Future?' 13th Annual Royal Marsden Head & Neck Conference, 11th November 2022, London.

Prizes & Awards

- BSODR Senior Colgate Prize (BSODR Meeting, September 2023) with invitation to represent the British Division in the Hatton Competition at IADR 2024 in New Orleans, USA
- ABAOMS Best Oral Presentation Prize (ABAOMS Conference, November 2022)
- BSODR Oral Pathology / Oral Medicine Research Prize (PER-IADR Congress, September 2022)
- BSOMP Potts Prize for Best Research Presentation (BSOMP Conference, April 2022)
- First year PhD Research Presentation Runner-Up (School of Clinical Dentistry, March 2022)

Grants & Fellowships:

- Research Fellowship award for project entitled 'Single cell sequencing and spatial biomarker analysis in oral epithelial dysplasia' funded by Faculty of Dental Surgery (Royal College of Surgeons of England) & ABAOMS (2023)

Abstract

Background: Oral epithelial dysplasia (OED) carries an increased risk of malignant transformation to oral squamous cell carcinoma (OSCC), which is amongst the leading cancers worldwide. The diagnostic gold standard for OED is histopathological assessment and grading, which is challenging with unreliable behaviour and progression prediction. Advancements in digital pathology and Artificial Intelligence (AI) provide opportunities to uncover novel data from whole-slide imaging (WSI) through automated detection, pattern recognition and quantitative analysis.

Aims: This research uses a range of digital, computational, and quantitative approaches to reveal novel insights into OED progression and develops OSCC risk prediction models which are compared to existing clinical grading systems.

Methods: Retrospective samples of OED and non-dysplastic WSI (with five-year clinical follow-up) were used to develop malignant transformation prediction models based on analysis of conventional architectural and cytological histological features (n=109), novel digital morphometric features (n=100) and mitotic features assessed on haematoxylin and eosin (H&E) WSI and immunohistochemistry for Phosphohistone H3 (n=68). Machine learning models were trained to detect dysplastic, immune, and stromal cells in OED (n=220-248). Deep learning neural networks were trained to segment OED epithelium (n=434) and validated on external unseen datasets. Prognostic relationships were explored, and spatial analysis conducted.

Results: Six conventional histological features were significant for transformation ($p<0.036$) and recurrence ($p<0.015$). Significant differences in cytoplasmic eosin, nuclear eccentricity and circularity were seen in basal epithelial cells of OED ($p<0.05$). Nucleus circularity was associated with recurrence ($p=0.018$) and epithelial perimeter with malignancy ($p=0.03$). The developed models demonstrated better predictive strength for malignant transformation risk (AUROC:0.74-0.81) compared to 'gold-standard' histological grading (AUROC:0.60-69) with superior performance maintained on unseen external datasets. Trained AI models segmented and classified OED epithelium, immune and stromal cells with good accuracy (F1 scores:0.80-0.87). Peri-epithelial lymphocyte count was associated with malignant transformation and reduced progression free survival ($p<0.05$).

Conclusions: This novel research shows correlations between individual OED histological features, digital morphometric features, and prognosis for the first time on the largest digital multicentre cohort to date.

Table of Contents

List of Abbreviations

AC	Alternating current
AI	Artificial Intelligence
ANN	Artificial Neural Network
ANN-MLP	Artificial Neural Network-Multilayer Perceptron
ANOVA	One-way Analysis of Variance
AUROC	Area Under the Receiver-Operator Characteristic
CAMELYON	Cancer Metastases in Lymph Nodes
CNN	Convolutional Neural Network
CT	Computed Tomography
DCE-MRI	Dynamic Contrast Enhanced Magnetic Resonance Imaging
DL	Deep Learning
DNA	Deoxyribonucleic acid
EBV	Epstein-Barr Virus
EIS	Electrical Impedance Spectroscopy
FDA	Food and Drug Administration
FFPE	Formalin Fixed Paraffin Embedded Tissue
H&E	Haematoxylin and Eosin
HER2	Human Epidermal Growth Factor Receptor 2
HNC	Head and Neck Cancer
HPV	Human Papilloma Virus
HR	Hazard Ratio
HSI	Hyperspectral Imaging
IARC	International Agency for Research on Cancer
IHC	Immunohistochemistry
InHANSE	Institute of Head and Neck Studies and Education
LN	Layer Normalisation
ML	Machine Learning
MSA	Multi-head self-attention

NPV	Negative Predictive Value
OD	Optical Density
OED	Oral Epithelial Dysplasia
ONS	Office of National Statistics
OR	Odds Ratio
OSCC	Oral Squamous Cell Carcinoma
PANDA	Prostate Cancer Grade Assessment
PELs	Peri-epithelial Lymphocytes
PET-CT	Positron Emission Tomography
PHE	Public Health England
PHH3	Phosphohistone H3
PPV	Positive Predictive Value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROM1	Prominin-1
QUADAS	Quality of Diagnostic Accuracy Studies
RF	Random Forest
RNA	Ribonucleic acid
ROI	Region of Interest
SCC	Squamous Cell Carcinoma
SD	Standard Deviation
SE	Standard Error
TB	Toluidine Blue
TCGA	The Cancer Genome Atlas
TGF-β	Transforming Growth Factor-β
TNOM	Total Number of Mitoses
TRIPOD	Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis
US	Ultrasound
WHO	World Health Organisation
WSI	Whole Slide Image

List of Figures

List of Tables

Chapter 1 Introduction

1.1 Research Summary, Aims & Hypothesis

Oral epithelial dysplasia (OED) is associated with an increased prevalence of oral squamous cell carcinoma (OSCC) which is amongst the top ten cancers worldwide, and a leading cause of death. Since a large proportion of OED lesions can progress to OSCC (2.6% - 36%), early detection is critical to reduce cancer burden and improve patient outcomes.

Despite research advancements, there are still no proven biological or molecular markers predictive of malignant transformation of OED. Clinical and histopathological features, though informative, are unreliable indicators of behaviour and progression. Despite being the 'Gold Standard', OED grading has been criticised as being subjective, poorly reproducible and associated with significant intra-observer inconsistencies, making it inadequate for clinical management. Furthermore, little is known about the prognostic significance of individual histological features used in grading, and there are an increasing number of other cytological and architectural abnormalities in OED, which have not been quantitatively analysed or correlated to clinical outcomes before. As such, there is a clinical need to identify better predictive markers of transformation to guide treatment and aid early cancer diagnosis/detection.

Several diagnostic aids (such as vital tissue staining, chemiluminescence, exfoliative cytology, salivary diagnostics, impedance spectroscopy) have been developed and explored for early recognition of oral potentially malignant disorders (OPMD) and OSCC. However, due to limitations in their sensitivity and specificity, none of these tools have been employed as point-of-care tools and they lack strong supportive evidence for routine clinical use. Furthermore, these tools do not provide new knowledge or insight to advance our understanding of OED progression to malignancy.

Advancements in Artificial Intelligence (AI) provide unique opportunities to uncover novel diagnostic and prognostic patterns from whole slide images (WSI) of haematoxylin and eosin (H&E) stained tissue, through automated feature detection and quantification. The use of digital methods to aid early cancer detection has been recognised as an important area of research by the Department of Health, Cancer Research UK, and the TOPOL review[1,2,3] . More specifically, exploring effective methods for increasing early detection/diagnosis of OSCC and the utilisation of digital technologies to improve oral health has been highlighted amongst the top 10 Oral and Dental Health priorities by the NIHR James Lind Alliance[4,5]. This emphasises the importance and timeliness of this research, particularly in the current NHS workforce climate[6].

This research aims to use a range of digital, quantitative and computational approaches to study OED progression (Figure 1) and identify novel digital markers predictive of malignant transformation and recurrence. The research hypothesis is that AI and digital analysis can reveal novel prognostic information about OED progression to malignancy.

An overview of the study methods presented in this thesis are shown in Figure 2.

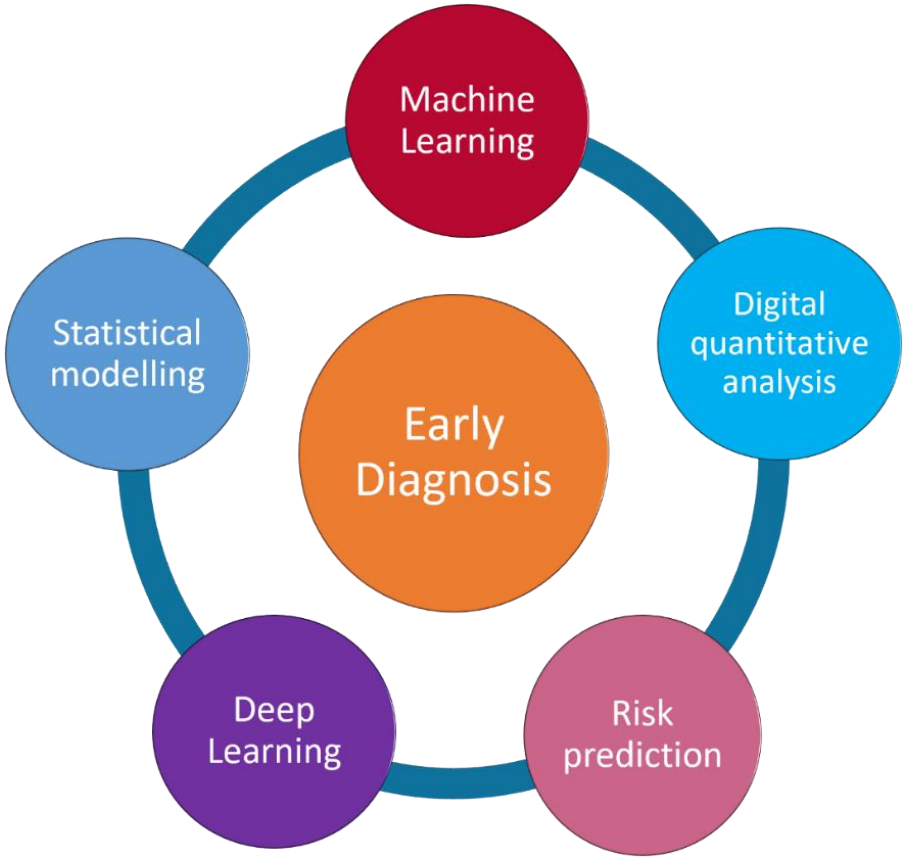

Figure 1. *Range of digital, computational, and quantitative methods used for early diagnosis.*

1.1.1 Research Objectives

The research hypothesis will be tested via the following objectives:

1. To perform a systematic review of the existing literature on the application of AI/ML methods for detection and grading of precancerous and cancerous head and neck lesions.
2. To explore the prognostic relationships of established histological features in OED using digital pathology and explore feature-specific scoring models.
3. To evaluate mitotic activity in OED based on H&E and immunohistochemistry for PHH3, and evaluate the prognostic importance of mitosis number, type and intraepithelial location.
4. To conduct digital quantitative analysis of novel morphometric features in OED and identify features important in prognosis prediction.
5. To develop multivariable models for prediction of malignant transformation and OED recurrence (for each objective 2, 3, 4), evaluate the impact of clinical variables and compare performance against existing clinical grading systems.
6. To train ML models for classification of dysplastic, immune and stromal cells, explore spatial patterns and conduct digital stromal analysis in OED.
7. To develop and test a novel DL pipeline for OED segmentation, classification and transformation prediction in H&E-stained whole slide images.

1.1.2 Research Methods

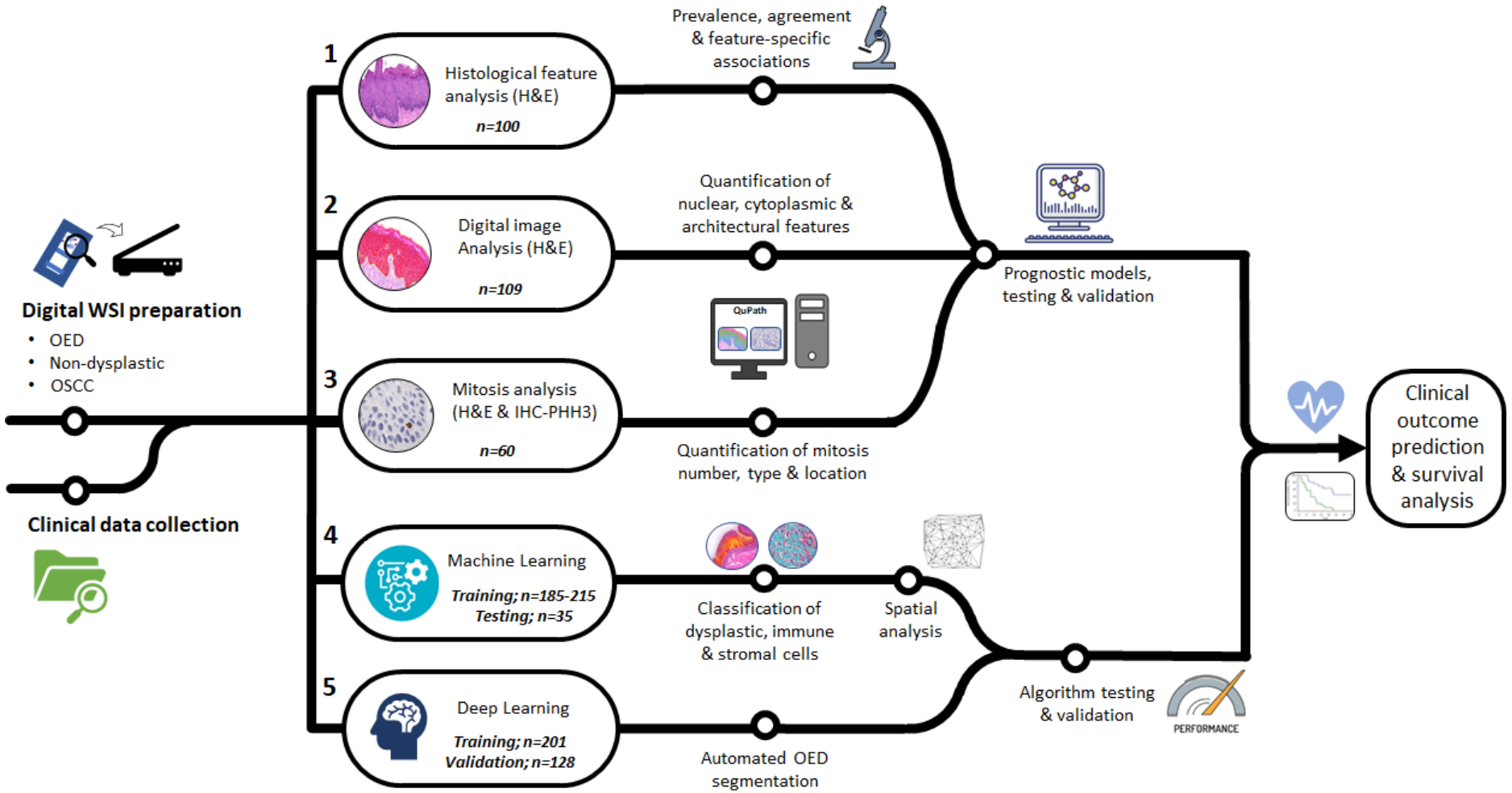

Figure 2. *Overview of research methods.*

(1) *Individual histological feature analysis in OED and feature-specific associations with clinical outcomes;* ***(2)*** *Digital quantification of novel morphometric features in OED;* ***(3)*** *Mitotic activity analysis in OED based on H&E and IHC-PHH3 analysis;* ***(4)*** *ML application for classification of cells and spatial analysis in OED;* ***(5)*** *DL application for automated OED detection and digital feature exploration.*

1.2 Thesis Contributions

This thesis takes the form of an alternative 'publication format thesis'. It is composed of nine chapters, including a collection of peer-reviewed publications, manuscripts prepared for journal submission and conventional monograph chapters. In brief, Chapter 2 provides a background to the research topic; Chapters 3-6 are reproduced by five peer-reviewed publications and a prepared manuscript; Chapters 7-8 comprise manuscripts prepared for journal submission incorporating materials from related publications in which the candidate has made significant contributions. Chapter 9 provides a general thesis discussion and conclusive summary to the presented research and highlights study limitations and future directions for research.

A chapter-by-chapter overview has been provided below with citations of the published works and the explicit contributions of the candidate. For clarity, this information has been re-presented at the beginning of each chapter.

The candidate confirms that the work submitted in this thesis is original and their own, and that they are aware of the University's Guidance on the Use of Unfair Means. Where reference has been made to the work of others, due credit has been given.

Chapter 2

This chapter provides a background to the subject matter including the global scale and epidemiology of oral cancer, complexities related to oral epithelial dysplasia (OED) diagnosis, emerging tools for early cancer detection and introduces digital pathology and artificial intelligence (AI).

Chapter 3

This chapter has been reproduced from three published narrative literature reviews. The first, a scoping review on AI-based methods in head and neck cancer diagnosis[7] published in the British Journal of Cancer (a Nature Journal) under the terms of a CC BY 4.0 license. The second, a systematic review conducted in conjunction with the International Agency for Research on Cancer (IARC) on the application and diagnostic accuracy of AI methods for detection and grading of potentially malignant and cancerous head and neck lesions[8] published in Oral Oncology. This article was reproduced in full print for non-commercial purposes, as permitted by Elsevier. The candidate is the first author for both these articles. Under the guidance of supervisors, the candidate developed the reviews and was responsible for conducting the electronic literature searches, article screening, data collection, narrative synthesis, and led on writing and editing of the manuscripts with contributions from co-authors.

The third paper is a review published in Journal of Pathology and has been reproduced under the terms of a CC BY-NC-ND 4.0 license. The candidate shares joint first authorship with Dr Paul

Hankinson, an Academic Clinical Fellow Oral Pathology at the University of Sheffield. The article discusses OED histological features and provides an overview of the common mimics of this condition. It also highlights the paucity of evidence defining these features and offers suggested definitions. The whole slide image examples from this article have been shared as a cloud-based open access dataset, available at: https://www.pathogenesis.co.uk/r/demystifying-dysplasia-histology-dataset.

Mahmood H, *Shaban M, Rajpoot N, Khurram SA. Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. British Journal of Cancer (2021)*

Mahmood H, *Shaban M, Indave BI, Santos-Silva AR, Rajpoot N, Khurram SA. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: a systematic review. Oral Oncology (2020)*

Hankinson P, ***Mahmood H,*** *Walsh H, Speight PM, Khurram SA. Demystifying oral epithelial dysplasia: a histological guide. Pathology. 2023 Nov 16.*

Chapter 4

This chapter is formed by two studies. The first, a study reproduced from an article published in Modern Pathology (a Nature Journal) under the terms of a CC BY 4.0 license. The study explores the prognostic relationships of conventional histological features in OED and develops feature-specific prognostic scoring models for clinical outcome prediction[9]. The second study aims to validate these feature-based prognostic scoring models and has been presented in a format ready for journal submission.

The candidate is the first author for both studies. Under the guidance of supervisors, the candidate was responsible for study design, obtaining ethical approval, sample retrieval and preparation, re-analysis of histological grading and clinical data collection. The candidate conducted the statistical analysis with support from a statistician at the Clinical Trials Research Unit, University of Sheffield and led on writing and editing of the manuscript with contributions from co-authors.

Mahmood H, *Bradburn M, Rajpoot N, Islam NM, Kujan O, Khurram SA. Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models. Modern Pathology (2022)*

Chapter 5

This chapter has been reproduced from an article published in the British Journal of Cancer (a Nature Journal) under the terms of a CC BY 4.0 license. The study explores the prognostic relationships of

novel digital morphometric features in OED, develops and externally validates a multivariable digital model for prediction of malignant transformation and OED recurrence[10].

The candidate is the first author for this publication. Under the guidance of supervisors, the candidate was responsible for study design, obtaining ethical approval, sample retrieval and preparation, clinical data collection, digital quantitative and statistical analyses, multivariable model development and validation and led on writing and editing of the manuscript with contributions from the co-authors.

Mahmood H, *Shephard A, Hankinson P, Bradburn M, Araujo AL, Santos-Silva AR, Lopes MA, Vargas PA, McCombe KD, Craig SG, James J, Brooks J, Nankivell P, Mehanna H, Rajpoot N, Khurram SA. Development and validation of a multivariable model for prediction of malignant transformation and recurrence of oral epithelial dysplasia. British Journal of Cancer (2023)*

Chapter 6

This chapter has been reproduced from an article published in Virchows Archiv under the terms of a CC BY 4.0 license. The study evaluates mitotic activity in OED assessed on haematoxylin and eosin (H&E) stained whole-slide images (WSI) and immunohistochemical staining for Phosphohistone H3 (PHH3)[11]. The candidate shares joint first authorship with Hrishikesh Sathyamoorthy, a previous MSc student at the University of Sheffield, for whom the candidate was a co-supervisor.

The candidate co-designed the study with the senior author, and was responsible for obtaining ethical approval, sample retrieval/preparation, clinical data collection and conducted the final statistical analysis that led to the development of the transformation risk prediction models presented in this paper. The candidate also led on writing and editing of the manuscript with contributions from the co-authors. Hrishikesh Sathyamoorthy is acknowledged for undertaking the immunohistochemistry and manual mitosis counting.

Sathyamoorthy H, ***Mahmood H,*** *Zubir AZ, Hankinson P, Khurram SA. Prognostic importance of mitosis quantification and PHH3 expression in oral epithelial dysplasia. Virchows Archiv. 2023 Oct 26:1-3.*

Chapter 7

This chapter takes the form of a manuscript in preparation for a journal submission, for which the candidate is the first and primary contributing author. The first part of the study explores the application of machine learning (ML) for classification of dysplastic, stromal and immune cells in OED. The second part evaluates spatial patterns in OED and conducts a digital quantification of stromal cellularity with analysis of nuclear features and exploration of prognostic relationships.

Under the guidance of supervisors, the candidate designed the study and was responsible for retrieval and preparation of the digital dataset, clinical data collection, training and testing of the ML models, quantitative analysis and downstream statistical evaluation. The candidate led on writing and editing of the manuscript with contributions from the co-authors.

Chapter 8

This chapter takes the form of a manuscript ready for journal submission, for which the candidate shares joint first authorship with Dr Adam Shephard, a data scientist at Warwick University. This study develops a novel deep learning (DL) pipeline for OED segmentation, classification and transformation prediction using H&E-stained whole slide images.

Under the guidance of supervisors, the candidate co-designed the study and was responsible for obtaining ethical approval, digital dataset preparation, clinical data collection and ground-truth slide annotations. The candidate worked closely with Adam Shephard to co-develop, train and test the DL models and conduct downstream statistical analysis. The candidate made a significant contribution to the writing and editing of the manuscript.

Chapter 9

This chapter provides a general thesis discussion, incorporating salient points from all aspects of the presented research. It describes the study limitations, challenges, ongoing work, and highlights opportunities for future research.

Chapter 2 - Background

2.1 Oral squamous cell carcinoma

2.1.1 State and scope of condition

Oral squamous cell carcinoma (OSCC) forms part of a larger group of diverse head and neck cancers (HNC) which are amongst the top ten leading groups of cancers worldwide, accounting for more than 650,000 cases and 33,000 deaths annually[12]. The majority of HNC are squamous cell carcinomas (SCC) (90%) which predominantly arise from the epithelial tissues of the oral cavity, nasal cavity, paranasal sinuses, pharynx and larynx[13]. OSCC are the most common type of HNC[14] with an increasing incidence and worsening prognosis[15,12]. It can affect any part of the oral cavity (tongue, floor of mouth, buccal/labial mucosa, palate, gingivae, tonsils)[14] (Figure 3) and smoking[16], betel quid chewing[17] and high alcohol intake[18] are the three most well-recognised risk factors by the World Health Organisation (WHO). OSCC is often preceded by abnormal maturation of the epithelial lining and stratification of the surface epithelium, a disorder known as oral epithelial dysplasia (OED) [19]. OED describes a spectrum of histological changes, both architectural and cytological, which is usually triggered by an accumulation of genetic alterations, chronic exposure to environmental carcinogens or viral infections[19].

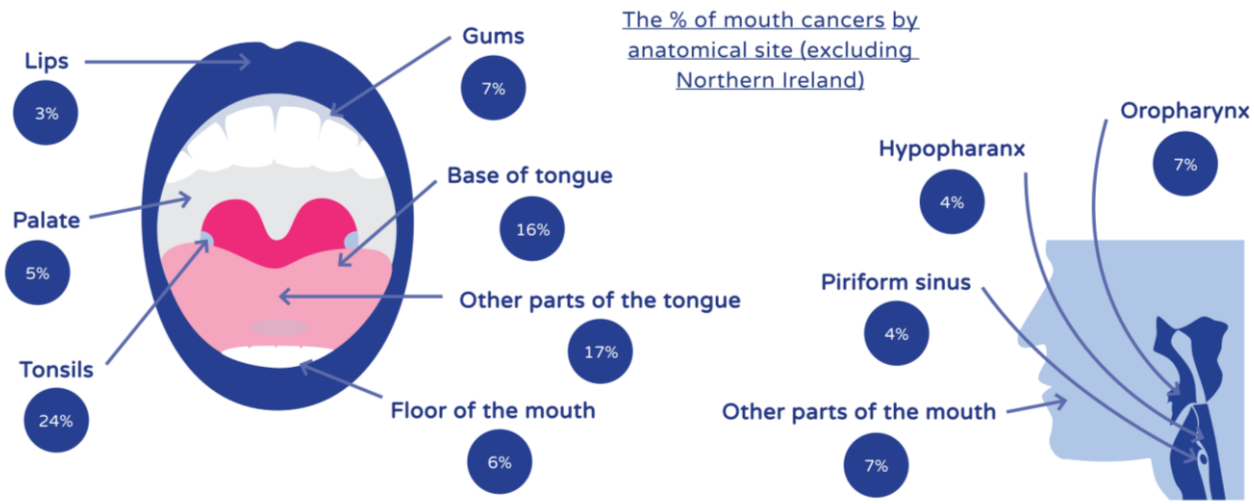

Figure 3. *The percentage of mouth cancer by anatomical site.*

Original source: State of Mouth Cancer UK Report, 2022. Reproduced with permission of the rights holder, Oral Health Foundation.

In early stages, OSCC can be treated with surgical resection and targeted radiotherapy. If, however, there are delays in diagnosis and/or treatment, OSCC can rapidly progress causing significant destruction of tissues and bone in the oral and maxillofacial region[20]. Survival is directly linked to its clinical stage and spread (metastasis) to regional or distant structures, therefore early recognition and treatment is critical[6]. The effects of surgery and oncological treatment can be severe and debilitating, resulting in a range of life-changing morbidities including difficulties in mastication, speech, swallowing, altered facial appearance and a reduced quality of life[21].

2.1.2 Epidemiological trends

Almost half a million new cases of OSCC are diagnosed globally each year. The Office for National Statistics (ONS) registry reports OSCC figures as a single grouping of 'lip, oral cavity, and pharynx' (pharynx including oropharynx, nasopharynx and hypopharynx) and Table 1 demonstrates age-standardised incidence rates per 100,000 person-year using the 2013 European Standard Population[22]. Statistics reveal a total of 42,997 cases of oral and oropharyngeal cancers between 2000 and 2016[22]. Although due to the continuing accrual of late registrations and 'opt out' registration by some individuals, in addition to the exclusion of data from private hospitals, primary care services and nursing homes, these figures are likely to underrepresent the actual numbers.

The recent State of Mouth Cancer UK report (2022) published by the Oral Health Foundation uses ONS data and reports >8,000 new diagnoses of OSCC each year in the UK, reflecting a 34% rise in the last decade and almost 103% rise compared with 20 years ago[23]. With time, it is expected that OSCC incidence will continue to increase, with 9,200 annual cases predicted in the UK by 2030[24].

The International Classification of Diseases version 10 (lip, oral cavity and pharynx, C00-C14 and oral cavity, C00-C06), Public Health England (PHE) reports 10,908 deaths from OSCC alone between 2012 and 2016[22,25], excluding head and neck precancers and other malignancies of bone, connective tissue and soft tissue. Despite the rising mortality from OSCC, there are few studies reporting the effect of independent risk factors on mortality rates in England. A cluster-randomised controlled trial conducted in India showed that regular oral screening could lower death rates amongst high-risk individuals with the potential to prevent approximately 37,000 deaths from OSCC globally[26]. The study showed greater death rates amongst those who smoke and drink excessive alcohol (mortality rate ratio 0·66 [95% CI 0·45–0·95])[26].

Table 1. *Age-standardised incidence rates per 100,000 person-year.*

Original source: European Standard Population. Office for National Statistics, England (2000-2016), reproduced under an Open Government Licence v3.0.

			2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Oral cavity cancer	**Sex**																		
	Male	Count	1,177	1,174	1,067	1,233	1,219	1,237	1,369	1,435	1,511	1,584	1,601	1,603	1,776	1,843	1,779	1,838	1,779
		Rate	6.2	6.0	5.4	6.2	6.0	6.1	6.6	6.8	7.0	7.3	7.2	7.1	7.8	8.0	7.5	7.7	7.3
	Female	Count	807	807	757	837	856	878	948	972	1,116	1,071	1,144	1,152	1,249	1,292	1,309	1,248	1,309
		Rate	3.5	3.4	3.2	3.5	3.6	3.7	3.9	4.0	4.5	4.3	4.5	4.5	4.8	4.9	4.9	4.6	4.8
	Person	Count	1,984	1,981	1,824	2,070	2,075	2,115	2,317	2,407	2,627	2,655	2,745	2,755	3,025	3,135	3,088	3,086	3,088
		Rate	4.7	4.7	4.2	4.8	4.8	4.8	5.2	5.3	5.7	5.7	5.8	5.8	6.3	6.4	6.2	6.1	6
Oropharyngeal cancer																			
	Male	Count	745	790	814	905	912	992	1,128	1,137	1,299	1,381	1,540	1,584	1,740	1,823	1,891	2,140	2,265
		Rate	3.7	3.9	4.0	4.4	4.4	4.6	5.2	5.2	5.9	6.1	6.7	6.8	7.4	7.6	7.8	8.7	9.1
	Female	Count	284	283	307	309	340	369	385	387	402	468	533	505	570	591	594	674	712
		Rate	1.2	1.2	1.3	1.3	1.5	1.6	1.6	1.6	1.7	1.9	2.1	2.0	2.2	2.3	2.3	2.6	2.7
	Person	Count	1,029	1,073	1,121	1,214	1,252	1,361	1,513	1,524	1,701	1,849	2,073	2,089	2,310	2,414	2,485	2,814	2,977
		Rate	2.4	2.5	2.6	2.8	2.8	3.0	3.3	3.3	3.7	3.9	4.3	4.3	4.7	4.9	4.9	5.5	5.8

2.1.3 Public awareness and early detection

In recent years there has been an attempt to improve patient and public awareness of OSCC through campaigns such as 'mouth cancer action month', and government legislations such as public smoking bans and the uptake of Human Papillomavirus (HPV) vaccinations[23]. That said, an alarming proportion of the UK population (80%) do not ever recall seeing a public health message around oral cancer[23]. Whilst general awareness may be slowly increasing (88% of UK adults have at least now heard of mouth cancer), awareness of the major signs and symptoms of OSCC are as low as 17%, and awareness around risk factors is as low as 9%[23] indicating the need for improvements.

Aside from poor public awareness, there are several other major barriers to dental access, including the rising cost of living which may deter people from making dental appointments, the dwindling NHS dental workforce resulting in a national dental access crisis, and the ongoing difficulties faced by special population groups such as the elderly, people with mobility difficulties, and individuals living in geographically isolated areas or from minority ethnic backgrounds[27]. These factors all contribute to late detection, which means OSCC may have already developed or progressed to a higher clinical stage (with metastasis) at the time of presentation, significantly reducing the success of treatments and patient survival[14,23]. Statistics indicate that >3,000 people die from OSCC each year in the UK[23] and the ten-year survival rate could be as low as 19% depending on the stage of disease at initial presentation[28]. Early detection of OSCC has been directly linked to improved survival[23] with 80% chance of survival for stage I OSCC compared to 20-30% for stage IV OSCC[29].

Improving access to NHS dental care, recognising early signs of cancer, and improving national public health policies remain a key focus in the prevention and early detection of OSCC.

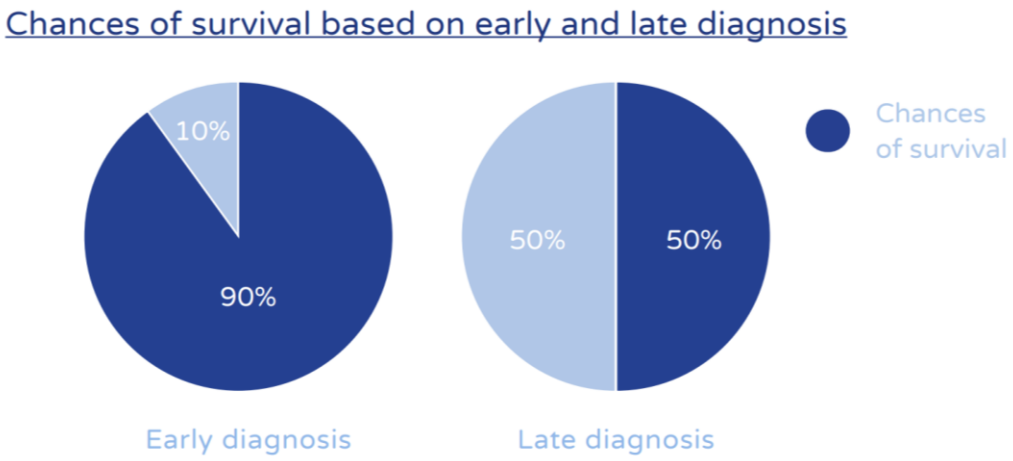

Figure 4. *Chances of survival based on early and late diagnosis.*

Original source: State of Mouth Cancer UK Report, 2022[23]. Reproduced with permission of the rights holder, Oral Health Foundation.

2.1.4 Risk factors

Tobacco smoking/chewing, high alcohol consumption and use of areca (betel) nut or paan are three major risk factors for OED and OSCC development[17,18]. Secondary risk factors include gamma and ultraviolet radiation, heightened exposure to sunlight, immunodeficiency, and a positive family history of HNC[30]. The Human Papilloma Virus (HPV) and Epstein-Barr virus (EBV) have been implicated in the development of oropharyngeal[31] and nasopharyngeal SCC[32] and may also play a role in the development of some cases of OSCC[16].

The reported incidence of OSCC amongst males (68%) is greater than females, and there is a strong age-related association with people aged >55 (80%)[23]. The latter is thought to be related to a greater exposure to environmental risk factors, genetic alterations and immunosenescence over time[33,34]. Living in areas of greater social deprivation and lower socioeconomic groups also presents an increased risk, with a rise in incidence by 68% for those living in the most deprived parts of the UK[23]. This is likely to be related to a lack of awareness of OSCC risk factors, signs and symptoms[23].

The effect of independent risk factors on OSCC is well documented. Pooled analysis from large case-control studies included in the International Head and Neck Cancer Epidemiology Consortium (n=25,500+ HNC patients and 37,100 controls) confirms the mutagenic effects of tobacco and alcohol to be dependent on dose, frequency and duration of use with increased risk with concurrent use[35,36]. Smoking duration was shown to be more important than smoking frequency, although alcohol frequency was shown to be more important than duration of alcohol use[37]. Those drinking in excess of 1.5 – 6 units per week were shown to have an increased risk by 81%[38]. Another multicentre case-controlled study which included fourteen different England hospitals found significantly raised odds ratios (OR) (14.3 [95% CI: 1.1–178.8]) amongst males (mean age of 38.5 years, SD ¼ 7.0) with OSCC if they had started smoking before 16 years of age, and there was a significant risk reduction also seen in ex-smokers (OR ¼ 0.2; 95% CI: 0.5–0.8) [39]. Consumption of alcohol showed an increased risk amongst males ≤45 years (OR ¼ 8.1; 95% CI: 1.6– 40.1) as did excessive alcohol drinking or having ever smoked (OR ¼ 4.4; 95% CI: 1.1–17.7)[39]

2.2 Oral epithelial dysplasia

OED is a precursor epithelial disorder of the oral mucosa, associated with a statistically increased risk of malignant progression to OSCC[19]. The literature reports 2.6% - 12.1% of all OED lesions progress to OSCC[40,41], highlighting the importance of early and accurate detection.

Clinically, OED can have a variable presentation which can make diagnosis challenging[42]. It can arise at the margins of ulcers or present as a persistent white patch (leukoplakia), red patch (erythroplakia) or a mixed white-red patch (erythroleukoplakia) (Figure 5). Leukoplakia is regarded as the most common clinical presentation, with OED present in almost 50% of biopsied lesions[42]. Malignant transformation rates of leukoplakia vary between populations but are reported at approximately 9.5%

(99% CI 5.9% to 14.00%) or 1.56% per year[43]. OED may also be seen in other oral potentially malignant disorders (OPMD) such as oral submucous fibrosis, actinic keratosis, and oral lichen planus[44,45]. The presence of OED in these lesions is associated with an increased risk to malignancy[46]. Whilst attempts have been made to correlate certain clinical features in OED to malignant transformation (lesion size ⩾200 mm, speckled, nodular or verrucous appearance, presence of multiple lesions, lateral/ventral tongue or floor of mouth subsites)[47,48,49] the supportive evidence is weak, and there is no single definitive feature shown to be reliably predictive[50,51]. OSCC may also arise from seemingly healthy, non-dysplastic oral epithelium[46]. The matter is further complicated by the lack of clinically proven predictive biomarkers and uncertainty of OED grading[54].

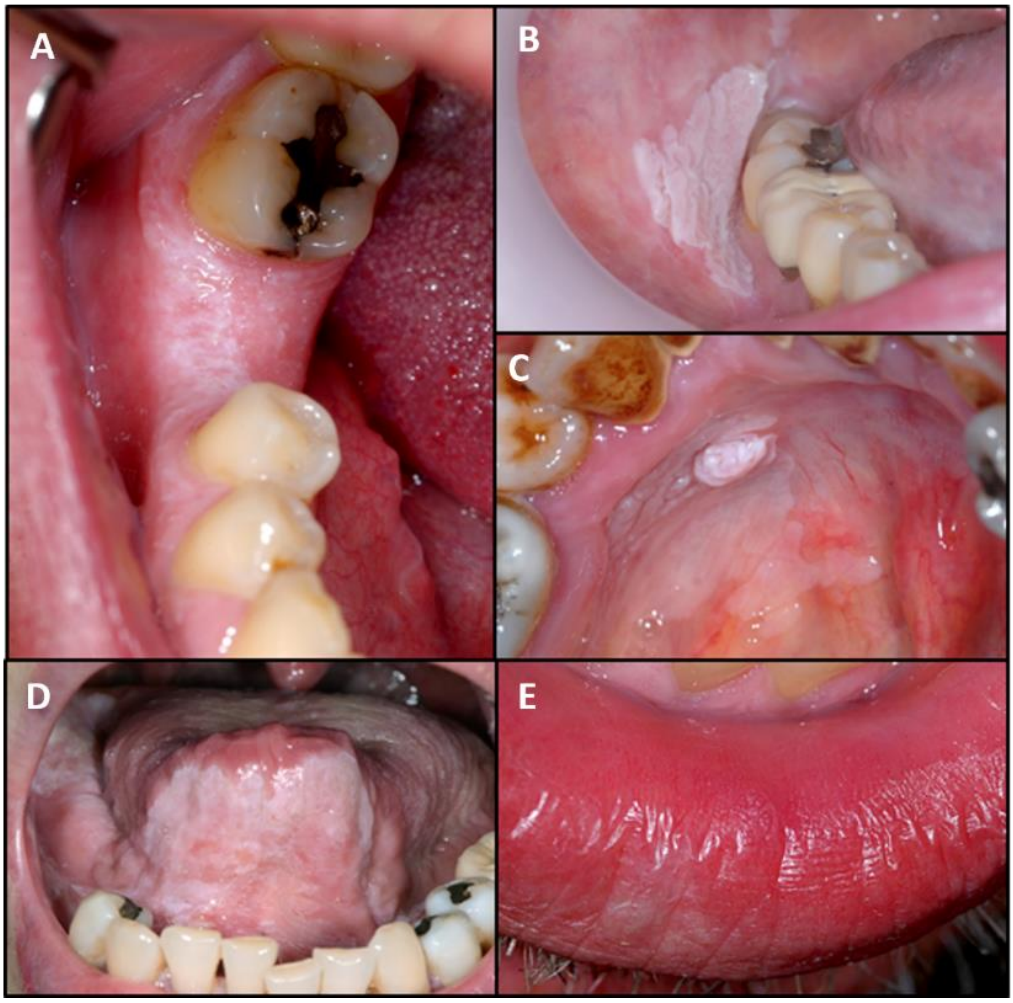

Figure 5. *Clinical photographs demonstrating variable appearances of OED.*

A) Diffuse leukoplakia involving gingival tissues; B) Focal raised leukoplakia with corrugated surface on right buccal mucosa; C) Raised leukoplakia with nodular area in floor of mouth; D) Diffuse erythroleukoplakia on ventral tongue extending to floor of mouth; E) Ill-defined leukoplakia with surrounding erythema on lower lip. Images acquired for purpose of thesis, with ethical approval (Ref: 18/WM/0335).

2.2.1 Histological grading

Diagnosis of OED is currently achieved through histopathological analysis of a biopsy tissue sample using light microscopy. Over the years, there have been several grading systems proposed with numerous iterations. The most widely accepted system is the World Health Organisation (WHO, 2017) classification which categorises lesions into 'mild', 'moderate' and 'severe' grades based on the presence and severity of a wide range of histological abnormalities and the extent to which these features extend upwards through the epithelial layers[52]. This is achieved by 'splitting' the epithelium into 'thirds' (Figure 6) [53]. Using this approach, a diagnosis of mild dysplasia is made if the dysplastic changes are confined to the basal and parabasal layers (lower third), whereas in moderate dysplasia the changes extend to involve the middle third of the epithelium, and in severe dysplasia the changes extend through the entire thickness (or at least more than half) of the epithelium (upper third) (Figure 6). However, this approach is arguably over simplistic, since the presence of a single feature in abundance, irrespective of its location in the epithelium, may be sufficient to upgrade a lesion. In the most recent iteration of the WHO classification of head and neck tumours, the range of features used to grade OED has been expanded to 28 features in total[52]. Not only is it difficult and impractical for pathologists to analyse such a wide variety of features, but it also increases the risk of inaccuracies and widens variability and grading consensus between pathologists[54].

In the early stages of OED, cessation of risk factors can result in regression of the lesion, but if there is constant exposure to carcinogens that promotes genetic alterations and chromosomal instability, histological atypia may progress to a higher grade or transform to malignancy. A recent meta-analysis showed that moderate/severe OED was associated with a greater risk of malignant transformation risk compared to mild OED with an odds ratio of 2.4 (99% CI1.5-3.8)[43]. However, due to lack of consensus on the most clinically appropriate grading system[55] histological grading is not a reliable indicator of malignant transformation. The mechanisms by which dysplasia progression (and regression) occurs is still unclear, and there remains a lack of understanding about which individual histological features have the greatest (and least) prognostic value.

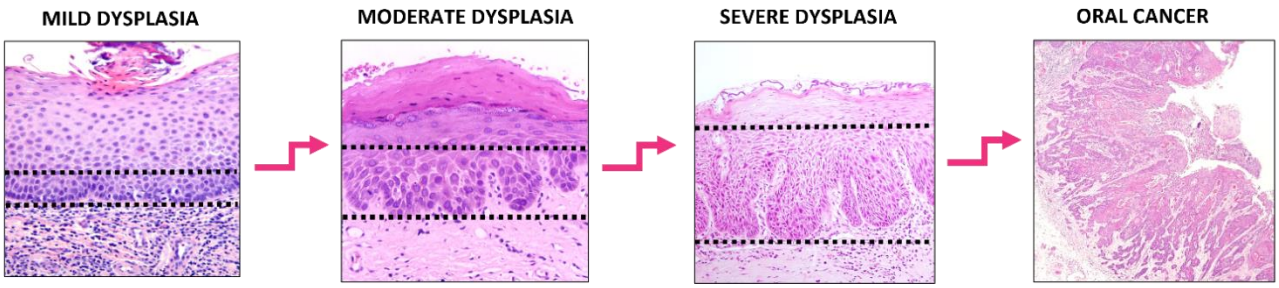

Figure 6. *A simplified overview of histological progression of OED to malignancy.*

Black dotted line demonstrates division of the epithelium into 'thirds' for grading lesions. Note: lesions do not always need to progress through to the worst OED grade for transformation.

2.2.2 Diagnostic challenges and alternative systems

There are several limitations of the WHO (2017) classification, despite it being in routine use[56]. Firstly, there is significant inter- and intra-observer variability amongst pathologists. This variability may arise due to the wide range of histological features (Figure 7, Figure 8, Figure 9) that can be seen in OED, subjectivity in dividing the epithelium into 'thirds' (Figure 6) as well as an inherent ambiguity of certain ill-defined features. As a result, there is likely to be differences in interpretation and agreement of features, which may lead to inconsistencies in grading and risk stratification of OED lesions[57].

Secondly, the classification system does not ascribe prognostic weight to individual features. This is likely due to a limited knowledge of which histological features (individually or in combination) are important in malignant progression. This means that even when grading is done correctly, it may not reliably predict prognosis, since mild grade lesions may transform, whereas some severe grade lesions may remain static[4,10].

Thirdly, the WHO (2017) system relies on the histological identification of architectural and cytological abnormalities, but many of these features are relatively non-specific and overlapping and may be present in a host of other conditions which mimic dysplasia, such as reactive atypia in inflammatory and ulcerative conditions, or fungal infections[28]. In addition to this, whilst some cases of OED are straightforward with obvious cytological abnormalities, the so called 'differentiated' or 'architectural dysplasia' tends to lack frank cytological atypia with architectural changes being the predominant feature, making these cases more challenging to diagnose[54]. A more detailed review attempting to demystify OED features is presented in the Appendix.

Finally, not all OSCCs are preceded by OED[46], suggesting that grading is less important than the biological mechanisms underpinning the disease process itself. It is well documented that OED development occurs due to a complex interaction between genetic and molecular alterations and histological atypia, but the exact mechanism of its progression to cancer is still poorly understood[58,59].

To overcome grading irreproducibility, an alternative binary grading criteria has been proposed[60]. This system uses the same histological features as that listed in the WHO criteria, but grades lesions based on the total number of features present ("low" risk: < 4 architectural features; < 5 cytological features and "high" risk: ≥ 4 architectural features; ≥ 5 cytological features). This system does not consider verrucous lesions though, or the extent of features present, limiting its clinical use. Some studies have demonstrated this system to have improved reproducibility in comparison to the WHO classification (unweighted and weighted kappa agreements for WHO grading Ks = 0.22 [95% CI: 0.11-0.35], Kw = 0.63 [95% CI: 0.42-0.78], respectively, versus K = 0.50 [95% CI: 0.35-0.67] for binary system) [60,61]. However, the strength of evidence to support its clinical utility is weak and conflicting. A recent systematic review and meta-analysis, comparing the binary classification with the WHO classification showed inconclusive results with regards to its prognostic value[62]. As such, this system is not deemed robust enough for routine clinical use on its own at present.

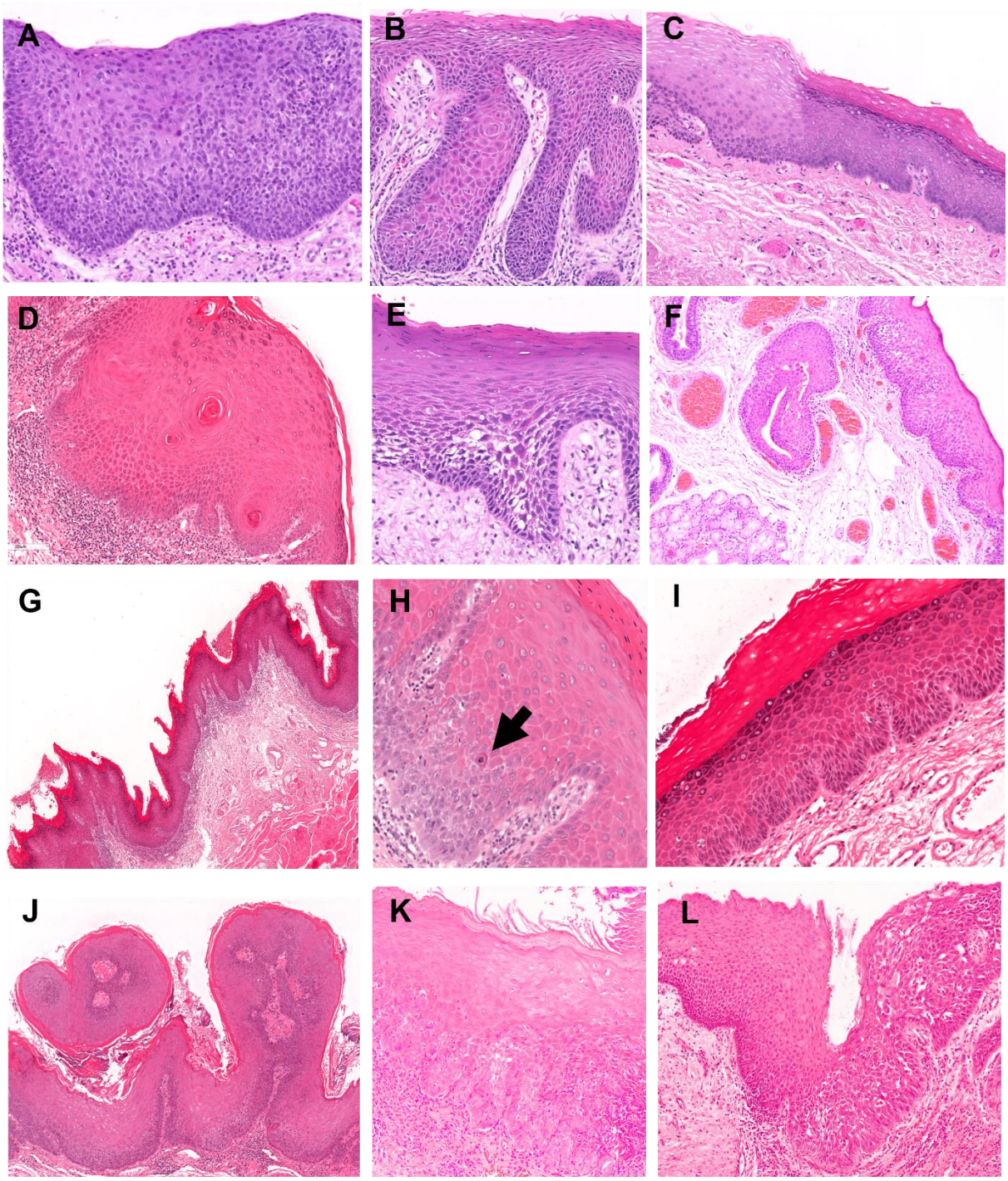

Figure 7. *Architectural features seen in OED.*

A) Irregular stratification; B) Premature keratinisation; C) Abrupt transition; D) Keratin pearl formation; E) Loss cell epithelial cohesion and basal cell polarity; F) Extension of OED along a salivary duct; G) Verrucous surface architecture; H) Superficial mitosis – black arrow; a mitotic figure is identified outside of basal compartment. I) Bulbous rete processes and generalised premature keratinisation; J) Papillary architecture; K) Basal cell nesting and clustering; L) Multiple patterns of dysplasia. Original magnifications: A, B, C, D, H, I, K, L (x10), E (x20), F, G, J (x4).

Original source: Hankinson P, Mahmood et al (2023)[63] reproduced under a CCBY-NC-ND license.

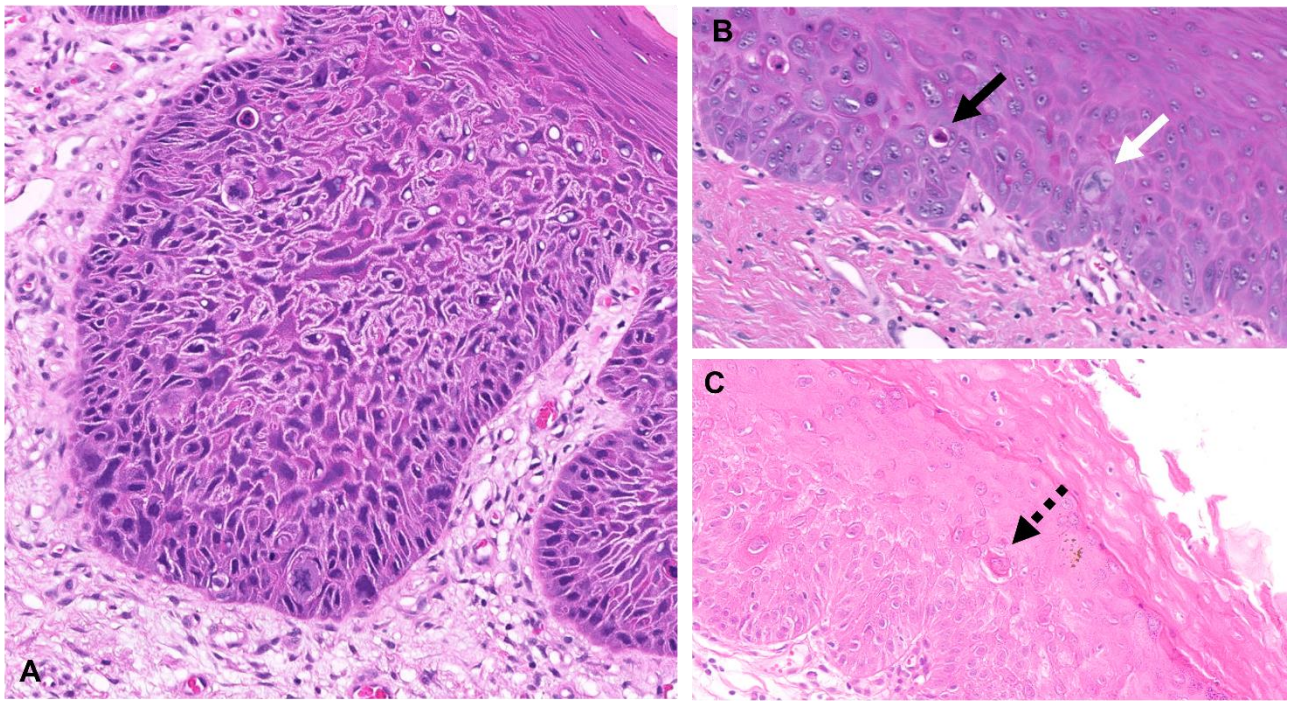

Figure 8. *Cytological features seen in OED.*

Severe OED; an abundance of cytological features of OED are seen in this bulbous rete process including: abnormal variation in nuclear size and shape, abnormal variation in cell size and shape, increased mitotic activity, increased nuclei: cytoplasm ratio, atypical mitotic figures, increased number and size of nucleoli, nuclear hyperchromasia and abnormal mitoses, B) Abnormal mitosis and apoptosis; black arrow – shows an apoptotic cell with pyknotic nucleus, brightly eosinophilic cytoplasm and retraction from neighbouring keratinocytes, white arrow – showing an abnormal mitotic figure with asymmetrical chromatin, C) Single cell keratinisation – black dotted arrow; there is generalised premature keratinisation seen in this example of OED giving the epithelium a strongly eosinophilic appearance, the black arrow highlights single cell keratinisation with even more eosinophilic cytoplasm and retraction from adjacent keratinocytes. Original magnifications: A, B (x20), C (x10).

Original source: Hankinson P, Mahmood et al (2023)[63] reproduced under a CCBY-NC-ND license.

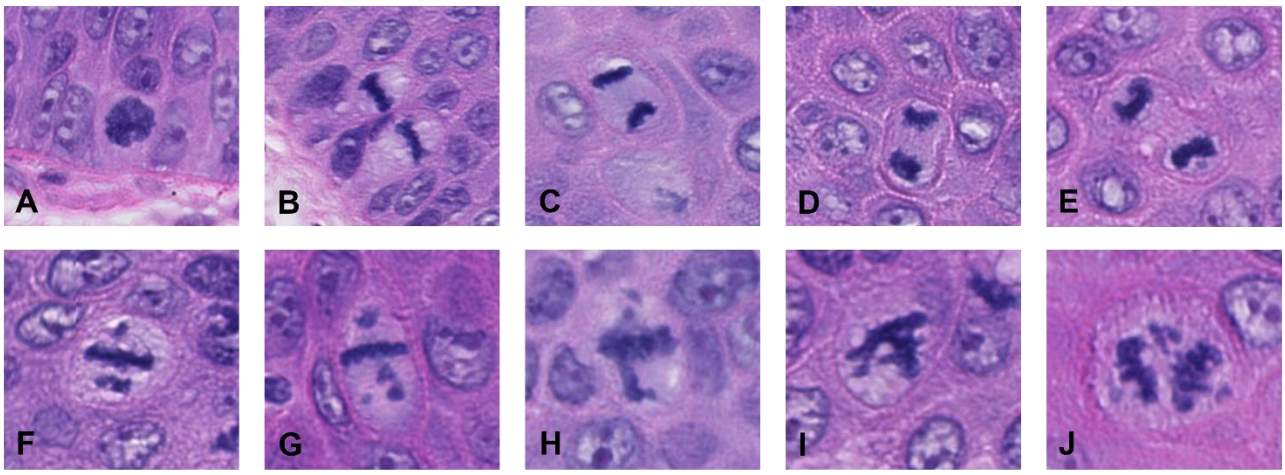

Figure 9. *Variations in mitotic figures.*

Normal (A-E) and abnormal mitotic figures (F-J) from OED (original magnifications x40). A) Prometaphase mitosis in a basal keratinocyte, B) Two basal keratinocytes in metaphases of mitosis, C+D) Anaphase mitoses in prickle cell layer keratinocytes, E) Early telophase mitosis in prickle cell layer F+G) Abnormal mitosis in OED showing abnormal chromosome segregation, fragments of chromatin are separate from the larger aggregate, H-J) Abnormal mitosis in OED showing mitotic/polar asymmetry with chromosomes aligning along more than two spindle poles, if these cells survive mitosis they will become aneuploid if not already as the distribution of chromosomes across daughter cells is uneven.

Original source: Hankinson P, Mahmood H et al (2023)[63] reproduced under a CCBY-NC-ND license.

2.2.3 Biomarkers

The process by which OED transforms to cancer is poorly understood[54]. It is thought that OSCC occurs through an aggregation of progressive molecular changes in key tumour suppressor genes and oncogenes. This is supported by the marked genetic alterations and chromosomal derangements that occur in OED as the lesion progresses, until eventually there is a final malfunction in the genotype to trigger cancer and invasion[58]. Whilst an increasing number of biomarkers have been implicated in OED, none of these have been clinically proven to predict oral carcinogenesis[40,64].

An extensive systematic review and meta-analysis identified four biomarkers in OED to be statistically associated with OSCC development[64]. This included loss of heterozygosity particularly at the 3p ± 9p loci (RR 17.60 (2.77, 108.37) $p < 0.001$), survivin (RR 30 (4.25, 197.73), $p \leqslant 0.001$), matrix metalloproteinase (RR 19.00 (1.56, 209.38) $p = 0.02$) and DNA content (RR 12.00 (1.17, 82.10) $p = 0.03$). Other analysed biomarkers (p53, p73, MMP 1 and 2 and cathepsin L mRNA) were weakly associated with malignant transformation[64]. These findings should be treated with caution though, due to methodological weaknesses among the included studies, which largely comprised small, retrospective, single-centre studies.

In another systematic review, the over-expression of ALDH1A1, PROM1 and PDPN[65] was linked to an increased malignant potential of OED. However, this review was based on the inclusion of just four studies with a weak overall strength of evidence[66]. ALDH1A1 has demonstrated a predictive role in advanced dysplasia of Barrette's oesophagus as well as high-grade cervical dysplasia and lung, oesophageal and breast malignancies[67,68,69]. Inhibition of PROM1, a membrane glycoprotein, has been shown to interfere with angiogenesis and cell proliferation which could be relevant in cancer development[70]. Similarly, PDPN, a specific marker for lymphatic endothelial cells, has been shown to play a role in the promotion of cancer cell clonal capacity, invasion, and metastasis[71]. However, further work is still required to validate the significance of these proteins in the prediction of OED progression to OSCC.

To summarise, the existing body of research exploring predictive biomarkers in OED is largely of mechanistic exploration, with weak clinical application and contribution to the overall understanding of OED progression. The lack of proven biomarkers means that histological grading is the main measure of cancer risk prediction, and therefore used by surgeons to inform treatment decisions[4]. This reinforces the need to explore new predictive markers to help improve the prognostic strength of existing diagnostic criteria's and optimise management protocols for patients with OED.

2.3 Diagnostic aids for early cancer detection

Visual and tactile oral screening is regarded an important step in the early detection of OPMD. However, as clinical appearances of OPMD are variable, it can be challenging for clinicians to know which lesions are at higher risk of progression to malignancy. As such, patients will often undergo further investigation in the form of a surgical biopsy to confirm the histopathological diagnosis. Whilst oral biopsies are regarded as minor surgical procedures, they can be unfavourable to patients, and are associated with surgical risks such as post-operative pain, swelling, bleeding, bruising, infection, scar and nerve injury. It is also possible that a biopsy sample does not provide the necessary information to confirm a definitive histological diagnosis. For example, if the biopsy sample is too small, of poor quality, or does not capture the severity of the lesion due to a sampling error. Multiple biopsies can be difficult to perform, can increase surgical risk, and are often not tolerated well by the patient.

Several suggestions have been proposed to overcome the reliance on histological grading, such as the use of molecular markers[72,73], morphological descriptors[74] and computer aided analyses[75]. To aid oral visual examination, several adjunctive techniques have been explored (Table 2). These methods include vital tissue staining with toluidine blue (TB) or Lugol's iodine, acetowhite staining of tissues, chemiluminescence, veloscope visualisation, OralCDx brush test and impedance spectroscopy. Some of these methods, along with their limitations, have been described in further detail below.

Table 2. *Diagnostic aids for early oral cancer detection.*

Modified from Masthan et al. (2012)[76]

Clinical Methods
Vital staining - Toludine Blue, Lugol's Iodine
Cytopathology – Exfoliative Brush Biopsy (Oral CDX)
Salivary transcriptome diagnostics
Molecular methods - tumour markers & biomarkers
DNA Ploidy
PCR-based diagnostics
Visualisation/Photo Aids
Chemiluminescence -ViziLite
VELscope
In vivo Confocal Microscopy
Autofluorescence Spectroscopy
Electrical Methods
Electrical Impedance Spectroscopy
Digital Methods
Digital Pathology
Quantitative digital analysis
Computational approaches – Artificial Intelligence/Machine Learning/Deep Learning

2.3.1 Toluidine blue staining

The use of TB staining as an adjunctive tool for early OSCC detection was first proposed in the early 1980s. TB is an acidophilic metachromatic dye which forms part of the thiazine group. It selectively stains acidic tissue components and is partially soluble both in water and alcohol. Since neoplastic cells contain higher concentrations of nucleic acid, the dye works by positively binding to deoxyribonucleic acid (DNA) or ribonucleic acid (RNA) within these cells to give a dark blue colour, which can improve visualisation of high-risk OPMD and OSCC[77].

Whilst a number of small-scale studies have reported high sensitivity rates for TB, this relates mainly to the assessment of 'high-risk' oral lesions, many of which will already have carcinoma in situ (Figure 10). A systematic review demonstrated a variable sensitivity of TB staining from 38% to 98% and an even more variable specificity ranging from 9% to 93%[78]. TB was associated with a high false-positive rate, a positive predictive value (PPV) between 33% to 93% and negative predictive value (NPV) between 22% to 92%[78]. As such, TB is not widely employed in clinical practice, and may have greater use as a diagnostic aid in remote regions where access to dental care is more restricted.

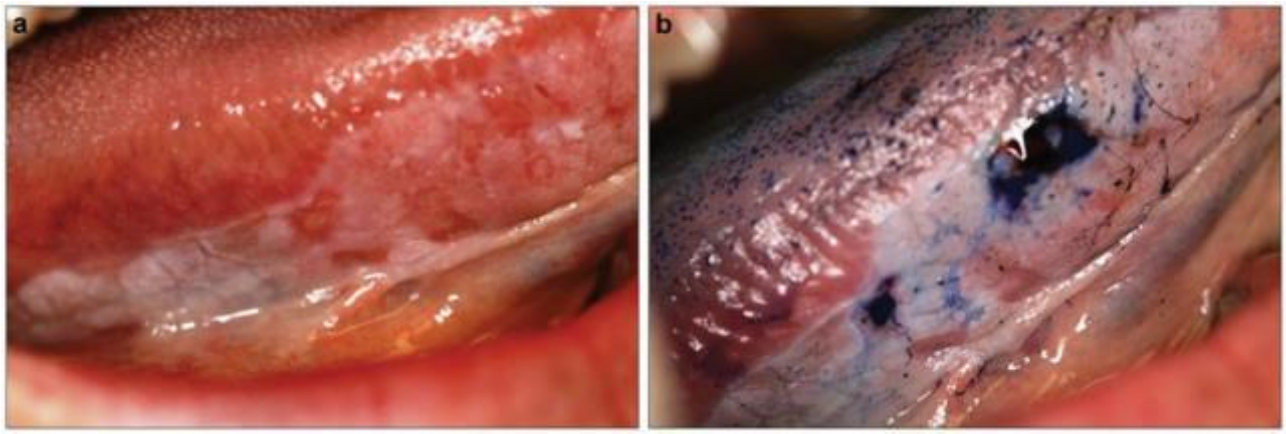

Figure 10. *TB staining of a suspicious oral lesion.*

Clinical photos demonstrating a) Erythroplakia on right lateral border of tongue; (b) Same lesion following application of TB stain which is retained (dark blue colour) in localised areas. Biopsy of these stained regions confirmed as a well differentiated squamous cell carcinoma. TB (toluidine blue).

Original source: Messadi DV (2013)[77]. Reproduced with permission of the rights holder under CC BY-NC-ND 3.0 DEED license.

2.3.2 Exfoliative cytology

Cytopathology is the microscopic study of cells collected from the surface of mucosal tissues, traditionally via smears, scrapings, lavage or from internal sites via fine-needle aspiration. The oral brush biopsy (OralCDx Brush Test) uses the concept of cytopathology or 'exfoliative cytology' to analyse a transepithelial sample of cells from suspected oral lesions. For analysis, the extracted sample of cells is fixed onto a glass slide and stained with a modified Papanicolaou test and analysed using a computer-based imaging system[77]. Results are reported as "negative or benign", "positive or atypical". Reported results are variable across studies with inconsistencies in specificities and PPVs of OralCDx. A systematic review reports varied sensitivities of 71% to 100%, and specificities of 27% to 94%. The reported PPVs ranged from 38-88% and NPVs from 60-100%[78]. Findings also indicated that OralCDx was better at detection of 'low-risk' lesions and in populations where oral malignancy is less common, in comparison to 'high-risk' lesions which was associated with higher false-positive rates.

Whilst the OralCDx has the advantage of being a minimally invasive chairside intervention and may be effective as an adjunct to visual oral screening, it does not replace the need for a surgical biopsy for a definitive histological diagnosis. Another limitation of this method is that exfoliative cytology samples can only be taken from oral lesions already visible to the naked eye, and therefore long-standing or submucosal lesions less visible would still require a standard biopsy[79,80]. As such, this method does not have a significant place in clinical practice.

2.3.3 Salivary diagnostics

Saliva contains constituents that can reflect disease and the physiological state of the human body and is therefore considered a useful medium for extracting cancer biomarkers[77]. Saliva has already demonstrated good promise with respect to detection of patients at risk of dental caries, periodontal disease, salivary gland disorders and other systemic conditions such HPV and Hepatitis C infections[81]. It has several useful properties in that it is an inexpensive, non-invasive, readily accessible, and a quick approach. More than 100 salivary biomarkers have already been implicated in OSCC development[82], however, none of these have yet been validated in large-scale prospective studies. The lack of strong evidence for many of these disease markers and their relatively small concentrations in saliva compared to serum, means the diagnostic value of saliva in oral cancer detection is not yet fully established.

2.3.4 DNA Ploidy

Ploidy refers to the number of complete sets of chromosomes within a cell, and DNA ploidy is the measure of a cell's nuclear DNA content that indicates the extent of genetic damage and chromosomal instability. During mitosis, if the chromosomes are not uniformly distributed to the daughter cells, some parts of chromosomes become detached, and chromosomal segregation becomes unbalanced, resulting in aneuploidy. Chromosomal instability and DNA aneuploidy are fundamental to the development of malignancy[83], and the presence of aneuploidy in precancerous oral lesions increases the risk to transformation[84]. These changes can be measured using DNA flow cytometry, image analysis and molecular analysis of loss of heterozygosity[85].

A study evaluating image-based DNA ploidy analysis in predicting malignant transformation risk of OED, demonstrated that combining DNA ploidy status with dysplasia grading gave a higher predictive value than either of these techniques alone[85]. The PPV for malignancy risk based on DNA aneuploidy status was 38.5% (sensitivity 65.2% and specificity 75%) in comparison 39.5% (sensitivity 30% and specificity 98%) for severe OED prediction. DNA diploid and tetraploid status had negative predictive values of 90% to 96%[85]. In another study, a significant correlation was found between DNA index and mitoses, cellular response and the degree of oral cancer differentiation, but not the ploidy status[86]. The authors concluded that DNA analysis may be a useful tool to support pathological analysis of HNC by reducing subjectivity, but ploidy status alone was not associated with tumour differentiation.

Whilst DNA ploidy analysis may have potential, research is ongoing in this field, and more work is needed to understand the molecular characteristics of OSCC to help develop more effective and reliable diagnostic tools. Clinical examination, imaging, and surgical biopsy, remain important measures for the diagnosis and staging of oral cancers.

2.3.5 Optical imaging

2.3.5.1 *Chemiluminescence: ViziLite*

ViziLite is a hand-held, single-use chemiluminescent light stick which emits light at varying wavelengths, intended to aid visualisation of healthy tissue from leukoplakia. Typically, healthy epithelial tissue will absorb more light and appear darker in comparison to hyperkeratinised or dysplastic lesions which will penetrate less light and appear whiter. The differences in light absorption, and therefore visual colour, is related to differing thickness of the lesions and may also be linked to altered nuclear structure and content that preferentially reflects the low energy blue-white light generating an “acetowhite” change. ViziLite Plus with TBlue systems have been approved for use by the Food and Drug Administration (FDA) as an adjunct to visual oral screening but the evidence to support its use is lacking and demonstrates conflicting results[77]. Several case series/convenience sample studies have explored the role of ViziLite in high-risk populations only, in cases where the oral lesions are already present. The reported sensitivity rate in these studies was 100% but due to the high level of sample bias, these figures may have been over-estimated and cannot be extrapolated to the general population. The other accuracy values were also inconsistent: 0-14% specificity, PPVs of 18-80% and NPVs of 0-100%[78].

2.3.5.2 *VELscope*

The VELscope system is a hand-held scope which allows screening of tissues visually to assess for changes in tissue fluorescence. The concept behind this system is that mucosal tissues have a reflective and absorptive pattern reflecting the naturally occurring fluorophores in the tissue. This method has already been well documented for the early detection of precancerous lesions in the lung, uterus, cervix and skin[87,88,89]. However, there are several factors that can affect variability in tissue fluorescence in the oral cavity, including alterations in epithelial tissue structure and composition, metabolism, vessel dilatation and inflammation[90,91]. The loss of autofluorescence in dysplastic and cancerous tissue is thought to be related to alterations to intrinsic tissue fluorophore distribution due to tissue remodelling, as well as alterations to metabolism.

The evidence for use of VELscope in patients in OPMD thus far has been to support detection of lesion margins, and therefore has a purpose for surgical planning. Aside from this, there is very little by means of research to support the usefulness of this system as a diagnostic adjunct in the general population, and particularly in lower-risk patient groups[78]. Given the variability of tissue fluorescence of the oral mucosa and the risk of false positives with benign lesions that may show loss of fluorescence, further research is still needed in this area.

2.3.6 Electrical impedance spectroscopy

Electrical impedance spectroscopy (EIS) is a technique based on the measurement of the frequency-dependent electrical response (or impedance) when a small alternating current (AC) is passed through a material. If the frequency of the AC signal is varied, it is possible to detect the response of multiple components within a material, depending on their capacitance and resistance[92]. All biological tissues have some degree of electrical impedance based on the tissue's components (cells, matrix, etc.) that have both resistive and capacitive (charge storage) properties. Since the size of the impedance and the dependence of impedance on frequency are related to tissue composition, different tissue structures are associated with different frequencies within an impedance spectrum.

EIS has been developed as a point-of-care screening tool for the detection and assessment of Barrett's oesophagus[93], cervical intraepithelial neoplasia[94], cutaneous malignancies[95,96], prostatic changes[97] and bladder pathology[98] but there is a paucity of high-quality evidence supporting its role as an adjunct for assessment of OPMD. In one study, EIS data was taken from a range of lesions including OSCC, OED, benign pathology (n=47) and compared with control sites (n=51). Findings demonstrated significant differences in the EIS of OSCC and high-risk OPMD versus low-risk PML and controls. There were no significant differences observed between benign lesions and normal controls[99]. EIS has also shown reduced potential in differentiating between individual OED grades, and between OED and OSCC.

Whilst there are some potential advantages of EIS, in that it is a non-invasive technique and provides real-time data allowing for immediate analysis during diagnostic procedures, it is associated with several practical limitations, such as equipment cost, the need for user training and results interpretation. These factors make EIS a less suitable point-of-care tool, and further research is needed to support its application and effectiveness in clinical practice.

2.3.7 Digital pathology and artificial intelligence

Over the last decade there have been significant technological advancements in computational software, artificial intelligence (AI) and digital image analysis. AI is a branch of computer science which relates to the use of smart machines that simulate processes which would typically require human intelligence, for example language processing, voice recognition, visual perception. The application of AI in everyday life is rapidly increasing, and more recently, the use of AI in precision medicine has seen a surge of interest[100-101].

There is a growing amount of evidence supporting the role of AI and digital analyses in removing subjective variability in cancer classification by ensuring standardisation, in addition to providing quantifiable outputs for cancer risk prediction and prognosis. The increasing ubiquity of patient data, whole-slide digital scanners and evolution of computational power has provided opportunities to uncover novel data from whole-slide images (WSI) through automated detection, pattern recognition

and quantitative analysis[100,102]. AI can obtain 'big data' from digital WSIs (Figure 11) and various image analysis platforms have been developed which allow for automated cell nuclei detection and extensive feature evaluation for more objective histological and morphometrical feature analysis.

Many studies have highlighted improved accuracy and efficiency of AI models in predicting diagnosis, treatment response, recurrence and cancer survival in a range of malignancies[103,104,105,106,107108,109]. CAMELYON (Cancer Metastases in Lymph Nodes) is a landmark study in which AI algorithms were shown to perform better than 11 experienced pathologists in detection of metastatic breast cancer in lymph node WSIs[110]. In another study, human epidermal growth factor receptor 2 (HER2) expression in breast cancer was more accurately scored by AI algorithms compared to the expert analysis by several pathologists[111]. Several pioneering AI-based cancer diagnostic algorithms have also been clinically deployed including the prostate cancer grade assessment (PANDA) tool[112] and an Image Biomarker Explorer (IBEX) software platform for detection and grading of breast malignancies[113]. These algorithms demonstrate the promise of such methods, and their potential to achieve better than pathologist-level accuracies. A more detailed review of the existing literature for application of AI in detection and diagnosis of precancerous and cancerous head and neck lesions is presented in Chapter 3.

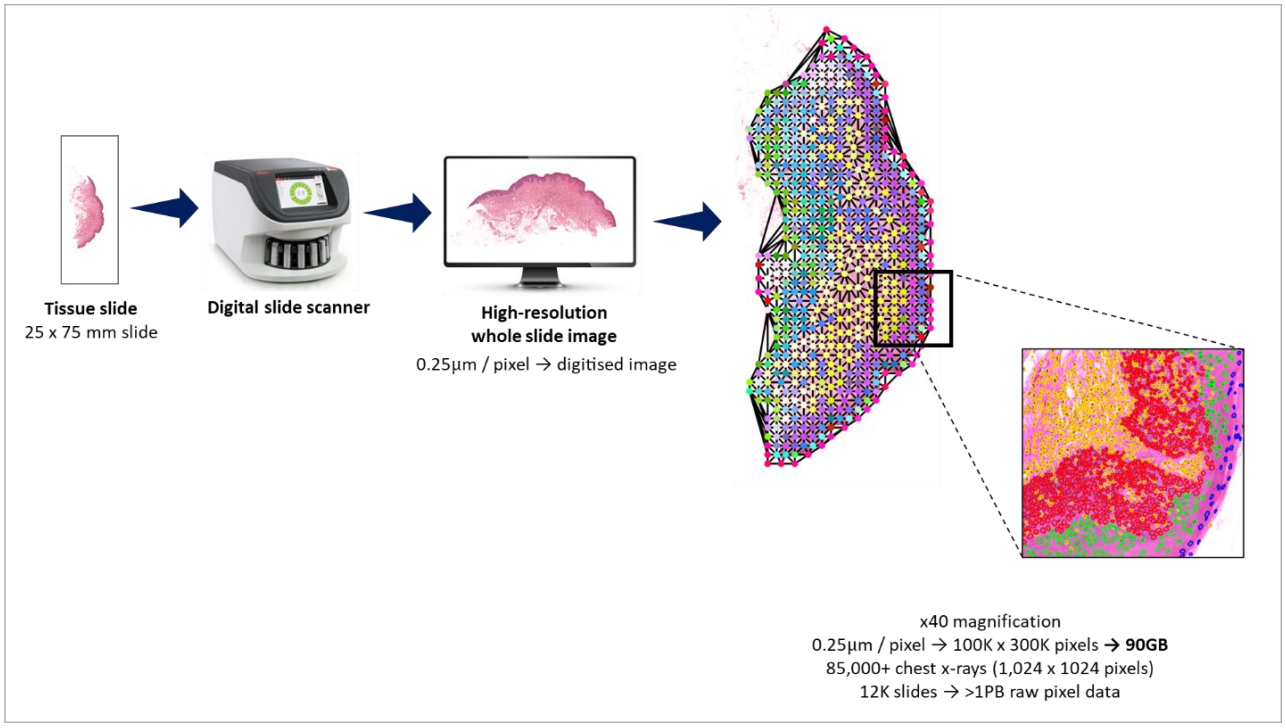

Figure 11. *Conversion of tissue section to digital whole slide image.*

Each whole slide image provides multi-gigapixel-level information for data leverage and algorithm training. A single 2.5x7.5cm slide converted at resolution of 0.25µm/pixel at x40 can generate 100,000 x 300,000 pixels which equates to ~90GB of raw data. A single WSI yields approximately 85,000+ pixels compared to conventional chest x-ray.

2.3.7.1 *Machine learning*

Machine learning (ML) is a branch of AI in which computational algorithms 'learn' information and patterns directly from data. There are various types of ML algorithms including *supervised, unsupervised* and *semi-supervised* (Figure 12) [114]. *Supervised* ML involves the training of models where the data input and output are already known, whereas *unsupervised* ML involves mining and extraction of hidden patterns from input data without any pre-defined information or linked examples[115]. Semi-supervised learning combines both labelled and unlabelled examples to enable the machine to generate an output. As more training data is provided, ML algorithms adaptively improve their performance through experiential learning.

Classification is an example of a traditional ML approach, which involves organising objects or observations based on matched attributes and characteristics and using this learning to predict a class label[8]. This technique is commonly used in computer vision tasks and has become increasing popular in cancer research, due to its ability to compute reliable quantifiable outputs for downstream statistical analyses from large scale imaging data[114]. Examples of commonly used classification algorithms include:

i. Linear Classifiers: Logistic Regression, Naive Bayes Classifier
ii. Nearest Neighbour
iii. Support Vector Machines
iv. Decision Trees
v. Random Forest
vi. Neural Networks

The application of these methods is the diagnosis of head and neck precancerous and cancerous lesions is presented in Chapter 3. Some of these approaches will also be discussed in further detail in Chapters 7 and 8.

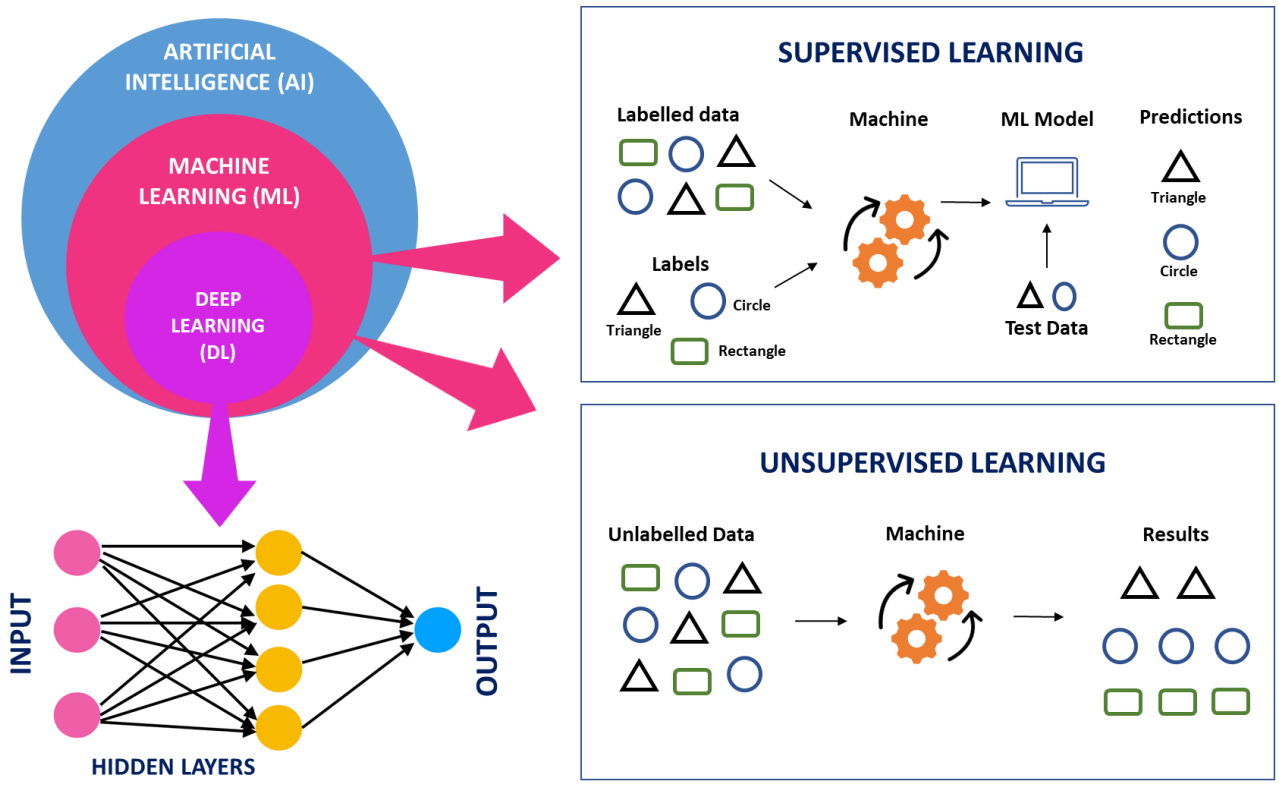

Figure 12. *Branches of Artificial Intelligence.*

Machine learning (ML) is a branch of AI and Deep learning (DL) is a further subfield of ML. ML models can be supervised or unsupervised.

2.3.7.2 *Deep Learning & Semantic Image Segmentation*

Deep learning (DL) is a further subfield of ML (Figure 12) in which algorithms learn high-level abstractions in data by utilising hierarchical architectures[116]. It works by using highly structured set of algorithms which model the brain's neural network system to form input layers, output layers and multiple layers in between[117]. Artificial neural networks (ANNs) and convolutional neural networks (CNNs) are well-established models used in a diverse range of computer vision applications[116].

Image classification, semantic segmentation, object detection/localisation and instance segmentation are all examples of DL tasks used in the field of computer vision and digital image analysis. Whilst these terms are used interchangeably, there are some distinct differences in these techniques.

Image classification has been described in 2.3.7.1 and essentially involves ascribing labels to image features based on the organisation of characteristics/objects. *Semantic segmentation* aims to classify individual pixels of an image by grouping together objects (or regions) which have similar characteristics[115]. To achieve this, WSI are split into regions of interest (ROI) and tessellated into

smaller sub-images ('patches') by a process known as 'patch extraction'. More precisely, the end point of semantic image segmentation is to label each pixel of the image into a predefined set of classes[118]. Some of the well-known algorithms that will be explored and discussed in Chapter 8 include CNN-based models (U-Net, DeepLab) and Transformers.

Object detection does not operate at the pixel level, but rather classifies objects in an image by localising it and enclosing all detected objects in a bounding box and giving a label to it. *Instance Segmentation* aims to combines *object detection* and *semantic segmentation* algorithm, which can be relatively difficult[118]. The method aims to classify the pixels of interest in the image and box-select the location of each object simultaneously. It should be noted that the same category will also be divided into different objects. The differences between *semantic segmentation* and the *instance segmentation* algorithms are illustrated in Figure 13.

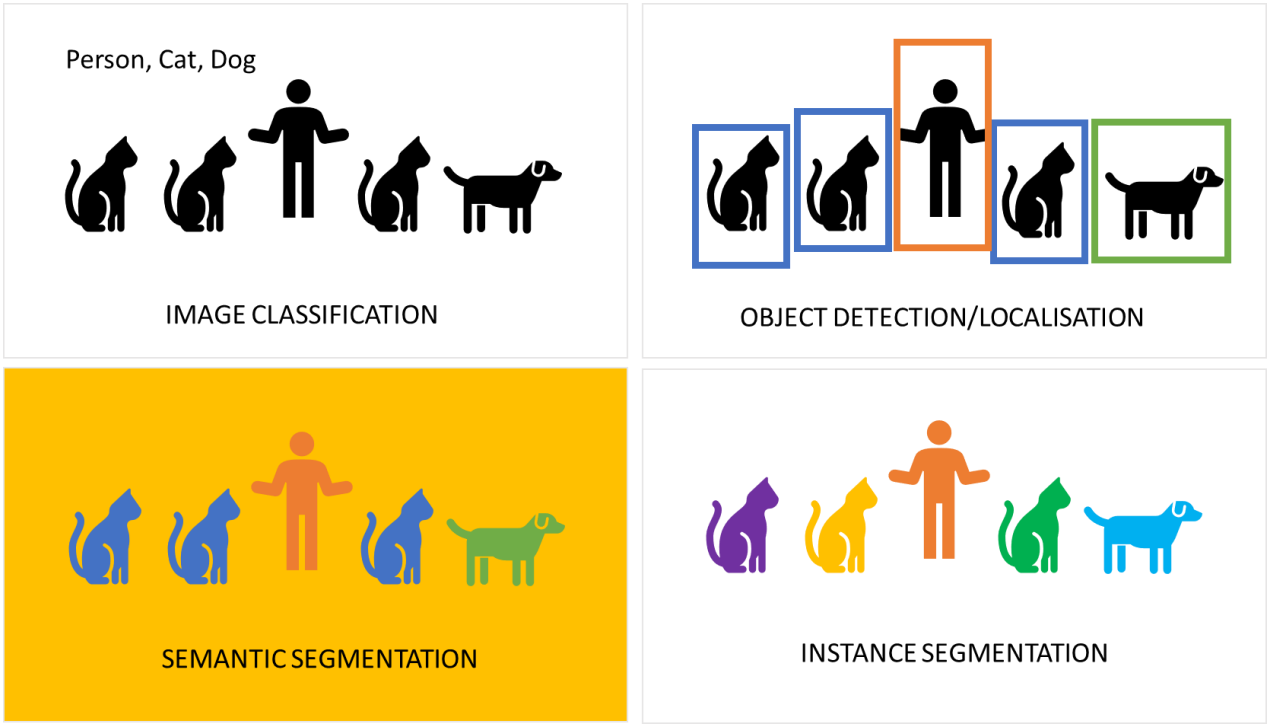

Figure 13. *Different image analysis approaches.*

Image classification, semantic segmentation, object detection/localisation and instance segmentation are all examples of DL tasks used in the field of computer vision and digital image analysis. These methods involve ascribing labels to image features based on the organisation of characteristics/objects.

Chapter 3 – Literature Review

Chapter 3 – Literature Review

This chapter has been reproduced from three peer-reviewed publications for which the candidate is first and primary contributing author for the first two articles (3.1 and 3.2) and shares joint first authorship for the third presented article (3.3).

The first paper provides an overview of the published literature (between 2009–2020) relating to the application of AI/ML methods to aid diagnostic evaluation of head and neck cancers (HNC). An outline of the different anatomical sites for HNC lesions, the type of diagnostic imaging modality and the AI/ML method used is reviewed.

The second paper is a focussed systematic review conducted with Blanca Iciar Indave Ruiz, a Systematic Reviewer for WHO/IARC Classification of Tumours Group. This review follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for review of the scientific literature (between 2009-2020) on the application and diagnostic accuracy of AI/ML methods for detection and grading of potentially malignant and cancerous head and neck lesions.

The third paper provides a comprehensive discussion of OED histological features and an overview of the common mimics. It highlights the paucity of evidence defining these features while offering suggested definitions. The whole slide image examples presented in this article has been shared as a cloud-based open access dataset, available at: https://www.pathogenesis.co.uk/r/demystifying-dysplasia-histology-dataset.

The candidate's contributions for both articles include:

i. performing the electronic literature search
ii. performing the screening and selection of eligible articles for inclusion in the reviews
iii. performing risk of bias and quality assessment for all the included articles
iv. data extraction and synthesis
v. writing up the manuscripts

For the first two articles (3.1 and 3.2), Dr Muhammad Shaban assisted with article screening and risk of bias assessment, and the senior authors confirmed final article selection. The search strategy was developed in collaboration with a medical information specialist from the Health Sciences Library, University of Sheffield.

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

BJC British Journal of Cancer

www.nature.com/bjc

ARTICLE

AI Applied to Cancer

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

Hanya Mahmood[1], Muhammad Shaban[2], Nasir Rajpoot[2] and Syed A. Khurram[3]

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 annually[11] 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

[1]Academic Unit of Oral & Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield, Sheffield, UK; [2]Department of Computer Science, University of Warwick, Coventry, UK and [3]Unit 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

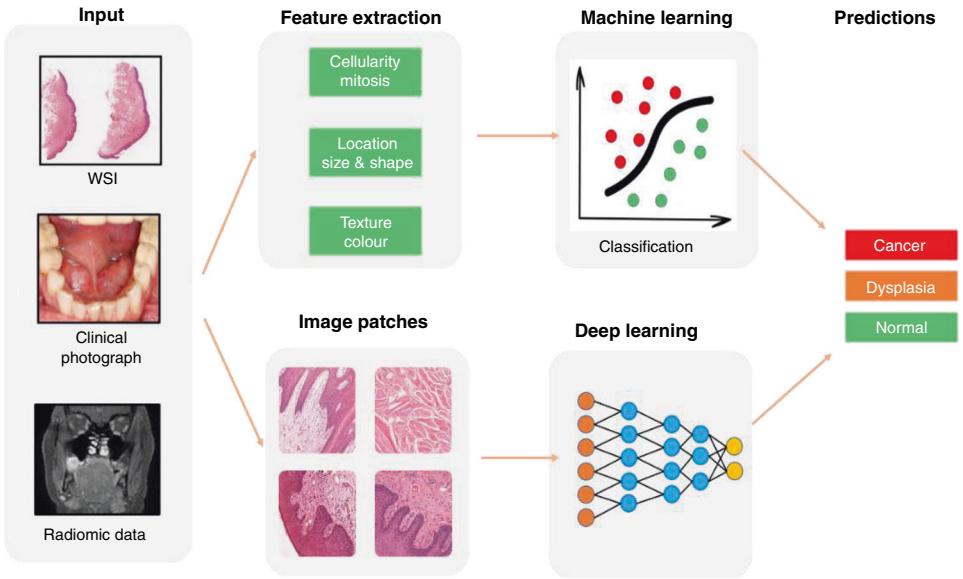

Fig. 1 Feature extraction from primary diagnostic imaging modalities to train ML/DL algorithms to aid outcome prediction. Source: The schematic diagrams were prepared in line with the journal's artwork guidelines; the clinical/histological/radiological images were obtained with appropriate consent from 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 steps[28] (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

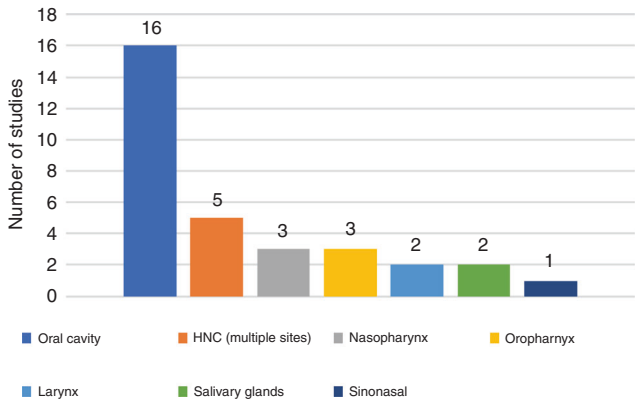

Fig. 2 HNC anatomical site distribution. A bar chart showing the proportion of included studies based on head and neck cancer anatomical subtype.

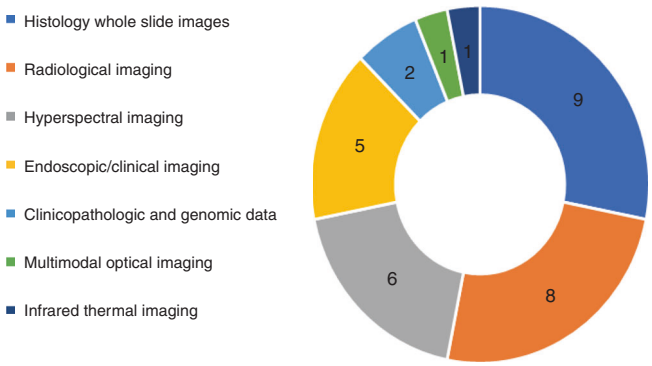

Fig. 3 Type of imaging modality/input data used. A pie chart showing the proportion of identified studies based on the type of maging modality/input data used to train AI/ML algorithms.

(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 nasopharyngeal[31] 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 ± 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] oropharyngeal[34] 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 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]

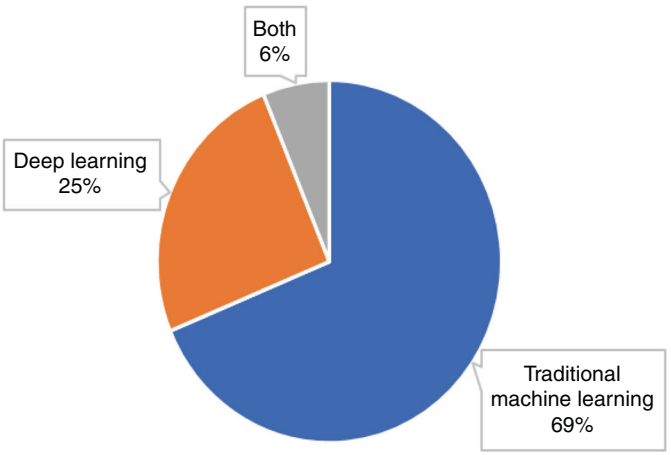

Fig. 4 Types of AI methods used. Proportion of studies using traditional ML methods, modern deep learning methods and a combination of both types of methods.

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.

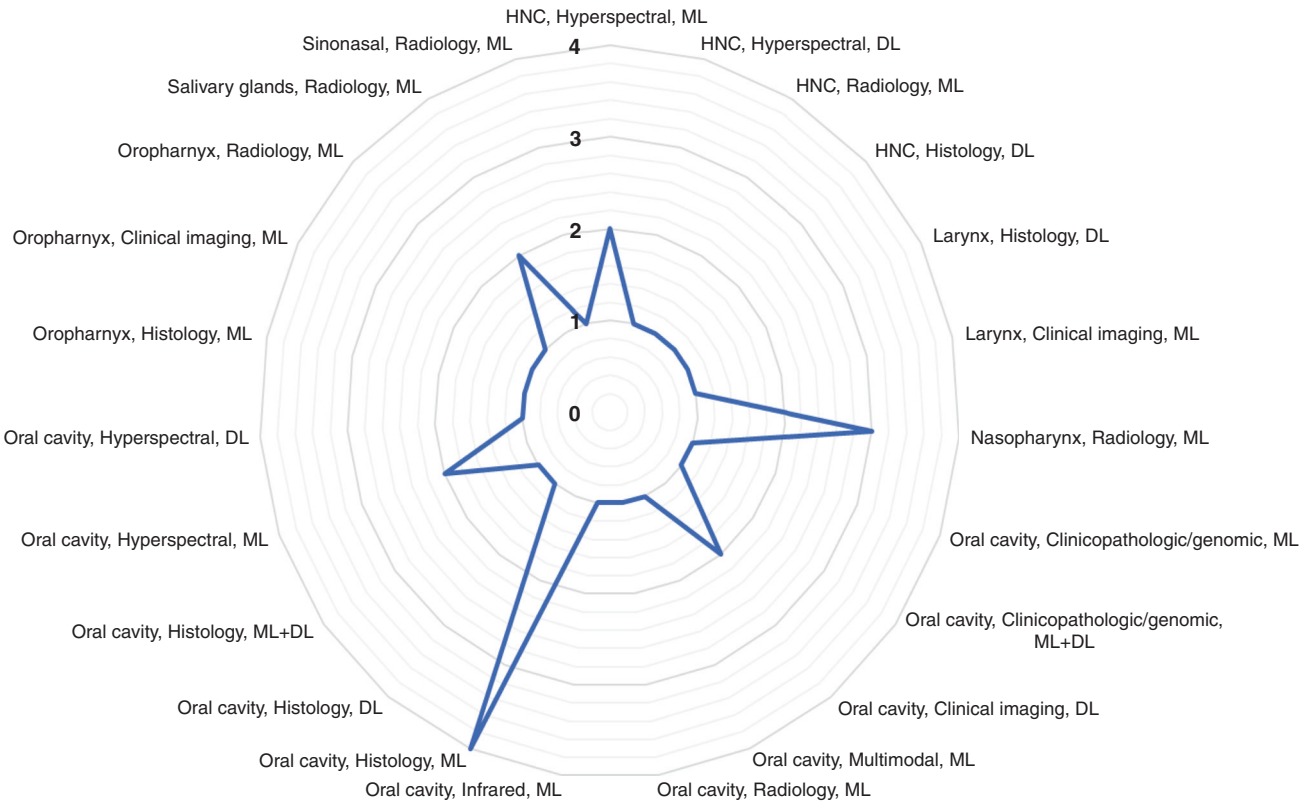

Fig. 5 Overview of recently published literature. A diagram illustrating an overview of the recently published literature based on anatomical site of lesion, imaging modality and AI methods used.

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.

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/healthprofessional/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. Mehlum, 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., Ciompi, 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 multienergy 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 laserendomicroscopy 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., Kolur, T. et al. Automatic classification of dual-modalilty, 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. Ægisdóttir, 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).

3.2 Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review

Oral Oncology 110 (2020) 104885

Contents lists available at ScienceDirect

Oral Oncology

journal homepage: www.elsevier.com/locate/oraloncology

Review

Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: A systematic review

H. Mahmood[a,*], M. Shaban[b], B.I. Indave[c], A.R. Santos-Silva[d], N. Rajpoot[e], S.A. Khurram[f]

[a] Dr Hanya Mahmood (NIHR Academic Clinical Fellow in Oral Surgery), Academic Unit of Oral & Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield, 19 Claremont Crescent, S10 2TA, UK
[b] Muhammad Shaban (Research Student), Department of Computer Science, University of Warwick, Coventry, UK
[c] Blanca Iciar Indave Ruiz (Systematic Reviewer), WHO/IARC Classification of Tumours Group, International Agency for Research on Cancer, Lyon, France
[d] Alan Roger Santos-Silva (Associate Professor in Oral Medicine & Pathology), Oral Diagnosis Department, Piracicaba Dental School, University of Campinas, Piracicaba, São Paulo, Brazil
[e] Nasir Rajpoot (Professor of Computational Pathology), Department of Computer Science, University of Warwick, Coventry, UK
[f] Syed Ali Khurram (Senior Clinical Lecturer & Honorary Consultant in Oral & Maxillofacial Pathology), Academic Unit of Oral & Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, 19 Claremont Crescent, UK

ARTICLE INFO

Keywords:
Artificial intelligence
Machine learning
Head and neck cancer
Oral cancer
Pre-cancer
Oral potentially malignant disorders, dysplasia, squamous cell carcinoma, deep learning, systematic review

ABSTRACT

This systematic review analyses and describes the application and diagnostic accuracy of Artificial Intelligence (AI) methods used for detection and grading of potentially malignant (pre-cancerous) and cancerous head and neck lesions using whole slide images (WSI) of human tissue slides. Electronic databases MEDLINE via OVID, Scopus and Web of Science were searched between October 2009 – April 2020. Tailored search-strings were developed using database-specific terms. Studies were selected using a strict inclusion criterion following PRISMA Guidelines. Risk of bias assessment was conducted using a tailored QUADAS-2 tool. Out of 315 records, 11 fulfilled the inclusion criteria. AI-based methods were employed for analysis of specific histological features for oral epithelial dysplasia (n = 1), oral submucous fibrosis (n = 5), oral squamous cell carcinoma (n = 4) and oropharyngeal squamous cell carcinoma (n = 1). A combination of heuristics, supervised and unsupervised learning methods were employed, including more than 10 different classification and segmentation techniques. Most studies used uni-centric datasets (range 40–270 images) comprising small sub-images within WSI with accuracy between 79 and 100%. This review provides early evidence to support the potential application of supervised machine learning methods as a diagnostic aid for some oral potentially malignant and malignant lesions; however, there is a paucity of evidence using AI for diagnosis of other head and neck pathologies. Overall, the quality of evidence is low, with most studies showing a high risk of bias which is likely to have overestimated accuracy rates. This review highlights the need for development of state-of-the-art deep learning techniques in future head and neck research.

Introduction

Head and neck cancers (HNC) encompass a large group of cancers, most commonly squamous cell carcinomas (SCC) (90%) of the oral cavity, nasal cavity, sinuses, salivary glands, pharynx and larynx. Primary risk factors include tobacco and betel nut use [1], alcohol consumption [2], radiation [3], immunodeficiency [4] and specific viruses including Human Papillomavirus (HPV) 16 and 18 (for oropharyngeal squamous cell carcinoma, OPSCC) [5] and Epstein-Barr virus (for nasopharyngeal squamous cell carcinoma, NPSCC) [6,7]. Chronic exposure to these carcinogenic factors and/or infection status can result in dysplastic changes in the oral, oropharyngeal, nasal or nasopharyngeal mucosa, which may lead to the development of HNC. The incidence of HNC continues to rise, making it the sixth leading group of cancers worldwide [8,9]. In 2018, HNC accounted for more than 650,000 new cases and 33,000 deaths annually worldwide [10]. In the UK, the number of new patientshas increased by 22% over the last decade, with almost 12,000 new diagnoses every year (33 every day) [11].

Despite advancements in medical and surgical techniques, prognosis of HNC remains poor with a five-year survival rate between 28 and 67% [11]. Due to late presentation, even successful treatment of HNC is

* Corresponding author.
E-mail addresses: h.mahmood@sheffield.ac.uk (H. Mahmood), m.shaban@warwick.ac.uk (M. Shaban), indavei@iarc.fr (B.I. Indave), alan@unicamp.br (A.R. Santos-Silva), N.M.Rajpoot@warwick.ac.uk (N. Rajpoot), s.a.khurram@sheffield.ac.uk (S.A. Khurram).

https://doi.org/10.1016/j.oraloncology.2020.104885
Received 13 April 2020; Received in revised form 28 June 2020; Accepted 29 June 2020

associated with multiple functional problems including masticatory, speech and swallowing impairments which can significantly reduce the quality of life [12]. Early diagnosis of potentially malignant head and neck lesions can prevent cancer development in up to 88% of cases [11], however most patients are diagnosed at a late stage of disease (62% diagnosed at stage III or IV) [13]. The conventional diagnosis of suspicious head and neck lesions involves clinical, radiological and histopathological assessment. The latter is the gold standard providing important prognostic information (i.e. grade for dysplasia and cancers) which can guide clinical treatment decisions [14,15]. However, histological interpretation can be subjective with differences in interpretation [16], variation in consistency [17] and may not provide effective risk stratification or management guidance. This highlights the importance of novel methods and technologies for more consistent, efficient and accurate diagnosis to aid clinical decision-making and to improve HNC related patient survival.

Over the past decade, Artificial Intelligence (AI) has gained popularity in cancer research where it has been shown to increase diagnostic accuracy and efficiency by providing quantifiable outputs to predict cancer behaviour and prognosis [18–20]. Machine learning (ML), a branch of AI, has been shown to reduce variability in grading of dysplasia and cancers by ensuring standardisation and consistency in addition to informing treatment decisions [21]. ML uses computational methods to 'learn' information and patterns directly from data. This learning can be *supervised (*involving training of ML models on a known data input and output i.e. histology slides with associated diagnostic annotations) or *unsupervised (*which involves mining and extraction of hidden patterns from input data without any pre-defined information). ML algorithms adaptively improve their performance with an increasing number of 'learning or training' samples, enabling the computer to essentially 'learn from experience'. Classical supervised ML approaches include *semantic segmentation* and *classification. Segmentation* involves dividing high-resolution digital whole slide images (WSI) of human tissue into regions of clinical relevance followed by deconstruction of the WSI into smaller patches (sub-images) by a process known as 'patch extraction'. This enables ML algorithms to compute local and global features which can be explored for significance during the 'learning or training' phase. *Classification* involves organising and classifying new observations based on specific attributes (e.g. morphology of nuclei) learnt from previous data input. Both techniques are commonly used approaches in cancer research, providing useful diagnostic and prognostic outputs. Other relevant computational pathology terms have been described in Table 1.

With the evolution of computational power and image analysis algorithms, there is now an increasing amount of evidence demonstrating the success of AI-based image analysis from WSI of human tissue slides [22]. Several studies have demonstrated the potential for AI-based methods to reliably predict diagnosis, prognosis, mutational status, and response to treatment in a range of cancers including colorectal, lung, skin and breast malignancies [23–27]. These studies highlight the potential for AI-based methods in provision of faster, consistent, accurate and reproducible information regarding cancer diagnosis and prognosis which can complement the conventional (and largely subjective) light microscopy analysis by experienced pathologists.

The alarming rise in global HNC incidence and its poor prognosis makes it ideally suited for application of AI-based methods to aid objective diagnosis and provide valuable prognostic information. We performed a systematic review of literature published in the last ten years, to assess the application and diagnostic accuracy of AI/ML methods for detection and grading of potentially malignant and cancerous head and neck lesions. To the best of the authors' knowledge, this is the first study reviewing the use and application of AI-based methods for head and neck lesions.

Materials and methods

The systematic review was conducted using a predetermined protocol which followed the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement and checklist28. The protocol was registered in the International Prospective Register of Systematic Review (PROSPERO) database, CRD42019153023.

Outcome definitions

The primary outcomes were the specific histopathological features used for diagnosis and grading of the head and neck lesion under study, in addition to the methods and performance of the proposed AI/ML techniques. Descriptive analysis was conducted on these outcomes and the reported diagnostic performance measures (i.e. sensitivity, specificity, accuracy, F1-score) where possible.

Literature search

Electronic databases search of MEDLINE via OVID, Scopus and Web of Science was conducted to retrieve articles published in the English language between October 2009 and April 2020. The Cochrane library was also consulted. This time period was chosen due to the rapid evolution of AI methods and their application to cancer diagnostics over the last decade.

The search strategy was jointly developed by the multidisciplinary authorship team in collaboration with a medical information specialist (librarian from University of Sheffield, UK). Input from an expert Oral and Maxillofacial Pathologist (SAK, University of Sheffield, UK) and a

Table 1
Glossary of relevant computational pathology terms.

Term	Description
Artificial Intelligence (AI)	A branch of computer science concerned with building smart machines that can perform tasks which typically require human intelligence.
Machine Learning (ML)	The ability for machines to 'learn' information and patterns directly from data without being programmed explicitly.
Supervised Learning	Training of ML algorithms from labelled (e.g. annotations) input and output data.
Unsupervised Learning	Training of ML algorithms by mining and extracting hidden patterns from input data that has not been labelled.
Deep Learning (DL)	DL is a subfield of ML in which algorithms learn from input data through example (without supervision).
Neural Network	A highly structured set of algorithms which models the brain's neural network system (deep learning) designed to recognise patterns from input data.
Whole slide image (WSI)	A high-resolution microscopy image of human tissue section.
F1 Score	A statistical analysis of binary classification to measure the accuracy of a test, considering the weighted average of the precision (p) and recall (r) of the test to compute the overall score.
Precision (p)	The number of correct positive results divided by the number of all positive results returned by the classifier
Recall (r)	The number of correct positive results divided by the number of all relevant samples (all samples that should have been identified as positive
Classification	A ML technique which categorises a set of data (structured or unstructured) into classes based on certain attributes.
Patch extraction	Deconstruction of a WSI into smaller pixelated patches known as 'sub-images'.
Semantic segmentation	The process of dividing WSI's into regions of interest and clustering data into distinct groups based on similarities.

Professor of Computational Pathology (NMR, University of Warwick, UK) ensured adequate selection of clinical and technical terms and controlled vocabulary for optimal identification of articles.

Tailored search strings containing keywords and database-specific medical subject headings (MeSH) for the two major topics (AI/ML and potentially malignant or cancerous head and neck disorders/lesions) were developed. Multiple variations of search terms were combined to produce different sets of results and the final search strategy was pilot-tested and modified accordingly (Appendix 1). Grey literature and the reference lists of selected articles were also screened for relevant studies that may not have been identified through the database searches. The electronic databases search was conducted with assistance from an experienced Librarian at the University of Sheffield, UK. Article citations were exported to EndNote® reference manager software (Clarivate Analytics, Philadelphia, USA) and duplicate articles were removed.

Study selection

Two independent reviewers (HM, MS) retrieved the literature, and screened titles and abstracts. Where insufficient information was available to determine eligibility, the full report was obtained for further assessment. Articles that did not meet the eligibility criteria were excluded. In the second stage of study selection, the same two reviewers (HM, MS) independently assessed the full-text reports to obtain a shortlist of relevant articles. The shortlists were compared, and differences discussed, obtaining a final selection of studies. In case of any discrepancies in article selection, a discussion with senior members of the review team (NMR, SAK) took place to reach a mutual final decision. For relevant articles with overlapping datasets or results, the most recent publication was included.

The following criteria were applied for the selection of eligible studies for this review:

Inclusion criteria:

- Studies using AI for automated detection, grading and classification of potentially malignant and cancerous head and neck disorders/lesions.
- Studies exploring diagnostic accuracy of the applied AI/ML method providing sensitivity, specificity, accuracy, F1-scores as outcome measures.
- Studies published in indexed journals between October 2009 - April 2020.

Exclusion criteria:

- Studies not using WSI of human tissue slides.
- Studies not using histological image modalities (e.g. radiographic, photographic, cytology, genomic data etc.).
- Studies using AI/ML to predict disease progression, prognosis, metastasis, recurrence, survival or treatment efficacy (i.e. those not primarily investigating detection, grading and classification of head and neck lesions).
- Studies using AI/ML for detection and diagnosis of thyroid or oesophageal cancer.
- Narrative reviews, letters to editors, commentaries and conference abstracts.
- Studies not available in the English language.

Data extraction

Relevant data from selected articles was extracted, processed and tabulated into a pre-developed data collection form in Microsoft Excel® (Microsoft Corporation, Washington, USA) by two reviewers (HM, MS). The following information was recorded:

- Study details (authors, year and country of publication, aims)
- Study methods (design, dataset size and selection)
- Description of outcome variables (AI/ML methods used, head and neck lesion and histological parameter under study, training and validation sample details) and its outcome measures (reported diagnostic accuracy, effect measures)
- Other relevant details (funding information, sources of support, conflict of interest disclosure)

Methodological quality and risk of bias assessment

The methodological quality of individual studies and risk of bias was assessed using the Quality Assessment of Studies of Diagnostic Accuracy – Revised QUADAS-2 tool [29]. This tool is designed specifically for use in systematic reviews to evaluate the risk of bias and applicability of the primary diagnostic accuracy studies. The tool was adapted with input from an Oral and Maxillofacial Pathologist (SAK) and a Professor of Computational Pathology (NMR) to ensure relevant signalling questions were included to reliably and fairly assess the quality of included studies in relation to: 1) sample selection 2) index test and 3) reference standard. The tailored QUADAS-2 tool was piloted on five studies by two independent reviewers (HM, MS) and differences were resolved with consensus. The overall score for each study was determined by combining the number of satisfied criteria, with a higher score representing higher methodological quality. The outcome of the methodological quality assessment is presented graphically in Table 3 and the influence of bias risk on our results was discussed where applicable.

Data synthesis

A narrative synthesis of the main study findings is presented. Due to the large variation in outcome definitions and heterogeneity of retrieved data, a *meta*-analysis for calculation of adjusted pool estimates was not carried out.

Results

Search results

The electronic database search retrieved a total of 314 articles (MEDLINE via OVID: 154, Scopus: 81 and Web of Science: 79). In addition, one article was identified through citation searching. Following removal of duplicate studies, 288 articles were selected. After the first screen based on title and abstract, 259 articles did not satisfy the inclusion criteria and were therefore excluded. A comprehensive full-text examination of the remaining [29] articles excluded a further 18; resulting in 11 eligible articles for inclusion in this review paper (Figure 1).

Out of 315 articles, 277 were excluded, with a large proportion not eligible due to the use of imaging modalities other than histology (n = 157). Many studies did not address the research question directly (n = 67) as they focussed on the application of AI algorithms to predict disease prognosis, recurrence, metastasis or treatment success. Other reasons for exclusion included studies which did not use AI based methods (n = 52) or human tissue (n = 1).

Description of studies

Table 2 summarises the main findings for the included studies. In six studies, AI-based methods were used to detect oral potentially malignant disorders (OPMD) [52,54,55,57,58,62] with five of these focussing on the detection of oral submucous fibrosis (OSF) specifically. Four studies aimed to detect oral squamous cell carcinoma (OSCC) [53,56,59,60] and one study aimed to classify OPSCC [61]. Overall, seven studies were conducted in India [53,54,55,57,58,59,62], two in China [60,61], one in USA [52] and one in Germany [56]. Eight studies

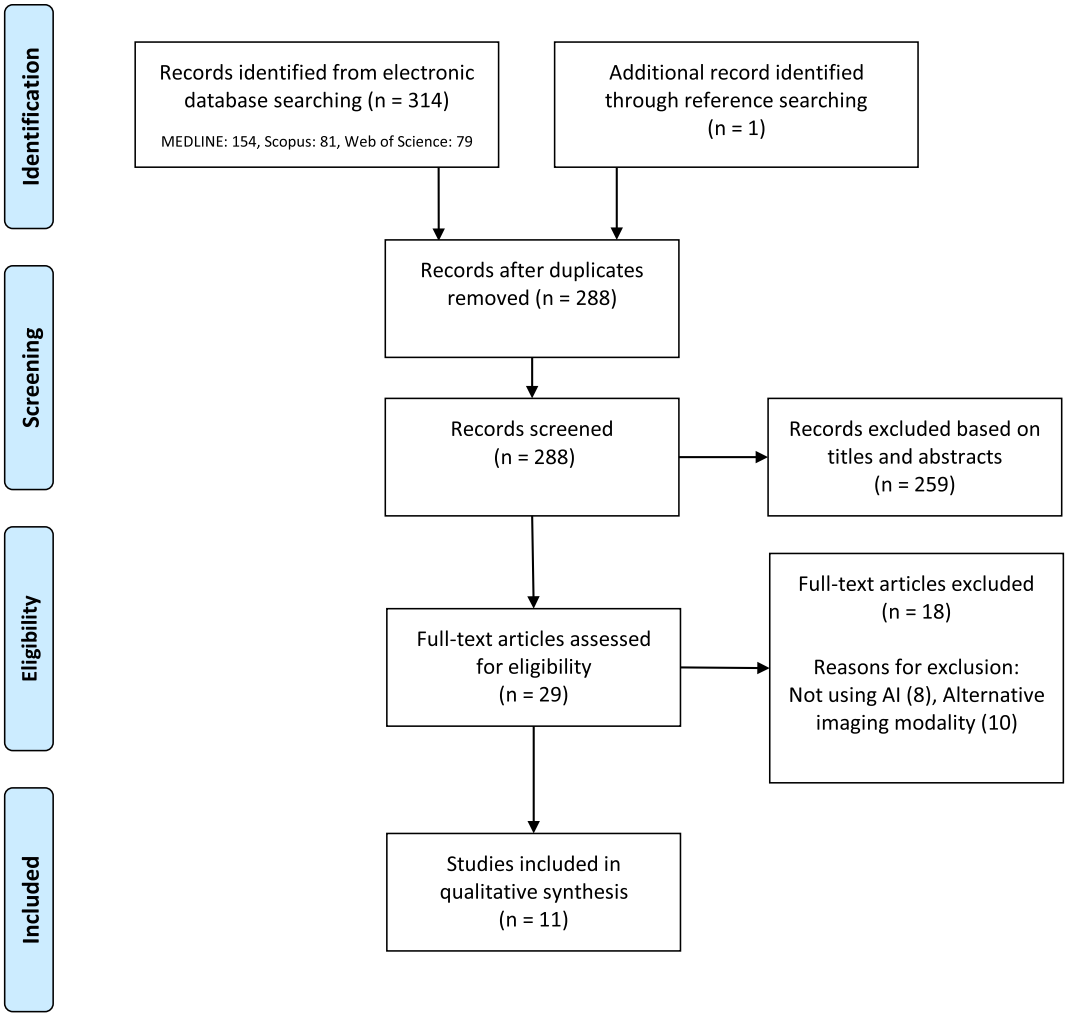

Fig. 1. PRIMSA flowchart (diagram adapted from PRISMA group, 2009. [28]) demonstrating the study selection process.

were published between 2009 and 2015 [52,54,55,57,58,60,61,62] and three were published after 2015 [53,56,59].

Results of the selected studies have been presented based on the type of head and neck lesion being analysed, which includes OPMD, OSCC, and OPSCC. Following this, the methods used in the selected studies will be presented, which will describe the type of AI/ML technique, dataset sample and the diagnostic performance for each.

Detection of OPMD

Baik *et al.* [52] quantified nuclear phenotypic changes in oral epithelial dysplasia (OED) lesions using an automated nuclear phenotypic score (a-NPS). The a-NPS was used to classify suspicious oral lesions based on the risk of progression to OSCC. The tissue samples used for algorithm training compared relatively normal oral mucosa (i.e. amalgam tattoo or melanotic macule, 34%) to OSCC from high risk intra-oral sites (floor of mouth and lateroventral tongue, 66%). Following training, the algorithm was tested on biopsies diagnosed as hyperplasia, mild or moderate dysplasia including almost an equal representation of transformed and untransformed lesions. The study used a robust experimental design to produce good accuracy (78% sensitivity and 71% specificity) highlighting the a-NPS as a potentially useful prognostic adjunct.

Five other studies focussed on detection of OSF [54,55,57,58,62] using a variety of supervised ML methods to differentiate between normal tissue, OSF with dysplasia or atrophy and OSF without dysplasia or atrophy (Table 2). Krishnan *et al.* [57] classified the number of sub-epithelial connective tissue (SECT) cells (excluding endothelial cells) in oral mucosa of normal and OSF tissue. Specific histological features including SECT cell shape, size and dimensions were evaluated with focus on round shaped cells (e.g. macrophages, lymphocytes, mast cells and neutrophils) and spindle shaped cells (fibroblasts, fibrocytes, histiocytes and endothelial). In addition, geometric properties such as the compactness and eccentricity of these cells were considered for classification. The results demonstrated a classification accuracy of 88.69%, although this was based on a small dataset. In another study, Krishnan *et al.* [58] used a texture-based method for segmentation of the constituent layers of the epithelium in OSF tissue (based on density and thickness of individual layers) to distinguish it from normal tissue. The standard performance measures were not reported in this study (Table 2).

Detection of OSCC

Das *et al.* [53] aimed to detect OSCC using a two-stage approach. This involved segmentation of constituent layers of the oral mucosa (into epithelial, subepithelial and keratin layers) followed by texture-based classification of keratin pearls from segmented keratin regions. The detection accuracy for keratin pearls was reported as 96.88% however this was based on a small dataset comprising small patches (sub-images) within WSI.

Rahman *et al.* [59] used a texture-based classifier to distinguish between normal and cancerous cells achieving an accuracy of 100% using small patches within WSI. In another study, Sun *et al.* [60]

Table 2
Summary of findings for selected studies.

Study	Study Reference	Study Aims	AI/ML Methods andImage Dataset	Reported Diagnostic Performance (%)	Risk ofBias
1	Baik *et al.*[52]	Classification of oral precancerous lesions into low and high-risk groups of progression into OSCC using a nuclear phenotypic score.	**Methods:**○ Random Forest**Datasets:**○ Training: 62○ Test: 71	○ Sensitivity: 78%○ Specificity: 71%	9
2	Das *et al.*[53]	Segmentation of OSCC histology images into epithelial, sub- epithelial, and keratin layers. Detection of keratin pearls from the segmented keratin layer.	**Methods:**○ Random Forest○ Neural Network**Datasets:**○ Training: 80○ Test: 20	○ Sensitivity 97.7%○ Dice-coefficient 95.3%	4
3	Krishnan *et al.* [54]	Classification of oral precancerous lesions into normal, oral sub-mucous fibrosis without dysplasia, and oral sub-mucous fibrosis with dysplasia.	**Methods:**○ Sugeno Fuzzy○ Decision Tree○ K-Nearest Neighbour○ Gaussian Mixture Model○ Neural Network**Datasets:**○ Training: 158○ Test: 0	○ Sensitivity: 94.5%○ Specificity 98.8%	7
4	Krishnan *et al.* [55]	Classification of oral precancerous lesions into normal and oral submucous fibrosis through segmentation of collagen fibres in the subepithelial connective tissue.	**Methods:**○ Bayesian Classifier○ Support Vector Machine○ Neural Network**Datasets:**○ Training: 89○ Test: 30	○ Accuracy: 91.70%	2
5	Lorsakul *et al.* [56]	Spatial quantification of brightfield-multiplex immunohistochemistry stained imaging for epithelial tumour cells and carcinoma-associated fibroblasts in tumour-associated stroma, through detection and classification of cells and segmentation of fibroblasts.	**Methods:**○ Random Forest○ L1-Logistic Regression○ Support Vector Machine**Datasets:**○ Training: 135○ Test: 35	○ Accuracy: 91.64%	2
6	Krishnan *et al.* [57]	Segmentation of sub-epithelial connective tissue cells followed by cell classification into normal and oral submucous fibrosis tissue.	**Methods:**○ Support Vector Machine**Datasets:**○ Training: 200○ Test: 20	○ Sensitivity: 90.46%○ Specificity: 87.54%○ Accuracy: 88.89%	3
7	Krishnan *et al.* [58]	Segmentation of the epithelial layer in normal and oral sub-mucous fibrosis histological images.	**Methods:**○ Watershed○ Otsu**Datasets:**○ Training: 158○ Test: 0	Standard performance measures not reported.	4
8	Rahman *et al.* [59]	Classification of oral histology images into normal and oral squamous cell carcinoma.	**Methods:**○ Support Vector Machine**Datasets:**○ Training: 216○ Test: 54	○ Accuracy: 100%	5
9	Sun *et al.*[60]	Segmentation of tumour in anti-CD34 antibody stained oral cancer histology images.	**Methods:**○ Clustering**Datasets:**○ Training: 8○ Test: 208	Standard performance measures not reported.	1
10	Fouad *et al.*[61]	Segmentation of oropharyngeal cancer tissue into epithelial and stromal regions.	**Methods:**○ Clustering**Datasets:**○ Training: 10○ Test: 45	F1-Score: 81%	5
11	Krishnan *et al.* [62]	Classification of oral premalignant lesions into normal, oral sub-mucous fibrosis with atrophy.	**Methods:**○ Otsu○ Linear Discriminant Analysis○ Neural Network**Datasets:**○ Training: 84○ Test: 28	○ Sensitivity: 92.31%○ Specificity: 100	3

developed an automated colour-based feature extraction system to segment and classify OSCC stained with anti-CD34 antibody. Specific histological features (vessel area/number/density and nuclei area/number) were computed to enable quantitative differentiation between different OSCC stages. Results demonstrated sensitivities of 49.11%, 64.17%, 58.55%, 79.60% for OSCC stages I-IV, respectively.

Detection of OPSCC

Fouad *et al.* [61] used unsupervised ML methods for automated identification of specific tissue compartments (cells and nuclei) in OPSCC tissue microarrays. Measurements of cell and nuclei colour and morphology were used for classification of epithelial and stromal tissue. This study compared their results with other standard segmentation methods and reported relatively low recognition accuracy (pixel-level F1 score of 80–81%) attributed to the lack of pre-defined manual annotations often used in supervised learning methods.

ML methods used in selected studies

ML algorithms can be divided into two groups: *classical* and *modern*. The classical methods require small amounts of training data and computational resources for pattern recognition in comparison to modern methods. However, modern methods often outperform classical methods in addressing most ML problems. Deep learning is a modern ML approach, in which algorithms mimic the brain's neural networks to learn without supervision however it can suffer from the 'black box' problem, unlike classical ML methods which are easier to interpret. A hierarchical classification of ML methods used in the selected studies is presented in Fig. 2.

In most of the selected studies, classical supervised ML approaches have been used although three classical unsupervised methods have also been employed, including Otsu and Watershed (for image segmentation into two or more classes) and Clustering (e.g. K-Mean and Agglomerative Hierarchical Clustering). The most frequently applied ML methods were from the classical supervised group, which included nine different techniques (Fig. 3). The majority of these supervised methods belong to the handcrafted feature-based classical ML group, although in four studies [53–55,62] modern ML methods (neural networks) were employed. These nine methods differ significantly in their learning strategies, as outlined below:

- *Sugeno* Fuzzy [30] involves ML of fuzzy rules from the training dataset.
- *Decision* Tree [31] generates a binary tree based on training features for classification.
- *Random* Forest [32] builds a classification model using a set of decision tree-based classifiers.
- *K-Nearest Neighbour* [33] classifies an input image based on its similarity with other training set images, which enables the most dominant class of K to be assigned to the input image.
- *Bayesian Classifier* [34] use the Bayes rules to calculate the probability of an input sample to be a member of a specific class where the final label is assigned to the most probabilistic class for the given input image.
- *Linear Discriminant Analysis* [35] learns a linear combination of the features from training images to predict the label of test images.
- *Support Vector Model (SVM)* [36] learns a set of parameters from the training image to find a hyperplane which splits the training images into two classes. Same parameters are then used to classify test images.
- *Gaussian Mixtures Model* [37] learns multiple models from the training images to classify it into multiple classes.
- *Neural* Network [38] methods learn the representation of the training images using a gradient descent-based learning method. These methods require large training datasets compared to other

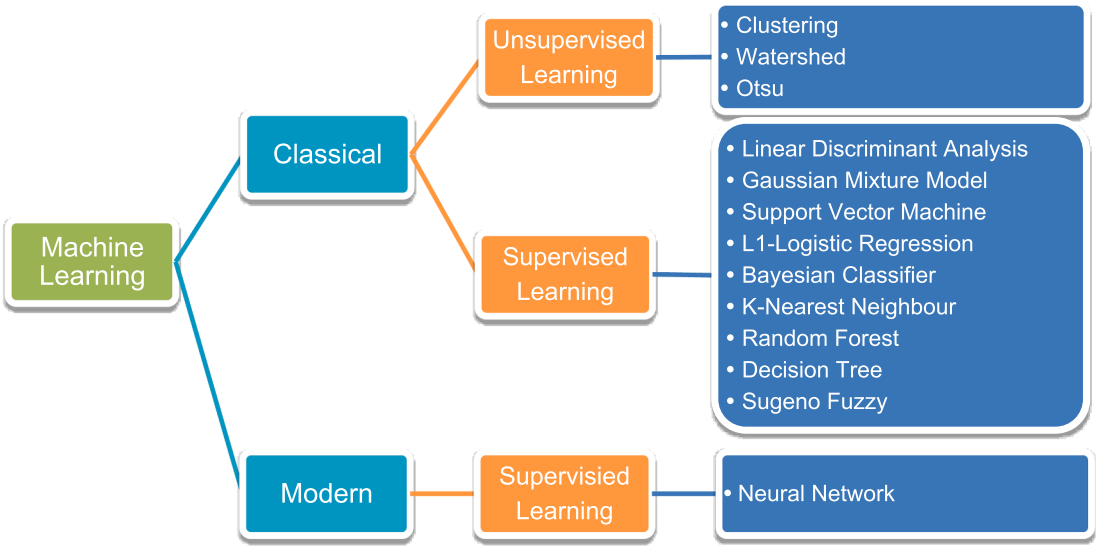

Figure 2. Hierarchical classification of the methods used in the selected studies.

aforementioned supervised based learning methods.

A combination of classification and segmentation methods were used to detect potentially malignant and cancerous head and neck lesions (Fig. 3). Overall, five studies used AI based classifiers [52,54,56,59,62], four studies used segmentation methods [53,58,60,61] and two studies used a combination of classification and segmentation methods [55,57]. The most frequently used methods included SVM [36], Neural Network [38], Random Forest [32] and clustering. In the majority of studies, multiple methods were used for intermediate and final stages of the proposed AI framework, although there were considerable variations in the overall number of methods used between studies. For example, in one study Krishnan *et al.* [54] compared five different classification approaches to obtain the best performing method, whereas in another study only two methods were trialled [58].

Image datasets used in the selected studies

Figure 4 illustrates the dataset sizes (for training and validation) and spatial dimensions of images (in pixels) for the selected studies.

There is apparent variability in sample sizes, with training samples ranging from 8 to 216 images and validation samples ranging from 0 to 208 images. The overall dataset size (including both training and test samples) ranges from 40 to 270 images (mean ~ 139 images).

The spatial dimensions of images ranged from 262,144 to 10,890,000 pixels; this excludes three studies where the image dimensions were not described [52,56,62]. Although the image sizes are measured in pixels, the actual size of the tissue sample (in microns) will differ due to the varying resolutions of different scanners and the magnification level chosen for the images. However, in five studies [54,55,57,58,62] the dataset samples were obtained from the same centre and in two studies [54,58] the same dataset was used.

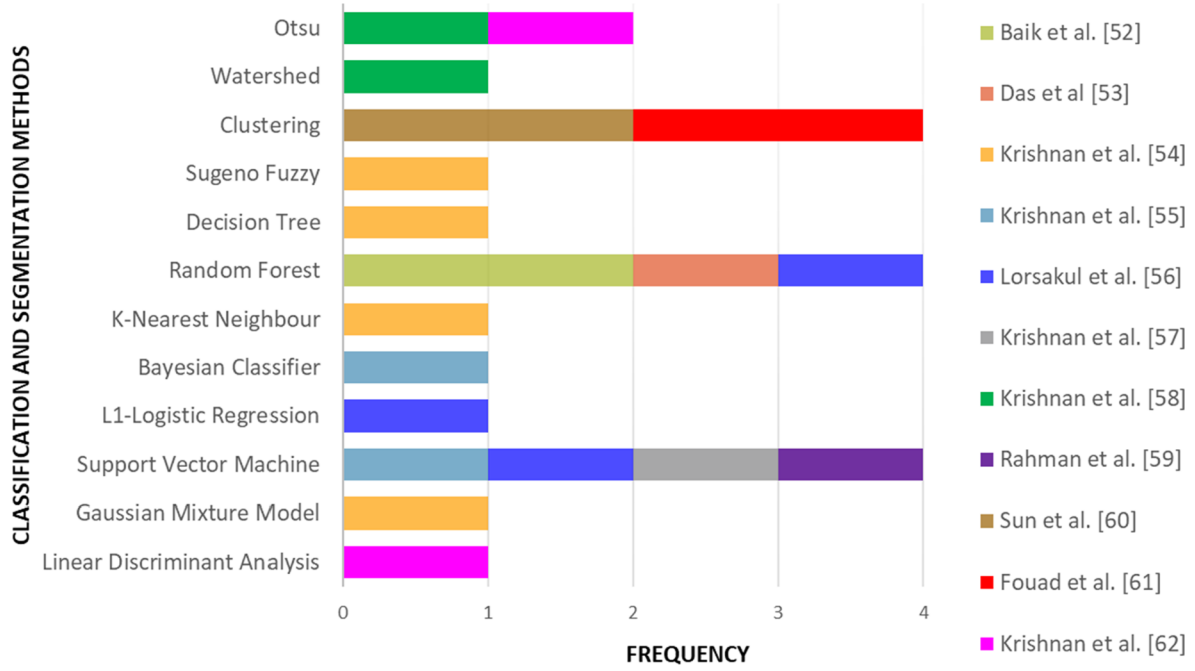

Figure 3. A bar chart representing the frequency of different methods used in the selected studies. The colours correspond to the individual studies (as per key on right).

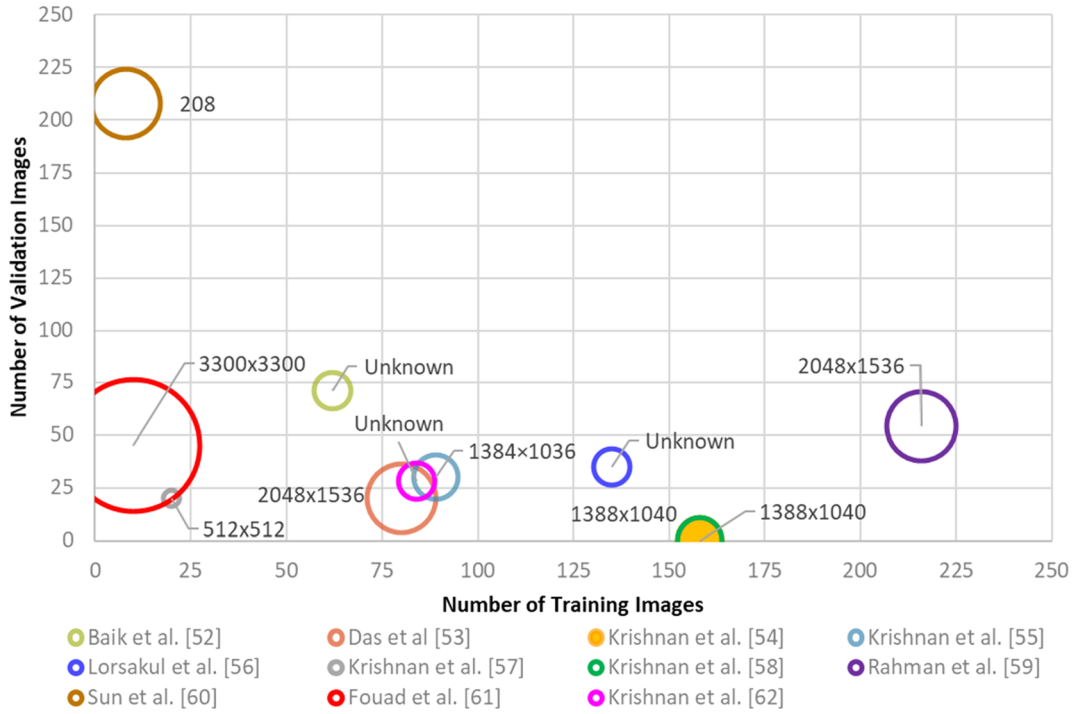

Figure 4. Graph demonstrating dataset sizes for selected studies. The area of the circle represents the spatial dimensions (in pixels) of the images used within the dataset whereas horizontal and vertical axis represent the number of images used for training and validation, respectively. Studies in which the image dimension is not provided have been marked as 'unknown.'

Quality appraisal assessment

The quality of selected studies was assessed using a tailored QUADAS-2 tool. The overall score for each study was determined by combining the number of satisfied criteria, with a higher score representing higher quality evidence (Table 3). There was considerable variability in the methodological quality of included studies. Baik *et al.* [52] scored the highest for including the use of a separate validation and test set for optimal model selection and evaluation [52], whereas Sun *et al.* [60] scored the lowest across the 13 areas of assessment, largely due to the description of their approach without the use of reasonable dataset and results [60]. In all applicable studies, except for Lorsakul *et al.* [56], the methods and intermediate results were clearly presented. Rahman *et al.* [59] was the only study to have used a multi-centric dataset [59].

Discussion

In order to safely and effectively implement automated AI-based methods in diagnostic and clinical practice, it is vital to validate these algorithms using a robust and fair experimental setup. This setup should include a clinically representative dataset and suitable evaluation metrics for validation.

The ideal dataset should represent clinical practice and take into account the whole tissue section. Tissue samples from multiple centres will enable greater diversity and biological variance through inclusion of cases from different geographical locations, patient populations and demographics. Furthermore, the ground truth should include meticulous annotations from multiple pathologists to minimise subjectivity and take into account inter-pathologist variation. In this review, seven studies described more than one expert to be involved in providing ground truth, however it is not clear whether these refer to experienced pathologists, trainees, non-clinical researchers or other allied

Table 3
Quality assessment of the selected studies using modified QUADAS-2 tool. The '✓' demonstrates a favourable response to the question and the 'X' demonstrates an unfavourable response to the question. The overall score reflects the quality and risk of bias for each study.

Quality Assesment Questions	Study Number 1	2	3	4	5	6	7	8	9	10	11
Does the dataset include the complete digitised biopsy section or a complete resection?	✗	✗	✗	✗	✓	✗	✗	✗	✗	✗	✗
Does the dataset consist of more 100 samples?	✓	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
If not, does it consist of more than 50 unique samples?	✓	✗	✓	✗	✗	✗	✗	✓	✗	✓	✗
Is the test dataset separate to the training and validation datasets?	✓	✗	✗	✗	✗	✗	—	✗	✗	✗	✗
Is the biopsy case selection representative of the condition being assessed in the study?	✓	✗	✓	✗	✗	✗	✓	✗	✗	✗	✗
Does the study involve more than one pathologist for annotations?	✓	✗	✓	✗	✓	✓	✓	✓	✗	✗	✓
Is the dataset multi-centric?	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
Does the study use an independent reference/test set?	✓	✗	✗	✗	✗	✗	—	✗	✗	✗	✗
Does the study use an independent validation set for optimal model selection?	✓	✗	✗	✗	✗	✗	—	✗	✗	✗	✗
Does the study fairly compare the outcomes of the AI methods to the existing methods?	✗	✓	✓	✗	✗	✗	✓	✗	✗	✓	✗
Are the compared methods state-of-the-art at the time of the publication of the article?	✗	✓	✓	✗	✗	✗	✗	✗	✗	✓	✗
Was the method described in sufficient detail to reproduce the presented results?	✓	✓	✓	✓	✗	✓	✓	✓	✓	✓	✓
Were the results of the intermediate stages reported?	✓	✓	✓	✓	—	✓	—	—	—	✓	✓
Overall Score	9	4	7	2	2	3	4	5	1	5	3

healthcare professionals. In most cases, the number of experts involved has also not been clearly stated. Furthermore, multi-centric data and WSI were used in only two studies (Rahman *et al.* [59] and Lorsakul *et al.* [56] respectively). The majority of studies therefore used uni-centric datasets mostly comprising smaller sub-images within WSI, which may have introduced bias and would offer limited applicability.

In supervised ML methods, the ideal dataset should be divided into three groups for 1) model training 2) optimal model selection and 3) validation or evaluation. In this review, most of the assessed studies used the same dataset for both optimal model selection and evaluation. This indicates a high risk of bias which is likely to have contributed to the high accuracy rates (ranging from 79 to 100% across all studies). This bias could have been easily avoided by dividing these datasets into the three defined sub-sets or adding more unseen cases to the test and validation sets. The overall size of a dataset mainly depends on the type of AI method used. Traditional AI methods require small datasets whereas modern machine/deep learning methods require a larger dataset for model training. This concept is also known as the model generalisability and is concerned with the replication of model accuracy when applied to a new and diverse cohort of cases.

Most of the reviewed studies used AI methods which were regarded as the state-of-the-art at the time of publication and described these in sufficient detail to ensure reproducibility. Various evaluation metrics were utilised (accuracy, sensitivity, specificity etc.) to report the overall performance on the test set. However, in one study (Krishnan et al. [55]) the performance of individual images was measured, which makes it difficult to gauge the overall average performance of the entire test set. In four studies [53,54,58,61] a direct comparison of proposed techniques has been made to existing methods which should give some credibility to their proposed techniques. However, only three studies [53,54,61] compared their methods with the best performing methods at the time of publication.

Most of the selected studies were published before 2015, therefore, the methods employed were mainly classical ML methods [32,33,36]. This was somewhat surprising as the ML and AI fields have significantly progressed in the last decade, resulting in the development of numerous state-of-the-art algorithms for different real-world problems such as object detection in natural images [39,40,41], human voice recognition [42] and natural language processing [43]. These methods have multiple applications including medical image analysis which have been used to reliably predict diagnosis [18], mutational status [44] and treatment response [45] in a range of malignancies including breast, lung and colorectal cancers. However, our review shows that these latest AI methods have not been applied for detection of head and neck lesions, despite the ever-increasing global incidence and poor prognosis of HNC.

Our review highlights that a huge opportunity (and need) for medical image analysis and computational pathology researchers to develop novel methods to aid HNC diagnosis using modern AI approaches such as deep learning. Early work in this field appears to show potential for reliable detection between normal, potentially pre-malignant and cancerous lesions from histology WSI using classical classification methods [46,47]. Customised deep learning techniques have been used for segmentation of the epithelium [48,49] and cell segmentation has shown successful morphological analysis in HNC [50,51].

Conclusion

This review provides early evidence to support application of supervised ML methods as an aid to detection and grading in a limited number and types of OPMD. Furthermore, there is limited evidence exploring the use of AI to aid diagnosis of other potentially pre-malignant and cancerous head and neck lesions. Having said this, most of the described AI/ML methods have the potential be modified for application to other clinical sites, including other head and neck lesions. The overall performance of the AI methods appears comparable to conventional light microscopic histopathological assessment but with added advantages of a faster, objective and reproducible evaluation. However, integration of these methods in the digital pathology workflow requires comprehensive evaluation of each method based on large multi-centric datasets. Future avenues include the use of deep learning methods for development of digital biomarkers and discovery of novel predictive features which will aid early detection of HNC and improve patient stratification. Ultimately, this will aid the development of targeted, patient-specific diagnostics and therapeutics to reduce HNC associated mortality.

Disclaimer: The content of this article represents the personal views of the authors and does not represent the views of the authors' employers and associated institutions. Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy or views of the International Agency for Research on Cancer/ World Health Organization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Anthea Tucker (University of Sheffield, UK) for her assistance with the database searches.

Appendix 1. . MEDLINE via OVID search strategy

1. artificial intelligence.mp. or exp Artificial Intelligence/
2. machine learning.mp. or Machine Learning/
3. deep learning.mp. or Deep Learning/
4. Image Processing, Computer-Assisted/ or automated detection.mp. or Diagnosis, Computer-Assisted/
5. "Neural Networks (Computer)"/ or neural networks.mp.
6. automated image analysis.mp.
7. digital image analysis.mp.
8. 1 or 2 or 3 or 4 or 5 or 6 or 7
9. Mouth Neoplasms/ or oral epithelial dysplasia.mp. or Leukoplakia, Oral/
10. oral leukoplakia.mp.
11. oral neoplasm.mp.
12. oral precancer.mp.
13. oral cancer.mp.
14. "head and neck cancer".mp. or "Head and Neck Neoplasms"/
15. "head and neck malignancy".mp.
16. 9 or 10 or 11 or 12 or 13 or 14 or 15
17. Diagnosis/ or diagnosis.mp.
18. diagnostic performance.mp.
19. 17 or 18
20. 8 and 16 and 19
21. limit 20 to (english language and humans and last 10 years)

References

[1] 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.
[2] IARC. Working Group on the Evaluation of Carcinogenic Risks to Humans. Alcohol consumption and ethyl carbamate. IARC Monogr Eval Carcinog Risks Hum 2010;96:3.
[3] International Agency for Research on Cancer. Solar and ultraviolet radiation. IARC monographs on the evaluation of carcinogenic risks to humans. 1992; 55.

[4] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Human immunodeficiency viruses and human T-cell lymphotropic viruses. IARC monographs on the evaluation of carcinogenic risks to humans. 1996.
[5] Papillomaviruses H. IARC monographs on the evaluation of carcinogenic risks to humans. Lyon, France: IARC; 2011.
[6] 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 2000;6(3):1046–51.
[7] Shaw R, Beasley N. Aetiology and risk factors for head and neck cancer: United Kingdom National Multidisciplinary Guidelines. J Laryngol Otology 2016;130(S2):S9–12.
[8] Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61(2):69–90.
[9] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7.
[10] Bray F, Ferlay J, Soerjomataram I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68:394.
[11] Cancer Research UK. Cancer Incidence in the UK in 2011.
[12] Liao LJ, Hsu WL, Lo WC, Cheng PW, Shueng PW, Hsieh CH. Health-related quality of life and utility in head and neck cancer survivors. BMC Cancer 2019;19(1):425.
[13] Registry NIC. Queens University Belfast, Incidence by stage 2010–2014. Belfast: NICR; 2016.
[14] Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E. Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. J Oral Pathol Med 2008;37(3):127–33.
[15] Lumerman H, Freedman P, Kerpel S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology. 1995;79(3):321–9.
[16] 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.
[17] Mehlum CS, Larsen SR, Kiss K, Groentved AM, Kjaergaard T, Möller S, et al. Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. Laryngoscope 2018;128(10):2375–9.
[18] Bejnordi BE, Veta M, Van Diest PJ, 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. JAMA 2017;318(22):2199–210.
[19] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Comput Struct Biotechnol J 2015;31(13):8–17.
[20] Graham, S. & Rajpoot, N. SAMS-NET: Stain-aware multi-scale network for instance-based nuclei segmentation in histology images. In Biomedical Imaging (ISBI 2018), 2018 IEEE 15th International Symposium on, 590–594 (IEEE, 2018).
[21] LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015;521(7553):436–44.
[22] Awan R, Sirinukunwattana K, Epstein D, Jefferyes S, Qidwai U, Aftab Z, et al. Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images. Sci Rep 2017;7(1):16852.
[23] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nat Rev Clin Oncol 2019;16(11):703–15.
[24] Zormpas-Petridis K, Failmezger H, Raza SE, 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 2019;9:1045.
[25] Wang S, Yang DM, Rong R, Zhan X, Fujimoto J, Liu H, et al. Artificial intelligence in lung cancer pathology image analysis. Cancers. 2019;11(11):1673.
[26] Sirinukunwattana K, Shan e Ahmed Raza, Tsang YW, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans. Med. Imaging. 2016 May 1;35(5):1196-206.
[27] Veta M, Van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A, et al. Assessment of algorithms for mitosis detection in breast cancer histopathology images. Med Image Anal 2015;20(1):237–48.
[28] Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg 2010;8:336–41.
[29] Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011;155(8):529–36.
[30] Yen J, Langari R, Zadeh LA, editors. Industrial applications of fuzzy logic and intelligent systems. IEEE; 1995 Apr 1.
[31] Quinlan JR. Induction of decision trees. Machine learning. 1986;1(1):81–106.
[32] Breiman L. Random forests. Machine learning. 2001;45(1):5–32.
[33] Altman NS. An introduction to kernel and nearest-neighbor nonparametric regression. Am Statistician 1992;46(3):175–85.
[34] Rish I. An empirical study of the naive Bayes classifier. InIJCAI 2001 workshop on empirical methods in artificial intelligence 2001 Aug 4 (Vol. 3, No. 22, pp. 41-46).
[35] Mika S, Ratsch G, Weston J, Scholkopf B, Mullers KR. Fisher discriminant analysis with kernels. InNeural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no. 98th8468) 1999 Aug 25 (pp. 41-48). Ieee.
[36] Suykens JA, Vandewalle J. Least squares support vector machine classifiers. Neural Process Lett 1999;9(3):293–300.
[37] Bilmes JA. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. Int Comput Sci Inst 1998;4(510):126.
[38] Fine TL. Feedforward neural network methodology. Springer Science Business Media 2006:6.
[39] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition 2016:770–8.
[40] Szegedy C, Vanhoucke V, Ioffe S, Shlens J. Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition 2016. p. 2818–26.
[41] Chollet F. Xception: Deep learning with depthwise separable convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition 2017. p. 1251–8.
[42] Graves A, Fernández S, Gomez F, Schmidhuber J. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. InProceedings of the 23rd international conference on Machine learning 2006 Jun 25 (pp. 369-376).
[43] Young T, Hazarika D, Poria S, Cambria E. Recent trends in deep learning based natural language processing. ieee Computational intelligenCe magazine. 2018 Jul 20;13(3):55-75.
[44] Kather JN, Pearson AT, Halama N, Jäger D, Krause J, Loosen SH, et al. Deep learning can predict microsatellite instability directly from histology in gastrointestinal cancer. Nat Med 2019;25(7):1054–6.
[45] 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):1–3.
[46] Graham S, Shaban M, Qaiser T, Koohbanani NA, Khurram SA, Rajpoot N. Classification of lung cancer histology images using patch-level summary statistics. InMedical Imaging 2018: Digital Pathology 2018 Mar 6 (Vol. 10581, p. 1058119). International Society for Optics and Photonics.
[47] Hou L, Samaras D, Kurc TM, Gao Y, Davis JE, Saltz JH. Patch-based convolutional neural network for whole slide tissue image classification. Proceedings of the IEEE conference on computer vision and pattern recognition 2016. p. 2424–33.
[48] Fraz MM, Khurram SA, Graham S, Shaban M, Hassan M, Loya A, et al. FABnet: feature attention-based network for simultaneous segmentation of microvessels and nerves in routine histology images of oral cancer. Neural Comput Appl 2019;3:1–4.
[49] Chan L, Hosseini MS, Rowsell C, Plataniotis KN, Histosegnet DS. Semantic segmentation of histological tissue type in whole slide images. Proceedings of the IEEE International Conference on Computer Vision 2019:10662–71.
[50] Raza SE, Cheung L, Shaban M, Graham S, Epstein D, Pelengaris S, et al. Micro-Net: A unified model for segmentation of various objects in microscopy images. Med Image Anal 2019;1(52):160–73.
[51] Graham S, Vu QD, Raza SE, Azam A, Tsang YW, Kwak JT, et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal 2019;1(58):101563.
[52] 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. Cellular Oncol 2014;37(3):193–202.
[53] Das DK, Bose S, Maiti AK, Mitra B, Mukherjee G, Dutta PK. Automatic identification of clinically relevant regions from oral tissue histological images for oral squamous cell carcinoma diagnosis. Tissue Cell 2018;1(53):111–9.
[54] Krishnan MM, Venkatraghavan V, Acharya UR, Pal M, Paul RR, Min LC, et al. Automated oral cancer identification using histopathological images: a hybrid feature extraction paradigm. Micron. 2012;43(2–3):352–64.
[55] Krishnan MM, Shah P, Chakraborty C, Ray AK. Statistical analysis of textural features for improved classification of oral histopathological images. J Med Syst 2012;36(2):865–81.
[56] Lorsakul A, Andersson E, Harring SV, Sade H, Grimm O, Bredno J. Automated wholeslide analysis of multiplex-brightfield IHC images for cancer cells and carcinoma-associated fibroblasts. InMedical Imaging 2017: Digital Pathology 2017 Mar 1 (Vol. 10140, p. 1014007). International Society for Optics and Photonics.
[57] Krishnan MM, Pal M, Bomminayuni SK, Chakraborty C, Paul RR, Chatterjee J, et al. Automated classification of cells in sub-epithelial connective tissue of oral submucous fibrosis—An SVM based approach. Comput Biol Med 2009;39(12):1096–104.
[58] Krishnan MM, Choudhary A, Chakraborty C, Ray AK, Paul RR. Texture based segmentation of epithelial layer from oral histological images. Micron. 2011;42(6):632–41.
[59] Rahman TY, Mahanta LB, Chakraborty C, Das AK, Sarma JD. Textural pattern classification for oral squamous cell carcinoma. J Microsc 2018;269(1):85–93.
[60] Sun YN, Wang YY, Chang SC, Wu LW, Tsai ST. Color-based tumor tissue segmentation for the automated estimation of oral cancer parameters. Microsc Res Tech 2010;73(1):5–13.
[61] Fouad S, Randell D, Galton A, Mehanna H, Landini G. Unsupervised morphological segmentation of tissue compartments in histopathological images. PloS one. 2017;12(11).
[62] Mookiah MR, Shah P, Chakraborty C, Ray AK. Brownian motion curve-based textural classification and its application in cancer diagnosis. Anal Quant Cytol Histol 2011;33(3):158–68.

3.3 Demystifying Oral Epithelial Dysplasia: A Histological Guide

Pathology (February 2024) **56(1)**, pp. 11–23

REVIEW

Demystifying oral epithelial dysplasia: a histological guide

Paul Hankinson[1,*], Hanya Mahmood[2,*], Hannah Walsh[1], Paul M. Speight[1], Syed Ali Khurram[1]

[1]*Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, England, UK;* [2]*Academic Unit of Oral and Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield, England, UK;* [*]*Joint first authors*

Summary

Oral epithelial dysplasia is a histologically diagnosed potentially premalignant disorder of the oral mucosa, which carries a risk of malignant transformation to squamous cell carcinoma. The diagnosis and grading of oral epithelial dysplasia is challenging, with cases often referred to specialist oral and maxillofacial pathology centres for second opinion. Even still there is poor inter-examiner and intra-examiner agreement in a diagnosis. There are a total of 28 features of oral epithelial dysplasia listed in the 5th edition of World Health Organization classification of tumours of the head and neck. Each of these features is poorly defined and subjective in its interpretation. Moreover, how these features contribute to dysplasia grading and risk stratification is even less well defined. This article discusses each of the features of oral epithelial dysplasia with examples and provides an overview of the common mimics, including the normal histological features of the oral mucosa which may mimic atypia. This article also highlights the paucity of evidence defining these features while offering suggested definitions. Ideally, these definitions will be refined, and the most important features identified to simplify the diagnosis of oral epithelial dysplasia. Digital whole slide images of the figures in this paper can be found at: https://www.pathogenesis.co.uk/r/demystifying-dysplasia-histology-dataset.

Key words: Oral epithelial dysplasia; oral pre-cancer; malignant transformation; oral cancer; oral squamous cell carcinoma; histological grading.

Received 7 July, revised 25 September, accepted 5 October 2023
Available online 16 November 2023

INTRODUCTION

Oral epithelial dysplasia (OED) is a disorder of oral mucosa that is diagnosed by histological identification of architectural and cytological abnormalities of the oral epithelium.[1] Some of these abnormalities are long established,[2,3] whereas others are described in more recent diagnostic criteria.[1,4,5] OED carries a risk of malignant transformation to oral squamous cell carcinoma (OSCC), though reported transformation rates vary.[6,7] OED develops as a result of genomic alterations which are often shared with the later carcinoma if transformation occurs.[8] However, most OED do not transform and many regress without intervention.[7] OED is usually initiated by chemical carcinogens such as tobacco smoke and alcohol,[3] with a smaller number driven by high-risk human papilloma virus (HPV) infection.[9] This is in contrast to cervical dysplasia where HPV is the most common cause,[10] or epidermal dysplasia where ultraviolet light exposure plays a significant role.[11]

OED manifests clinically as white, red or mixed lesions, though other presentations can be seen. Attempts have been made to correlate clinical appearances of OED to malignant transformation;[12] however, the supportive evidence is weak.[7]

Diagnosis of OED is complex and there are many mimics, due to the phenomenon of reactive atypia in many inflammatory diseases of the oral cavity. A diagnosis is reached by identification of a range of architectural and cytological features, though in some cases an individual feature, if extensive enough, may qualify a diagnosis. Some cases of OED are straightforward with obvious cytological abnormalities. However, the so called 'differentiated' or 'architectural dysplasias' tend to lack frank cytological atypia with architectural changes being the predominant feature making these cases more challenging to diagnose.[4] In the latest World Health Organization (WHO) classification of head and neck tumours[1] there are 28 histological features listed, most of which are poorly defined and with limited evidence of correlation to clinical outcomes.[4] OED is graded to assist with prognostication and treatment planning, with higher graded lesions believed to have a higher risk of malignancy.[6] However, OED grading is an unreliable predictor of cancer risk, complicated by the multiple proposed grading systems,[1,13] and wide inter- and intra-examiner variability.[14–17] Complicating matters further, there is no minimum agreed number or 'extent' of features required for a diagnosis in the WHO classification, and no consideration to the importance of individual features is given. The binary grading system proposed to simplify this by suggesting two categories with a minimum number of features for grading. However, the 'extent or abundance' of a feature, correlation of individual features and analysis of verrucous and differentiated dysplasia are still not considered, hence robust evidence to support its routine use is lacking.[13] The complexity of OED diagnosis often leads to referrals to specialist centres by pathologists inexperienced in examining oral mucosa.[18] However, variabilities and inconsistencies exist even amongst these specialists.[17]

The advent of digital pathology, artificial intelligence (AI) and machine learning (ML) has increased opportunities to explore tissue sections through automated and quantitative means, allowing more objective analysis of histological

Print ISSN 0031-3025/Online ISSN 1465-3931
DOI: https://doi.org/10.1016/j.pathol.2023.10.002

features of disease. These methods are increasingly used to study OED and may help with defining the histological criteria. At present, there remain many questions with regards to a diagnosis: how many basal cells constitute basal cell crowding; how superficial does a high mitotic figure need to be, and how much variation in nuclear size constitutes pleomorphism? Without strict definitions, there will continue to be subjectivity for the diagnosis and grading of OED. In Table 1 we have proposed working definitions for individual OED features listed in the 5th edition WHO classification, as these features are considered the 'gold standard' at present.[1] We believe that the list of proposed features in this classification is somewhat academic and excessive with poor scientific evidence and prognostic correlation. Though many of

Table 1 Proposed definitions for the architectural and cytological features of oral epithelial dysplasia presented in the upcoming World Health Organization 5th edition classification of head and neck tumours[1]

Features	Proposed definition
Architectural features (Fig. 4/5)	
1. Irregular stratification (Fig. 4A)	Disturbance of the stratified layers of the epithelium, with haphazardly organised and difficult to distinguish layers.
2. Loss of basal cell polarity (Fig. 4E)	Abnormal nuclear location (away from the basement membrane) and abnormal nucleus orientation (no longer parallel with other basal cell nuclei).
3. Drop shaped rete processes (Fig. 5C)	The rete process is broader at the base than at the apex.
4. Basal cell clustering/nesting (Fig. 5E)	This feature is better defined in skin with crowding of atypical basal cells, followed by budding into the lamina propria, eventually taking on an irregular outline and exceeding the thickness of the epithelium.[26,27]
5. Expanded proliferative compartment	Thickening of the basal cell layer with evidence of mitotic activity.
6. Mitoses high in the epithelium (Fig. 5B)	Mitotic figures present outside the basal cell layer. No consensus or evidence about the minimum number for dysplasia. There may be value in combining this feature with mitoses in maturing cells.
7. Mitoses in maturing cells (Fig. 5B)	Mitotic figure present in the prickle cell or granular cell layers. No consensus or evidence about the minimum number for dysplasia.
8. Generalised premature keratinisation (Fig. 4B/5C)	Increased prickle cell cytoplasmic eosinophilia due to keratinisation in excess of what is normally expected at that oral cavity site.
9. Keratin pearls in rete processes (Fig. 4D)	The formation of an intra-epithelial collection of keratin with no surface connection. It is better to consider any keratin pearl formation as dysplastic.
10. Reduced keratinocyte cohesion (Fig. 4E)	A spectrum of changes; begins with widening intercellular junctions and ends with acantholysis.
11. Altered keratin pattern for oral sub-site	An increase in the amount or change in the type of keratin from the normal at that subsite, in the absence of features of trauma.
12. Verrucous/papillary architecture (Fig. 5A/5C)	Verrucous: hyperkeratinised surface composed of sharp or blunt epithelial projections with keratin filled invaginations without fibrovascular cores. Papillary: exophytic projections of epithelium with fibrovascular cores.
13. Extension along salivary ducts (Fig. 4F)	Features of dysplasia observed within salivary ducts adjacent the dysplastic surface epithelium.
14. Sharply defined lesion (Fig. 4C)	Abrupt transition between the normal epithelium and one with the features of dysplasia.
15. Multiple patterns of dysplasia (Fig. 5F)	Multiple regions of dysplasia each with a distinct collection of dysplastic features, either adjacent each other or separated by areas of normal epithelium.
16. Multifocal or skip lesions	Multiple epithelial lesions with features of dysplasia clearly separated by areas of normal epithelium.
Cytological features (Fig. 6)	
17. Abnormal variation in nuclear size	Variation in the size of keratinocyte nuclei beyond that expected in a normal epithelium, there is no current definition of normal oral epithelial nuclear size variation.
18. Abnormal variation in nuclear shape	Variation in the shape of keratinocyte nuclei beyond that expected in a normal epithelium, usually deviation from a circular to oval or an irregular shaped nucleus.
19. Abnormal variation in cell size	Variation in the size of keratinocytes beyond that expected in a normal epithelium, there is no current definition of normal oral keratinocyte size variation.
20. Abnormal variation in cell shape	Variation in the shape of keratinocytes beyond that expected within that epithelial layer, basal cells are usually more cuboidal, prickle cells are usually more polygonal and granular cells are usually the most squamoid.
21. Increased mitotic activity	Readily identifiable, numerous, normal mitotic figures throughout the thickness of the epithelium.
22. Increased nuclear size	Larger nuclear size beyond what would be normal for that epithelial location, the normal nucleus size has not been defined. It may be best to combine increased nuclear size with increased nucleus to cytoplasmic ratio.
23. Increased nucleus: cytoplasm (N:C) ratio	Increase in the nuclear size leading to reduction of the cytoplasmic area normal for the epithelial location. The normal nuclear to cytoplasmic ratio has not been defined.
24. Atypical mitotic figures	Readily identifiable mitoses which do not have a normal morphology.
25. Increased number and size of nucleoli	A greater number of nucleoli or a larger size of at least one nucleolus, beyond what is normal in the oral epithelium. The normal number and size of nucleoli is not known.
26. Single cell keratinisation	Individual cells with keratinisation giving the cytoplasm a strongly eosinophilic appearance with retraction from neighbouring keratinocytes.
27. Nuclear hyperchromasia	Keratinocyte nuclei with greater basophilia than would be normal for the epithelial location, the normal degree of basophilia has not been defined.
28. Apoptotic mitoses	This relates to the observation of a mitotic catastrophe, though the histological appearance is not well defined in the literature.

these features can be seen in OED, future research should focus on determining which features are most common, the threshold of these features for diagnosis, and which of these are most important prognostically. By reducing the number in this list to only the most important, merging some of the overlapping cytological and architectural features and assessing the quantification/weighting of features with prognostic importance, some clarity may be gained not only for reporting pathologists but also for the surgical teams involved in treatment.

The main aim of this paper is to review the currently known and suggested histological features of OED and offer definitions which may be less opaque to pathologists, students, trainees and patient-facing clinicians. In this context, the authors also report on the histology of healthy oral tissue and other histological mimics to highlight the complexity of OED diagnosis. Digital whole slide images of all the figures provided in this paper can be found at: https://www.pathogenesis.co.uk/r/demystifying-dysplasia-histology-dataset. Finally, the authors comment on the challenges and limitations of the current diagnostic criteria and existing grading systems for prediction of malignant transformation risk.

NORMAL ORAL MUCOSA

The first step in OED diagnosis is recognising variations of the normal oral epithelium (Fig. 1–3). Each subsite has unique features that may be misinterpreted as OED. Oral mucosa can be placed into three broad categories: (1) masticatory mucosa (gingivae, hard palate), (2) lining mucosa (labial, buccal, ventral tongue, floor of mouth, soft palate), and (3) specialised mucosa (dorsal tongue, vermillion lip border)[19] (Fig. 1–3). Although oral stratified squamous epithelium has the same layers as the epidermis (basal cell layer, prickle cell layer, granular cell layer and keratinised layer) these vary in extent, and may not always be apparent.[19] The specialist mucosa of the tongue is highly variable and most likely to be confused with OED, though OED accounts for less than 5% of cases at this site.[20] In the anterior tongue the filiform papillae form sharp projections of parakeratin and the filiform papillae of the posterior tongue form broader projections with fibrovascular cores. These structures must not be mistaken for verrucous or papillary changes associated with some OED lesions.[21]

ARCHITECTURAL FEATURES OF OED

Irregular stratification

Irregular stratification is the first architectural feature listed in the recent WHO criteria.[1] This feature is long established,[2,3] though poorly defined. Keratinocytes mature as they approach the surface of the normal oral epithelium, having a distinct morphology in each of the layers. The thickness of each epithelial layer varies by site, but each layer should be easily identified. Irregular stratification describes disorder of this maturation, giving the epithelium a haphazard appearance. Cells from each layer become admixed making the layers harder to distinguish (Fig. 4A).[22] There is no definition for how much disorder is required to count this feature, though the epithelial thickness can be used to inform OED grading.[1,4] The tastebuds of the posterior tongue may give a disordered appearance to the epithelium, but this should not be mistaken for irregular stratification of OED (Fig. 3D,E). Irregular stratification is best identified at low power to allow comparison with adjacent normal epithelial layers.

Abnormalities of the basal compartment

Architectural abnormalities of the basal cell layer are common in OED and account for three of the listed features mentioned in the WHO criteria.[1] In the normal epithelium, basal keratinocytes are cuboidal with round to oval nucleus positioned adjacent to the basement membrane (Fig. 2B).[23] In OED, this regular arrangement is lost, and the nuclei of basal cells become abnormally located away from the basement membrane (known as loss of basal cell polarity) (Fig. 4E).[1–3] This feature gives a disordered appearance to the basal layer as the basal cell nuclei no longer occupy a consistent location. Unfortunately, basal cell palisading is variable in the oral cavity and may not be fully appreciated in crosscut sections, making interpretation difficult. Additionally, a lichenoid pattern of inflammation, often present in OED and inflammatory oral diseases, causes disruption of the basal cell layer, further complicating interpretation.[1,3,4,24]

Bulbous (or tear drop) rete processes are a long-described feature of OED (Fig. 4B, 5C).[1–4] These can be defined as an increase in the width of the rete processes, leading to a broader base with a narrow isthmus where the rete process joins the superficial epithelium. Normal rete processes

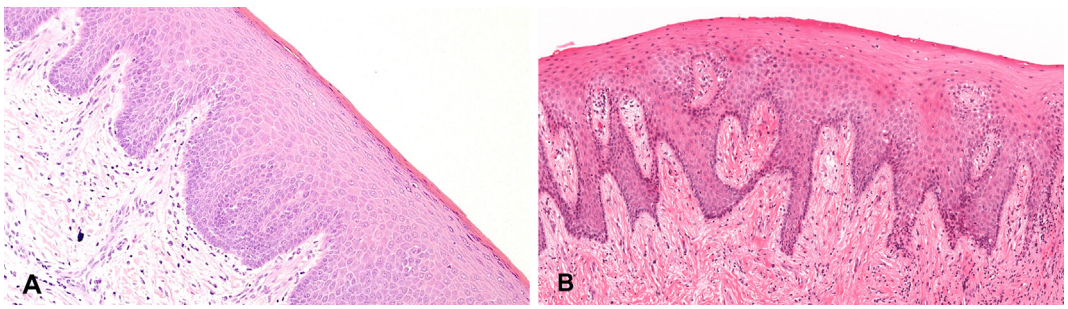

Fig. 1 Examples of the masticatory mucosa of the oral cavity. The masticatory mucosa is exposed to friction during mastication and possesses a parakeratinised or orthokeratinised surface, has a thicker epithelium than lining mucosa and a denser lamina propria. As a result of the keratinisation, masticatory mucosa is usually more eosinophilic than lining mucosa. If eosinophilia is seen in the lining mucosa, however, it may represent the premature keratinisation of OED. This is one example where variation in normal oral mucosa may be mistaken for OED or a potential OED missed. (A) Hard palate: the hard palate is a site subjected to high masticatory forces and comprises orthokeratinised stratified squamous epithelium. There is a visible granular cell layer in this epithelium, which is often not present in the oral epithelium. (B) Gingivae: this is a form of masticatory mucosa comprising parakeratinised stratified squamous epithelium with elongated and branching rete ridges. Note the lamina propria is more densely collagenous at this site.

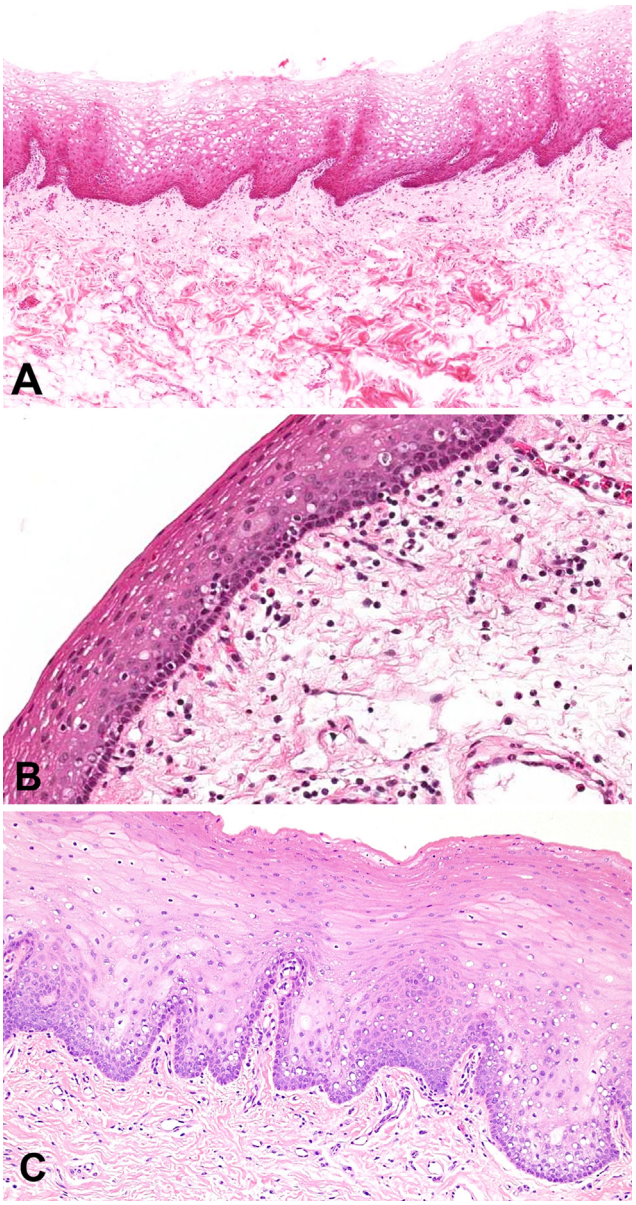

Fig. 2 Examples of lining mucosa of the oral cavity. The lamina propria of the lining mucosa is looser and the submucosa often contain abundant minor salivary gland lobules. The epithelium is non-keratinised, lacks a granular layer and the keratinocytes may contain significant cytoplasmic glycogen. The thickness of the epithelium and length of rete processes is variable with the labial and buccal mucosa being thicker (with longer rete processes) than the floor of mouth and ventral tongue which generally have a flat interface with the lamina propria. The buccal mucosa is also more prone to keratinisation because of trauma than other lining mucosa sites. This must not be mistaken for abnormal keratinisation in OED. (A) Buccal mucosa: this is a form of lining mucosa from an area not usually subjected to highly abrasive forces, therefore the epithelium is non-keratinised with mildly branching rete ridges. (B) Ventral tongue: similar to the buccal mucosa the ventral tongue is a type of lining mucosa which is non-keratinised and is approximately 10–14 cells in thickness with a flat basal compartment lacking rete ridges. The basal keratinocytes can be seen to have a regular arrangement and orientation towards the basement membrane. The underlying submucosa is loose and when combined with the epithelial features makes it a useful site for rapid absorption of various medications. (C) Soft palate: the soft palate is a form of lining mucosa composed of non-keratinised stratified squamous epithelium with small rete ridges and looser submucosa.

maintain an even width or become narrower with depth. Even in reactive or inflammatory lesions, rete processes may increase in length but do not significantly increase in width. At some oral sites (floor of mouth, ventral tongue) the normal basal layer is essentially flat with few or small rete processes (Fig. 2B).[25] As such, any degree of bulbosity should raise concern at these sites. There is no definition of how broad a rete process must be, or how narrow its isthmus, to consider it bulbous.

Basal cell clustering and nesting (Fig. 5E) is a recently proposed feature of OED.[1] This feature has been described in epidermal dysplasia, with three defined stages.[26] First, crowding of atypical basal cells within the epithelium, followed by slight budding of basal cell nests into the papillary dermis, and finally rounded nest of atypical basal cells extending beyond the deepest epidermis without features of frank invasion.[26,27] The presence of this feature in skin has shown to be more important for malignant transformation than upward spread of dysplasia throughout the epidermis.[26] This budding may be suggestive of the keratinocytes moving towards an invasive phenotype;[4] however, it has not been reported in OED to date, nor has its association with malignant transformation been proven in the oral cavity. Another feature common in epidermal dysplasia is extension along adnexal structures.[26,27] Rarely, OED can be seen extending downwards into and along salivary ducts[28] (Fig. 4F). This must not be confused with invasion; useful clues include identification of duct lumen and a lack of stromal reaction around the dysplastic duct.

Tangentially sectioned bulbous rete processes or nested basal cells may simulate invasion. It may be particularly difficult in some cases to distinguish between OED and early invasive OSCC. Further tissue levels are often of benefit as it may reveal that the apparently invasive islands are crosscut rete processes which join up with the epithelium, or alternatively that true invasion exists. In some cases, it is impossible to completely rule out invasion and so a report expressing this uncertainty must be issued. Communication with the surgical team can allow an appropriately conservative excision on which interpretation is often easier.

Number, type and location of mitoses

The next three architectural features of OED are related to mitoses. The first is expansion of the proliferative compartment.[1] In the normal epithelium, the basal cells are the only mitotically active keratinocytes. The basal cell layer is normally only a few cells thick but is often thicker in OED, leading to an increased thickness of the epithelium occupied by mitotically active cells. Some authors have suggested the use of adjuncts such as Ki-67 staining to highlight the altered distribution of keratinocytes in the cell cycle.[4] Positive Ki-67 staining can be seen in S, G2 and M phases of the cell cycle, with variable staining in the G1 phase.[29] Some evidence suggests suprabasal expression can be indicative of dysplasia.[30] However, Ki-67 as an adjunct to dysplasia diagnosis is not widely used due to lack of robust validation. Specific mitosis markers have been investigated in many other tumours and may be more valuable in highlighting

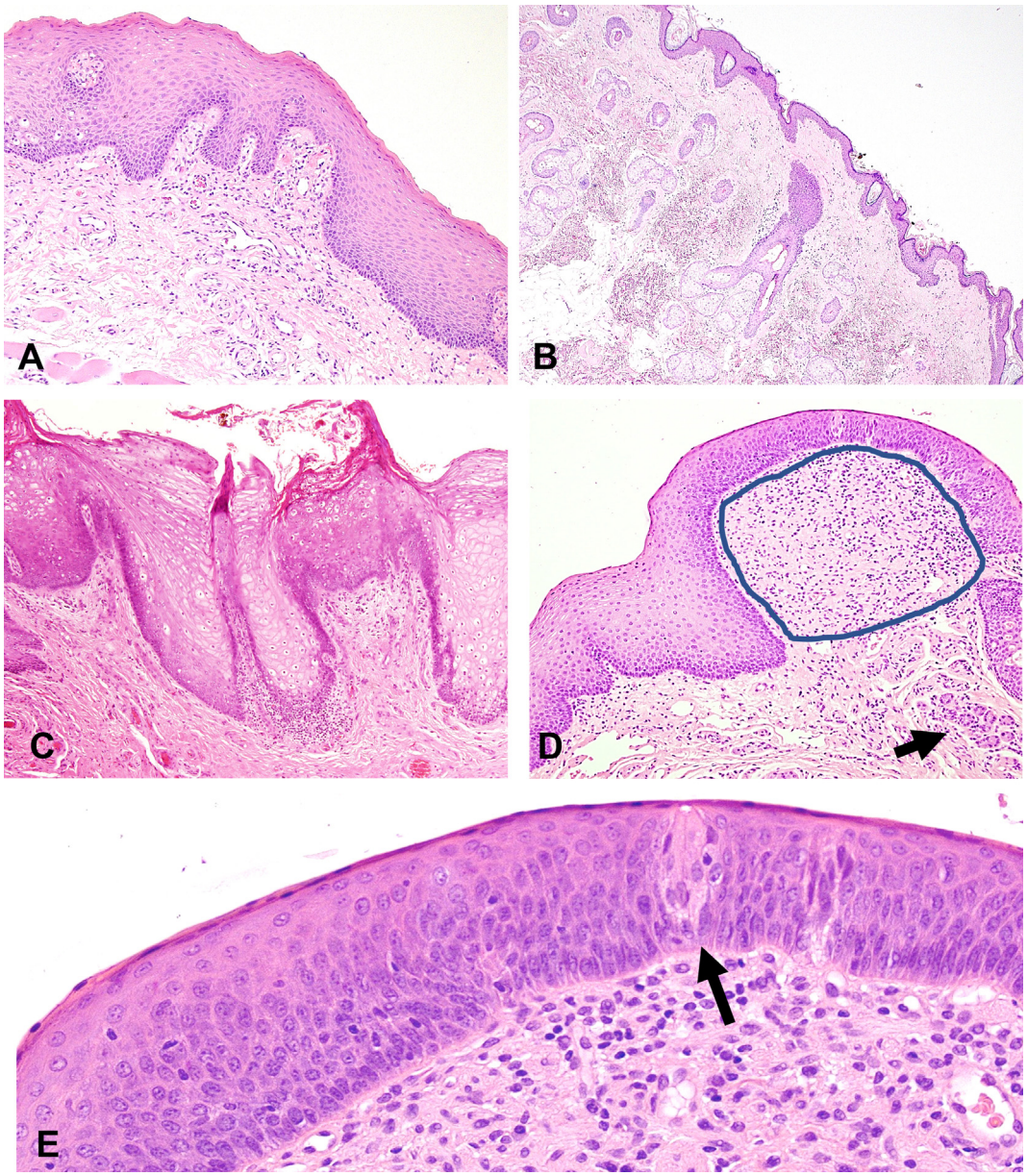

Fig. 3 Examples of specialised mucosa of the lip and tongue. The dorsal tongue epithelium is parakeratinised, thicker than at other sites with long rete processes and forms several types of papillae depending on location. In the anterior tongue, there are filiform papillae with sharp parakeratin projections. The more posterior fungiform papillae have broader projections with an associated specialised lamina propria rich in blood vessels. The epithelium of the fungiform papillae tends to be thinner, and taste buds may be present. Where taste buds are present, there is often disruption of the epithelium, giving it a somewhat disordered appearance, which may be mistaken for dysplasia. The identification of taste buds, however, helps rule out OED. (A) Lip vermillion border, (B) lip skin: the lip is a unique site of the oral cavity as on one aspect it is lined by non-keratinised lining mucosa and on the other aspect is lined by orthokeratinised skin. The transition between these two types of epithelium is called the vermillion border. The vermillion border comprises parakeratinised stratified squamous epithelium with branching rete ridges. The underlying lamina propria is thin and tightly bound to the underlying muscle. (C) Dorsum tongue: showing filiform papillae; the dorsum tongue is subject to masticatory forces and thus comprises parakeratinised stratified squamous epithelium overlying variably dense fibrous connective tissue. The dorsum tongue is also where structures such as filiform papillae are identified. Filiform papillae are heavily parakeratinised with a narrow core of connective tissue. (D) Posterior tongue: showing several structures of the normal posterior tongue; blue ring (subgemmal neurogenous plaque, arrow) serous glands of von Ebner. The subgemmal neurogenous plaque is a neural plexus that serves the overlying taste buds. The polypoid appearance of the mucosa is due to the presence of fungiform papillae. (E) Posterior tongue: showing the surface epithelium of the posterior tongue (taste bud, arrow). The adjacent epithelium is disrupted by the taste bud mimicking disordered stratification.

abnormally distributed mitoses in the oral epithelium.[31] Finally, it should be noted that an increased number of normal mitotic figures limited to the basal layer does not qualify a diagnosis of OED as this can be seen in reactive and inflammatory conditions.

Mitoses high in the epithelium (Fig. 5B), which are referred to as superficial mitotic figures are another feature of OED.[1–3] Any mitotic figure observed superficial to the basal layer can be considered aberrant and the location may range from immediately adjacent to the basal layer through to the surface. These superficial mitoses may be in maturing keratinocytes (as discussed below) or immature cells which have retained a basal cell morphology.

Finally, mitoses in maturing cells can be considered distinct from mitoses high in the epithelium.[1] For example, a mitosis immediately adjacent to the basal cells, and therefore not high in the epithelium, but occurring in a mature cell (such as a prickle cell keratinocyte) is abnormal and can be a

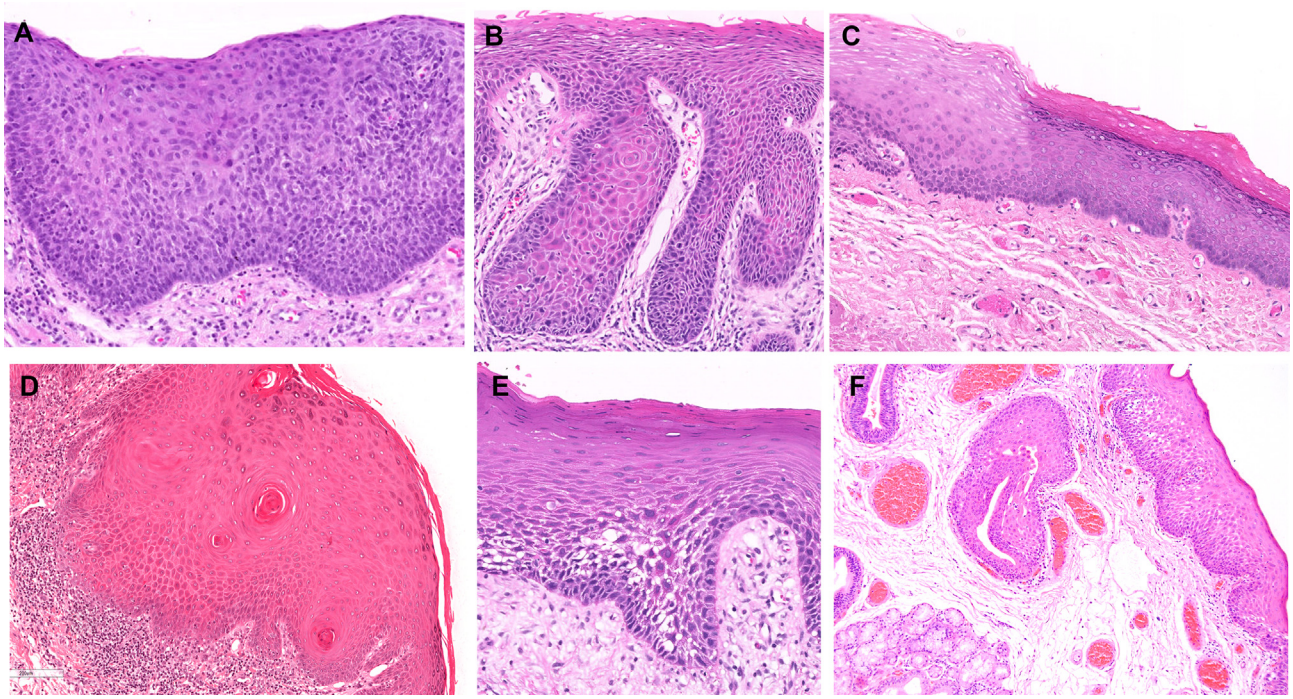

Fig. 4 Examples of the range of architectural abnormalities seen in OED. (A) Irregular stratification: in this example the regular stratification of the epithelium is disordered, and it is difficult to identify and distinguish the basal cell and prickle cell layers. (B) Premature keratinisation: a group of cells in the prickle cell layer show bright eosinophilic cytoplasm and retraction from neighbouring keratinocytes, highlighting premature keratinisation. Only cells within the superficial keratin layer should exhibit such an eosinophilic appearance. (C) Abrupt transition: a sharp and defined change from normal epithelium to dysplastic epithelium is seen. The dysplastic epithelium shows a change in keratin pattern (from parakeratinised to orthokeratinised), and shows bulbous rete ridge morphology alongside basal cell hyperplasia and nuclear and cellular pleomorphism in the basal compartment. (D) Keratin pearl formation: intra-epithelial spherical collections of keratin without extension to the overlying epithelium are identified within the spinous layer of the epithelium. (E) Loss cell epithelial cohesion and basal cell polarity: retraction between adjacent keratinocytes and subsequent loss of cohesion is identified in the prickle cell layer; in addition the nuclei in the basal compartment are seen in an abnormal location (i.e., away from the basement membrane). (F) Extension of OED along a salivary duct: OED is observed in the epithelium in this section with similar features identified in a salivary gland duct (such as loss of cell cohesion) in the underlying connective tissue. In deeper tissue sections the salivary gland duct would show communication with the surface epithelium.

feature of OED.[4] This is also the case for mitotic figures present in the granular layer. However, in order to provide clarity and reduce the extensive number of OED criteria it may be valuable to group superficial mitoses and mitoses in maturing cells as one feature (abnormally located mitotic figures).

It may be difficult to interpret these features in practice. Assessment of the location of mitotic figures may be hampered by cross cutting of the epithelium, giving mitotic figures a more superficial appearance. There is also no definition for the minimum number of abnormally located mitotic figures to qualify an OED diagnosis, therefore this feature needs to be taken in the context of other features. Finally, reactive and inflammatory conditions often lead to increased mitoses and basal cell hyperplasia which may mimic the expanded proliferative compartment of OED.

Abnormal keratinisation

Abnormal keratinisation is another important feature of OED that bridges both cytological and architectural changes. A change in the thickness or type of keratin abnormal for that oral mucosal site, for example, from parakeratin to orthokeratin or non-keratinised to keratinised, may be part of the spectrum of abnormal keratinisation. This feature is often present in differentiated dysplasias, where an abrupt 'clonal' change between parakeratin and orthokeratin is seen with limited or no atypia.[4,32,33] This feature must be interpreted with caution as it can appear similar in reactive processes such as traumatic (frictional) keratosis or inflammatory conditions such as lichen planus, although the change in keratinisation pattern in these reactive lesions is usually more subtle and not abrupt. Anecdotally, excessive keratinisation (in particular presence of orthokeratinisation) of the floor of the mouth is worrying, particularly when paired with atrophic epithelium.[5] Another useful clue for OED is generalised premature keratinisation within the prickle cell layer.[1,2,4] Where early keratinisation is present, cells in the prickle cell or even the basal cell layer will become larger with a more eosinophilic cytoplasm (Fig. 4B,D, 6C). This feature is relative, as some sites in the oral cavity are more keratinised than others, making comparison with the adjacent normal epithelium crucial. Keratin pearl formation in the rete processes (Fig. 4D) is a particularly alarming feature of OED[1,3,4] as it closely mimics the appearance of the invasive squamous cell carcinoma.[34] Keratin pearls appear as an eosinophilic acellular collection of keratin with no connection to the surface. Some authors have suggested that requiring keratin pearls to be within the rete processes is inappropriate and keratin pearls anywhere in the epithelium should be considered concerning.[3] Abnormal keratinisation may also take the form of single cell keratinisation (also referred to as dyskeratosis) (Fig. 6C), though it is listed as a cytological feature.[1] It is usually easily identified, as dyskeratotic keratinocytes have

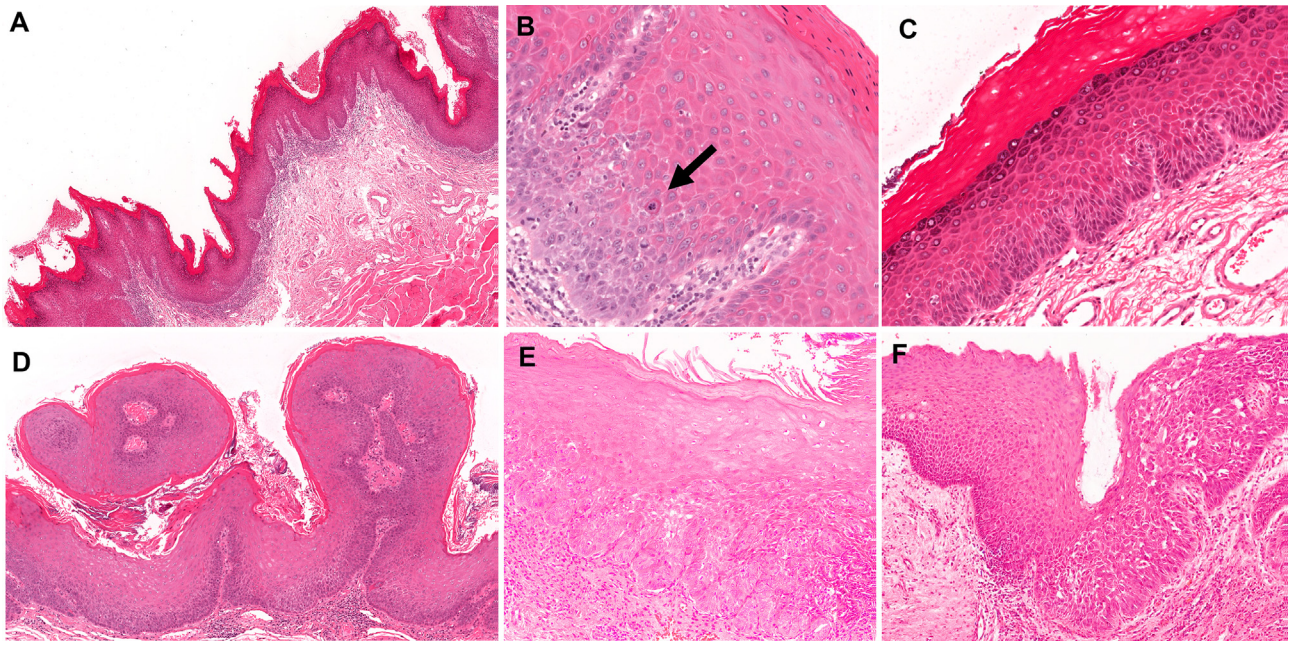

Fig. 5 Further examples of the range of architectural abnormalities seen in OED. (A) Verrucous surface architecture: in this example the epithelium is hyperplastic and orthokeratinised with sharp and pointed epithelial projections. (B) Superficial mitosis (black arrow): a mitotic figure is identified outside of basal compartment. There is no consensus on the number required to arrive at a diagnosis of dysplasia. (C) Bulbous rete processes and generalised premature keratinisation: this example demonstrates rete processes that are wider at the base than at the apex giving a bulbous and drop-shaped architecture. Cells within the spinous layer also show premature keratinisation with deeply eosinophilic cytoplasm, a feature that should only be seen in the surface keratin layer. (D) Papillary architecture: the epithelium is folded into exophytic projections with central fibrovascular connective tissue cores. Basal cell hyperplasia and nuclear and cellular atypia can be seen in the basal compartment of this example. (E) Basal cell nesting and clustering: the basal compartment of the epithelium shows the pleomorphic cells to be tightly packed together in small nests. This feature is better described and more frequently seen in skin biopsies. (F) Multiple patterns of dysplasia: within one specimen it is possible to see more than one pattern of dysplasia. Here we can see more 'conventional' dysplastic changes in the epithelium on the left side of the image with basal cell crowding, hyperchromatism and nuclear and cellular pleomorphism observed. Towards the right side of the image the epithelium shows marked loss of keratinocyte adhesion (acantholysis) alongside basal cell hyperplasia and nuclear and cellular pleomorphism.

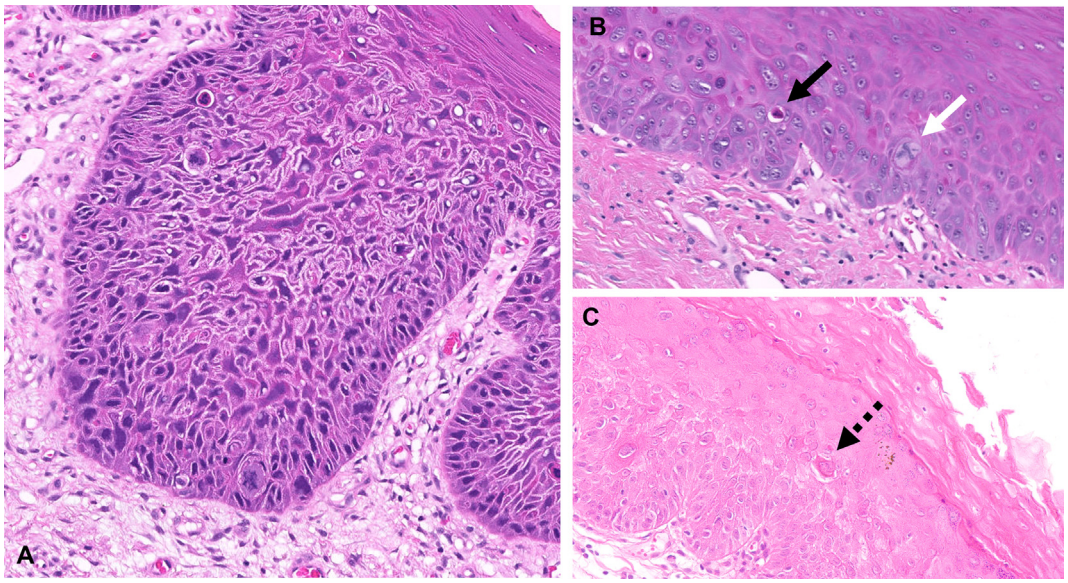

Fig. 6 Cytological abnormalities seen in OED. (A) Severe OED: an abundance of cytological features of OED are seen in this bulbous rete process including abnormal variation in nuclear size and shape, abnormal variation in cell size and shape, increased mitotic activity, increased nucleus:cytoplasm ratio, atypical mitotic figures, increased number and size of nucleoli, nuclear hyperchromasia and abnormal mitoses. (B) Abnormal mitosis and apoptosis: an apoptotic cell with pyknotic nucleus, brightly eosinophilic cytoplasm and retraction from neighbouring keratinocytes (black arrow); an abnormal mitotic figure with asymmetrical chromatin (white arrow). (C) Single cell keratinisation: there is generalised premature keratinisation seen in this example of OED giving the epithelium a strongly eosinophilic appearance. Black dotted arrow highlights single cell keratinisation with even more eosinophilic cytoplasm and retraction from adjacent keratinocytes.

a strongly eosinophilic cytoplasm and a clear halo around the cell, due to retraction from the neighbouring keratinocytes. However, if only a single cell is independently keratinising it is difficult to ascribe this to dysplasia; therefore, multiple instances of single cell keratinisation usually need to be identified alongside other features of OED.

Loss of epithelial cell cohesion

A loss of cohesion between keratinocytes (Fig. 4E) appears to be important in the prediction of malignant transformation risk and recurrence,[22] and has been described in OED for a long time.[1–3,5] This feature initially appears as a widening of the space between keratinocytes before leading to complete separation (acantholysis) of epithelial cells (Fig. 4E). Acantholysis is also seen in vesiculobullous diseases (i.e., pemphigus vulgaris). However, acantholysis is much more pronounced in vesiculobullous disease than OED and does not tend to show atypia. Contact inhibition of epithelial cell locomotion and proliferation through intercellular adhesion complexes such as E-cadherin is well described.[35] Loss of cohesion may give OED a proliferative advantage through loss of contact inhibition.

Verrucous and papillary architecture

A verrucous or papillary surface architecture on its own may be enough to make a diagnosis of dysplasia as these changes are highly abnormal for oral mucosa. Often these patterns, though dysplastic, may have only subtle cytological atypia.[3,5] A verrucous architecture is characterised by a hyperkeratinised surface composed of sharp or blunt epithelial projections with keratin-filled invaginations without fibrovascular cores[36,37] (Fig. 5A). A verrucous pattern may be seen in the keratin alone with the epithelial rete processes maintaining their normal shape. However, in some cases the epithelial morphology may also be affected, leading to an undulated appearance with broad and somewhat 'pushing' rete processes. These changes may be seen in proliferative verrucous leukoplakia (PVL) which will be discussed later.[1,38] A papillary pattern comprises projections of the epithelium supported by fibrovascular connective tissue cores (Fig. 5D). Though both of these patterns are features of dysplasia they may be present in other benign diseases of the oral cavity.[21] Attention must also be paid to the site of the specimen as the filiform papillae of the dorsal tongue may superficially resemble a verrucous surface architecture, and the fungiform papillae show a papillary pattern (Fig. 3C,D,E).[21] Similarly, it is important to be mindful of squamous papilloma like lesions in unusual locations (i.e., gingivae) and clinicopathological correlation is needed to rule out a verrucous lesion.

Abrupt transition, multifocal OED and multiple patterns of OED

The way that the architectural and cytological features manifest can be suggestive of OED. An abrupt transition from normal epithelium to abnormal is very telling of a clonal population and a feature highly suggestive of OED (Fig. 4C).[5,38] It is a recently added feature to the WHO classification.[1] Reactive atypia tends to taper off with increasing distance from the insult, rather than having an abrupt border.

Another feature recently included in the WHO classification is the presence of several different patterns of OED in one lesion (Fig. 5F).[1,4] This is suggestive of several competing clonal keratinocyte populations. Histologically, this is seen as regions of dysplasia each with a distinct collection of atypical features. For example, one area may be basaloid with obvious cytological atypia, whereas another area may show abnormal keratinisation and bulbous rete processes with limited cytological atypia. These regions may be adjacent to each other or separated by areas of normal epithelium. Another recently proposed feature is the presence of multiple epithelial lesions with a consistent pattern of OED separated by regions of normal oral epithelium.[1,4,5]

CYTOLOGICAL FEATURES OF OED

Most OED will display cytological changes, alongside architectural changes (Table 1). In normal oral epithelium, there is little variation in the appearance of the keratinocytes within each epithelial layer as each basal cell will be a similar size, shape, and colour to another basal cell. A well reported cytological feature of OED is pleomorphism.[1,2] This appears as variation in the shape and size of the keratinocytes and their nuclei[1] (Fig. 6A). It is useful to compare the suspected dysplastic keratinocytes to normal keratinocytes within the same epithelial layer when assessing pleomorphism. This accounts for variation in cell morphology over the maturing layers of the epithelium, where basal cells tend to be cuboidal, prickle cells polygonal, and granular cells flattened.[23] There is no defined degree of acceptable variation in normal epithelium, making assessment of pleomorphism highly subjective.

Nuclear variation accounts for several of the cytological features of OED and has many forms[1] (Table 1, Fig. 6). The nucleus may be hyperchromatic, appearing darker, due to increased chromatin content. Variation in nuclear haematoxylin staining within and between laboratories makes hyperchromatism difficult to interpret. Amongst all OED features, hyperchromasia has the lowest interobserver agreement,[22] but comparison with adjacent normal epithelium is useful to account for staining variation. There may be an increase in the nucleus size, either as an absolute increase or an increase compared to the whole cell size (increased N:C ratio).[1] However, both animal and human pathological studies have questioned the use of nuclear to cytoplasmic ratio as a feature of dysplasia.[39,40] Usually, an increased N:C ratio is due to an increase in nuclear size, so having both as separate criteria is likely superfluous.[3] Pleomorphism of nuclei is also an important feature, but similar to cellular pleomorphism is poorly defined. Both hyperchromatism and nuclear pleomorphism are features which have recently been correlated with risk of malignant transformation and recurrence of OED.[22] An increase in the number and size of nucleoli can also be seen in OED, though these changes may also manifest in reactive epithelium and are again poorly defined.

Atypical mitoses, regardless of location, are a feature of OED.[1,4,5] There are many forms of abnormal mitoses including: aneuploid mitoses (tripolar mitoses), asymmetrical mitosis, chromosome bridging and chromosome lagging[41] (Fig. 6; Supplementary Fig. 1, Appendix A). An abnormal mitosis suggests a cell has abnormal chromosome numbers, a failure of chromosome segregation or a general failure of DNA replication. Abnormal mitosis likely leads to further acquisition of oncogenic mutations.[42] Unfortunately there is no evidence to suggest how many abnormal mitoses are required before being ascribed as a diagnostic feature. A recent addition to the diagnostic criteria is the presence of apoptotic mitoses.[1,4] These are also known as mitotic

catastrophes, are generally recognised within the spectrum of abnormal mitoses, and are thought to be indicative of chromosomal instability.[4] To standardise the diagnosis and keep the list of OED features concise and clear, it would be best for this to be considered under the broader category of abnormal mitotic figures.

HPV-RELATED ORAL DYSPLASIA

It is estimated that 27% of OED harbour HPV infection with the majority having high-risk HPV (HPV16, 18).[9] These lesions have distinct histological features when compared with OED caused by traditional risk factors such as tobacco and alcohol consumption. The discussion so far has been focused on OED caused by traditional risk factors and the features discussed here (though not all are specific to HPV-related OED) are clues that further testing is warranted. HPV-related OEDs often have extensive architectural and cytological changes.[43] The keratinocytes are basaloid due to a high nuclear to cytoplasmic ratio and the epithelium is often parakeratinised, with the keratin being densely eosinophilic (Supplementary Fig. 2, Appendix A).[1,4,43] Karyorrhectic keratinocytes known as mitosoid bodies and apoptotic keratinocytes are characteristic of HPV-related OED.[1,4,43] A mitosoid body has condensed nuclear chromatin and a pericellular halo, giving the appearance of a mitotic figure.[43] Apoptotic keratinocytes will have a strongly eosinophilic cytoplasm with a pyknotic nucleus[43] (Supplementary Fig. 2, Appendix A). Though apoptotic cells may be seen in OED, the above-mentioned features are not common in conventional OED and can be helpful in distinguishing HPV-related OED from other causes.[43] In summary, if OED is generally basophilic with bright parakeratin, mitosoid bodies and apoptosis, HPV should be suspected as the causative factor rather than the traditional risk factors.

Ultimately, as HPV infection in the oral cavity is uncommon, it requires confirmation by a combination of p16 immunohistochemistry which is positive in 62% of HPV-infected cases,[9] and the detection of high-risk HPV by polymerase chain reaction (PCR) or *in situ* hybridisation (Supplementary Fig. 2D, Appendix A).[1] Almost 20% of HPV-negative OED can show p16 staining and up to 40% of HPV-positive OED are p16 negative, making p16 staining alone an unreliable surrogate for HPV infection in OED.[9] At present there is limited research into the outcomes of these lesions; as such, they should be graded and managed the same as conventional OED lesions.[1] The malignant transformation rate of these lesions has been reported to be approximately 15%.[43]

DIFFERENTIATED DYSPLASIA

Differentiated dysplasias present an area of diagnostic difficulty, as they tend to lack cytological atypia.[4,33] Despite this lack of cytological change, some suggest they account for a significant portion of OED which transforms to OSCC.[33] They have also been referred to as architectural dysplasias[4] and are often considered a subtype of dysplasia rather than a distinct entity.[4,33] Although there is little consensus on what features constitute a differentiated dysplasia, they can include hyperkeratosis in the presence of an atrophic epithelium, abrupt change when compared with the adjacent mucosa, multiple skip lesions,[32] premature keratinisation and loss of epithelial cell cohesion.[33,44] The requirement for cytological atypia to be present is debated and most classify lesions with similar features but a verrucous architecture as distinct.[4,32–44] Practically, the presence of an unusual pattern of keratin for the oral subsite being examined and an abrupt transition are incredibly useful clues for the presence of differentiated OED, even in the absence of any cytological atypia (Supplementary Fig. 3, Appendix A). However, clinicopathological correlation is important as similar changes are seen in the early stages of PVL.[1,32] Grading of these lesions is difficult as the thickness of the epithelium affected by atypia is often limited, with very few lesions having full thickness dysplasia despite undergoing malignant transformation.[33] As such, a good 'rule of thumb' is to assign the lesion one grade higher than it would be given based on traditional 'thirds' assessment, and clinical follow-up should also be recommended.

PROLIFERATIVE VERRUCOUS LEUKOPLAKIA

PVLs can only be diagnosed by careful clinicopathological correlation.[45] They are distinct entities from OED with a high risk of transformation to OSCC.[1] PVL is characterised clinically by flat or verrucous leucoplakias, a long clinical history, multiple lesions and a predilection for the gingivae (Supplementary Fig. 4A,B, Appendix A).[1,45,46] The histological features are variable and broadly separated into the categories 'hyperkeratotic lesions' and 'lesions with epithelial proliferation' (Supplementary Fig. 4C,D, Appendix A).[1,45] Although these two categories are often considered as early and late phases of the disease,[1] there is no evidence to suggest one category progresses to the other.[45] The hyperkeratotic lesions show hyperkeratosis possibly with a corrugated, or verrucous surface.[1,45] These changes may be accompanied by a verrucous morphology in the epithelium.[1,45] Alternatively, the verrucous pattern might be absent and there is hyperkeratosis with epithelial atrophy and an abrupt transition[45] in a similar pattern to differentiated dysplasia.[4,32] The 'proliferative epithelium' category often shows less significant keratinisation, but has bulky epithelial hyperplasia, often with both exophytic and endophytic expansion.[1,45] The rete processes are usually broad and may coalesce, and a lichenoid pattern of inflammation is also commonly encountered.[45] In both categories cytological atypia is often limited.[45,46] Distinguishing the proliferative epithelium category from verrucous OSCC is challenging. Extension of the lesion deeper than the adjacent epithelium is suggestive of transformation to OSCC,[1] but this may be difficult to assess or may not be always present.[45] Due to the histological similarities between OED and PVL, communication between the pathologist and the clinician managing the patient is essential where PVL is suspected by either party.

MIMICS OF OED

The features of OED can overlap with several other entities, primarily inflammatory diseases where the epithelium displays so called 'reactive atypia'.[38] In these instances, there is limited or no risk of malignant transformation, and the atypia will resolve with removal of the cause. A diagnosis of OED given to these mimics will lead to over-treatment, whereas failure to diagnose true OED in a lesion mimicking a reactive process will lead to under-treatment. It is often difficult in these contexts to distinguish mild OED from these reactive changes.

Reactive atypia is commonly seen in oral candida infections;[5] however, there are usually features of candida infection as well, and fungal hyphae visible with special stains such as periodic acid–Schiff (Supplementary Fig. 5, Appendix A). Testing on multiple tissue levels may be useful. However, it is important to be mindful of the fact that candida can get trapped on the surface of dysplastic lesions and the atypia present may not always be reactive. There is also a correlation between a diagnosis of OED and the presence of candida,[47,48] though this complex relationship and how one impacts the other is not fully understood. In difficult cases where the atypia is deemed too prominent to confidently call it reactive, clinicopathological correlation becomes vital. A useful tool is to defer the decision to a second later biopsy soon after treatment with antifungal therapy.[4,5,38] Such treatment will resolve or reduce fungal-related reactive atypia, whereas changes related to OED will persist.

Oral lichen planus (OLP) and oral lichenoid tissue reactions (LTR) are inflammatory diseases of the oral cavity. Both can display features which overlap with OED clinically and histologically[49] (Supplementary Fig. 6, Appendix A). OLP and LTR are characterised by a lichenoid pattern of inflammation. Unfortunately, OED and OSCC often have lichenoid inflammation and other features of OLP/LTR in up to 29% of cases, particularly in mild and moderate OED lesions.[1,4,24,50] Lichenoid dysplasia (OED with features of OLP) has been suggested as a distinct entity,[51] although this is not widely accepted, with recent molecular studies demonstrating significant transcriptional overlap between OLP and lichenoid dysplasia.[52] Ultimately, despite the presence of a lichenoid infiltrate or other features of OLP or LTR, if any features of dysplasia are seen, a diagnosis of OED should be given, as OLP/LTR should not show any 'true' features of dysplasia.[50]

Ulcers are common in the oral cavity and are frequently biopsied to rule out the possibility of OED or OSCC. Supplementary Fig. 7 (Appendix A) shows an example of reactive atypia in an ulcer, while Supplementary Fig. 8 (Appendix A) shows true OED with ulceration. Although there are limited studies which aim to distinguish reactive atypia from true atypia,[38] a useful clue is resolution of atypia with distance from the area of ulceration. This is a feature usually seen in reactive atypia; conversely, OED atypia will persist beyond the epithelium immediately adjacent to the ulcer.

Other mimics of OED include hyperkeratosis and epithelial hyperplasia (HK+EH) (Supplementary Fig. 9, Appendix A) which may have overlapping architectural features with OED such as altered keratinisation,[5,53] oral hairy leukoplakia (OHL) (Supplementary Fig. 10, Appendix A) which is driven by Epstein–Barr virus (EBV) infection[54] and multifocal epithelial hyperplasia which is cause by HPV13 and HPV32[1,55] (Supplementary Fig. 11, Appendix A).

HK+EH may be particularly difficult to distinguish from mild OED. Both usually present clinically as a leukoplakia and they may show similar histological architectural changes.[5] HK+EH should show hyperkeratosis associated with acanthosis, basal cell hyperplasia, and intracellular oedema[53] with an absence of disordered stratification and basal cell pleomorphism, though some basal cell hyperchromatism is acceptable.[38] It may be challenging to distinguish the basal cell hyperplasia seen in HK+EH from mild cytological atypia. The context of the lesion and other features suggestive of trauma are helpful in this scenario. In HK+EH the underlying lamina propria is often more densely collagenous in reaction to the traumatic causes of these lesions, whereas this may be absent in OED. Another consideration is the site of lesion, for example buccal mucosa and lateral tongue are often exposed to trauma, making them frequent sites for HK+EH. However, if thick keratin is seen, particularly in the absence of other traumatic features (epithelial acanthosis and oedema, denser extracellular matrix in the lamina propria) in a site not prone to trauma such as the floor of mouth or ventral tongue, dysplasia must be highly suspected, even where limited cytological atypia is seen.

GRADING ORAL EPITHELIAL DYSPLASIA

Grading of OED was first introduced in 1969[56] with the purpose of stratifying lesions based on risk of malignant transformation. Ironically, the value of histological grading in prediction of cancer risk has been somewhat limited due to subjectivity and lack of reproducibility.[57,58] Numerous grading classifications have been proposed, amongst which the WHO criteria remains the most widely accepted system.[59] This system originally had five tiers but now has three (mild, moderate and severe dysplasia). This simplification aimed to reduce bias and simplify management by grouping 'severe dysplasia' and 'carcinoma *in situ*' into the 'severe' category, and hyperplasia and mild OED into the 'mild' category.[59] This method is in part based on analysis of epithelial 'thirds' i.e., ascribing a grade based on the collective appearances of a wide range of features and their location (or height) within the epithelium. In mild dysplasia, the dysplastic changes are confined to the basal and parabasal layers, whereas in moderate dysplasia, the changes extend to the middle third of the epithelium, and in severe dysplasia, the changes extend through the entire thickness (or more than half) of the epithelium. However, this method is unreliable, and arguably over-simplistic. Furthermore, the presence of a single feature in abundance, irrespective of its location in the epithelium, may be sufficient to upgrade a lesion. The WHO grading system historically has also not been fit for grading verrucous lesions or differentiated dysplasia where marked architectural changes may arise without significant cytological atypia. This ambiguity and subjectivity results in wide inter- and intra-observer variability,[14–17] and consequently may lead to inaccurate diagnosis and inadequate management. However, this problem is not specific to oral dysplasia, and extends to other parts of the body including cervical intraepithelial neoplasia,[60] vulvar intraepithelial neoplasia[61] and Barrett's oesophagus.[62]

An alternative binary grading system was proposed in 2006 with the aim of increasing diagnostic reproducibility.[13] This classification graded lesions based on the total number of histological features ('low' risk: <four architectural features, < five cytological features; 'high' risk: ≥ four architectural features, ≥ five cytological features). A recent systematic review and meta-analysis, however, comparing this system with the WHO classification showed inconclusive results with regards to its prognostic value.[63] This system does not consider verrucous lesions or the extent of features present. As such, the binary system is not deemed robust enough for routine clinical use at present.

Historically, no prognostic weight had been ascribed to one feature over another. However, a recent study has indicated

that certain features may be associated with an increased risk of progression to OSCC, including bulbous rete processes, hyperchromatism, loss of epithelial cohesion, loss of stratification, suprabasal mitoses and nuclear pleomorphism.[22] These six features were also statistically associated with OED recurrence, in addition to dyskeratosis. In this study, two prognostic scoring models were developed and tested. The first, a 'six-point' model allocated one point for the presence of each of the six OED features which were associated with a greater incidence of transformation and recurrence. Using this model, a score of '4–6 points' produced the highest risk of malignant transformation and recurrence at five years, estimated at 38% and 49%, respectively. This model demonstrated greater prognostic performance than that achieved by the WHO (2017) grading system for both transformation and recurrence, but only a marginal improvement over binary grading.[22] The second 'two-point' model allocated a point for each of the two features that had the highest inter-rater agreement and were also associated with transformation and recurrence (loss of epithelial cohesion and bulbous/drop shaped rete pegs). The presence of both features was associated with an increased risk of malignant transformation at 5 years, in comparison to each single feature in isolation.[22] This study also evaluated the individual prognostic relationships of less conventional but commonly observed features of OED, including verrucous architecture, lymphocytic band (lichenoid-like inflammatory infiltrate) and abrupt orthokeratosis. Whilst the inter-observer agreement for these features was better (Cohen's kappa 0.60–0.73) than other conventional features, interestingly, none of these were associated with malignant transformation or OED recurrence.

Several suggestions have been proposed to overcome the reliance on grading, such as the use of molecular markers,[64–67] morphological descriptors[68] and computer-aided analyses.[69] The latter has seen a surge of interest, particularly with the increasing ubiquity of digital slide scanners in pathology laboratories. Various image analysis platforms have been developed which allow for automated cell nuclei detection, extensive feature evaluation and quantitative approaches for more objective histological and morphometrical feature analysis. Machine learning, a branch of AI, has been shown to reduce variability in classification of precancerous and cancerous lesions by ensuring standardisation and providing quantifiable outputs for risk stratification.[70,71] However, further research is needed to correlate histological features with OED progression to malignancy and to discover novel digital markers important in prognostication. This may support the development of new and improved prognostic models to assist with clinical decision making. However, no molecular, digital or histological features (singly or combined) have been well correlated with malignant transformation in prospective studies. Due to this lack of evidence, though the features of dysplasia have been extensively described, listed and here defined, OED is still poorly understood.

CONCLUSION

The diagnosis of OED is complicated by the great variety of features, most of which are poorly described and have limited support by good quality evidence. Therefore, it is not surprising that OED diagnosis and grading can show such significant inter- and intra-observer variations. The many features of OED must be interpreted in the context of other factors including the extent of any one feature, the presence of inflammatory disease which may mask or enhance the changes seen, and the clinical scenario in which the lesions have arisen. Strict definitions do not exist for most of the features listed in the WHO criteria, but it is hoped the descriptions provided in this paper will help improve the understanding of these features. Digital whole slide images of all figures have been provided to aid the reader in their understanding of the range of features which may be seen in OED (https://www.pathogenesis.co.uk/r/demystifying-dysplasia-histology-dataset). Efforts should be made to create clear definitions for all features to aid diagnosis, training and future research. There is great potential for automation and objective quantitative assessment of histological features using digital and computational methods. When properly assessed in real-world clinical settings, such approaches may assist decision making and improve patient management by yielding more reliable prognostic information to aid risk stratification. Future directions should be to work closely with the WHO and stakeholders to simplify the current criteria, unify understanding, and study the possibility of quantitatively using OED features as prognostic indicators.

Ethics approval and consent to participate: Ethical approval is in place for use of the histological images presented in this article (West Midlands - Edgbaston Research Ethics Committee, reference: 18/WM/0335).

Data availability: All data generated or analysed during this study are included in this published article. The whole slide images examples have also been shared as a cloud-based open access dataset. https://www.pathogenesis.co.uk/r/demystifying-dysplasia-histology-dataset.

Conflicts of interest and sources of funding: SAK is one of the inventors of the Pathogenesis Digital Pathology Platform (pathogenesis.co.uk) which has been used to create the whole slide image collection, and is a share holder in Histofy, an AI startup. Pathogenesis is a licensed platform developed at the University of Sheffield for pathology learning and education. HM is funded by the National Institute for Health and Care Research (Award ID: NIHR300904). PH is partially funded by a pre-doctoral bursary from Cancer Research UK and the Pathological Society of Great Britain and Ireland (reference: RCCPSB-Nov21\100001). SAK is partially funded by a Cancer Research UK Project Grant (reference: C63489/A29674). The authors state there are no other conflicts of interest to disclose.

APPENDIX A. SUPPLEMENTARY DATA

Supplementary data to this article can be found online at https://doi.org/10.1016/j.pathol.2023.10.002.

Address for correspondence: S.A. Kurman, Unit of Oral and Maxillofacial Pathology School of Clinical Dentistry, University of Sheffield, Sheffield, S10 2TN, UK. E-mail: s.a.khurram@sheffield.ac.uk

References

1. WHO Classification of Tumours Editorial Board. *Head and Neck Tumours*. 5th ed. Lyon: IARC, 2023 (beta version ahead of print).

2. Katz H, Shear M, Altini M. A critical evaluation of epithelial dysplasia in oral mucosal lesions using the Smith-Pindborg method of standardization. *J Oral Pathol Med* 1985; 14: 476–82.
3. Tilakaratne WM, Jayasooriya PR, Jayasuriya NS, De Silva RK. Oral epithelial dysplasia: causes, quantification, prognosis, and management challenges. *Periodontology* 2000 2019; 80: 126–47.
4. Odell E, Kujan O, Warnakulasuriya S, Sloan P. Oral epithelial dysplasia: recognition, grading and clinical significance. *Oral Dis* 2021; 27: 1947–76.
5. Woo SB. Oral epithelial dysplasia and premalignancy. *Head Neck Pathol* 2019; 13: 423–39.
6. Hankinson PM, Mohammed-Ali RI, Smith AT, Khurram SA. Malignant transformation in a cohort of patients with oral epithelial dysplasia. *Brit J Oral Max Surg* 2021; 59: 1099–101.
7. Speight PM, Khurram SA, Kujan O. Oral potentially malignant disorders: risk of progression to malignancy. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2018; 125: 612–27.
8. Wood HM, Daly C, Chalkley R, *et al*. The genomic road to invasion-examining the similarities and differences in the genomes of associated oral pre-cancer and cancer samples. *Genome Med* 2017; 9: 53.
9. de la Cour CD, Sperling CD, Belmonte F, Syrjanen S, Verdoodt F, Kjaer SK. Prevalence of human papillomavirus in oral epithelial dysplasia: systematic review and meta-analysis. *Head Neck* 2020; 42: 2975–84.
10. de la Cour CD, Guleria S, Nygard M, *et al*. Human papillomavirus types in cervical high-grade lesions or cancer among Nordic women-potential for prevention. *Cancer Med* 2019; 8: 839–49.
11. Rossi R, Mori M, Lotti T. Actinic keratosis. *Int J Dermatol* 2007; 46: 895–904.
12. Ho MW, Risk JM, Woolgar JA, *et al*. The clinical determinants of malignant transformation in oral epithelial dysplasia. *Oral Oncol* 2012; 48: 969–76.
13. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelia dysplasia for prediction of malignant transformation. *Oral Oncol* 2006; 42: 987–93.
14. Ranganathan K, Kavitha L, Sharada P, *et al*. Intra-observer and inter-observer variability in two grading systems for oral epithelial dysplasia: a multi-centre study in India. *J Oral Pathol Med* 2020; 49: 948–55.
15. Speight PM, Abram TJ, Floriano PN, *et al*. Interobserver agreement in dysplasia grading: toward an enhanced gold standard for clinical pathology trials. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015; 120: 474–82.
16. Abbey LM, Kaugars GE, Gunsolley JC, *et al*. Intraexaminer and interexaminer reliability in the diagnosis of oral epithelial dysplasia. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1995; 80: 188–91.
17. Kujan O, Khattab A, Oliver RJ, Roberts SA, 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* 2007; 43: 224–31.
18. Mullin MH, Brierley DJ, Speight PM. Second opinion reporting in head and neck pathology: the pattern of referrals and impact on final diagnosis. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2015; 119: 656–60.
19. Berkovitz BK, Holland GR, Moxham BJ. *Oral Anatomy, Histology and Embryology*. 5th ed. E-book. Elsevier Health Sciences, 2017.
20. Jaber MA, Porter SR, Speight P, Eveson JW, Scully C. Oral epithelial dysplasia: clinical characteristics of western European residents. *Oral Oncol* 2003; 39: 589–96.
21. Thomas GJ, Barrett AW. Papillary and verrucous lesions of the oral mucosa. *Diagn Histopathol* 2009; 15: 279–85.
22. Mahmood H, Bradburn M, Rajpoot N, Islam NM, Kujan O, Khurram SA. Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models. *Mod Pathol* 2022; 35: 1151–9.
23. Squier CA, Finkelstain MW. Oral mucosa. In: Nanci A, editor. *Ten Cate's Oral Histology*. Mosby, 2003; 329–407.
24. Muller S. Oral lichenoid lesions: distinguishing the benign from the deadly. *Mod Pathol* 2017; 30: S54–67.
25. Winning TA, Townsend GC. Oral mucosal embryology and histology. *Clin Dermatol* 2000; 18: 499–511.
26. Schmitz L, Gambichler T, Kost C, *et al*. Cutaneous squamous cell carcinomas are associated with basal proliferating actinic keratoses. *Br J Dermatol* 2019; 180: 916–21.
27. Schmitz L, Gambichler T, Gupta G, *et al*. Actinic keratoses show variable histological basal growth patterns – a proposed classification adjustment. *J Eur Acad Dermatol Venereol* 2018; 32: 745–51.
28. Daley TD, Lovas JGL, Peters E, Wysocki GP, McGaw TW. Salivary gland duct involvement in oral epithelial dysplasia and squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol Endod* 1996; 81: 186–92.
29. Gerdes J, Lemke H, Baisch H, Wacker HH, Schwab U, Stein H. Cell cycle analysis of a cell proliferation-associated human nuclear antigen defined by the monoclonal antibody Ki-67. *J Immunol* 1984; 133: 1710–5.
30. Gonzalez-Moles MA, Ruiz-Avila I, Rodriguez-Archilla A, Martinez-Lara I. Suprabasal expression of Ki-67 antigen as a marker for the presence and severity of oral epithelial dysplasia. *Head Neck* 2000; 22: 658–61.
31. Ribalta T, McCutcheon IE, Aldape KD, Bruner JM, Fuller GN. The mitosis-specific antibody anti-phosphohistone-H3 (PHH3) facilitates rapid reliable grading of meningiomas according to WHO 2000 criteria. *Am J Surg Pathol* 2004; 28: 1532–6.
32. Li CC, Almazrooa S, Carvo I, Salcines A, Woo SB. Architectural alterations in oral epithelial dysplasia are similar in unifocal and proliferative leukoplakia. *Head Neck Pathol* 2021; 15: 443–60.
33. Arsenic R, Kurrer MO. Differentiated dysplasia is a frequent precursor or associated lesion in invasive squamous cell carcinoma of the oral cavity and pharynx. *Virchows Archiv* 2013; 462: 609–17.
34. Speight PM, Farthing PM, Bouquot JE. The pathology of oral cancer and precancer. *Curr Diagn Pathol* 1996; 3: 165–76.
35. Mendonsa AM, Na T-Y, Gumbiner BM. E-cadherin in contact inhibition and cancer. *Oncogene* 2018; 37: 4769–80.
36. Mehrotra D, Goel M, Kumar S, Pandey R, Ram H. Oral verrucous lesions: controversies in diagnosis and management. *J Oral Biol Craniofac Res* 2012; 2: 163–9.
37. Shear M, Pindborg JJ. Verrucous hyperplasia of the oral mucosa. *Cancer* 1980; 46: 1855–62.
38. Müller S. Oral epithelial dysplasia, atypical verrucous lesions and oral potentially malignant disorders: focus on histopathology. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2018; 125: 591–602.
39. White FH, Gohari K. Variations in the nuclear-cytoplasmic ratio during epithelial differentiation in experimental oral carcinogenesis. *J Oral Pathol Med* 1981; 10: 164–72.
40. White FH, Jin Y, Yang L. An evaluation of the role of nuclear cytoplasmic ratios and nuclear volume densities as diagnostic indicators in metaplastic, dysplastic and neoplastic lesions of the human cheek. *Histol Histopathol* 1997; 12: 69–77.
41. Donovan TA, Moore FM, Bertram CA, *et al*. Mitotic figures-normal, atypical, and imposters: a guide to identification. *Vet Pathol* 2021; 58: 243–57.
42. Ganem NJ, Pellman D. Linking abnormal mitosis to the acquisition of DNA damage. *Cell Biol* 2012; 199: 871–81.
43. Lerman MA, Almazrooa S, Lindeman N, Hall D, Villa A, Woo S-B. HPV-16 in a distinct subset of oral epithelial dysplasia. *Mod Pathol* 2017; 30: 1646–54.
44. Kim E, Chung M, Jeong H-S, Baek C-H, Cho J. Histological features of differentiated dysplasia in the oral mucosa: a review of oral invasive squamous cell carcinoma cases diagnosed with benign or low-grade dysplasia on previous biopsies. *Hum Pathol* 2022; 126: 45–54.
45. Thompson LDR, Fitzpatrick SG, Müller S, *et al*. Proliferative verrucous leukoplakia: an expert consensus guideline for standardized assessment and reporting. *Head Neck Pathol* 2021; 15: 572–87.
46. Batsakis JG, Suarez P, El-Naggar AK. Proliferative verrucous leukoplakia and its related lesions. *Oral Oncol* 1999; 35: 354–9.
47. Barrett AW, Kingsmill VJ, Speight PM. The frequency of fungal infection in biopsies of oral mucosal lesions. *Oral Dis* 1998; 4: 26–31.
48. McCullough M, Jaber M, Barrett AW, Bain L, Speight PM, Porter SR. Oral yeast carriage correlates with presence of oral epithelial dysplasia. *Oral Oncol* 2002; 38: 391–3.
49. Cheng Y-SL, Gould A, Kurago Z, Fantasia J, Muller S. Diagnosis of oral lichen planus: a position paper of the American Academy of Oral and Maxillofacial Pathology. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2016; 122: 332–54.
50. Fitzpatrick SG, Honda KS, Sattar A, Hirsch SA. Histologic lichenoid features in oral dysplasia and squamous cell carcinoma. *Oral Surg Oral Med Oral Pathol Oral Radiol* 2014; 117: 511–20.
51. Krutchkoff DJ, Eisenberg E. Lichenoid dysplasia: a distinct histopathologic entity. *Oral Surg Oral Med Oral Pathol* 1985; 60: 308–15.
52. Farah CS, Fox S, Shearston K, Newman L, Babic S, Vacher M. Lichenoid dysplasia is not a distinct pathological entity. *Oral Oncol* 2021; 119: 105362.
53. Müller S. Frictional keratosis, contact keratosis and smokeless tobacco keratosis: features of reactive white lesions of the oral mucosa. *Head Neck Pathol* 2019; 13: 16–24.
54. Greenspan JS, Greenspan D, Webster-Cyriaque J. Hairy leukoplakia; lessons learned: 30-plus years. *Oral Dis* 2016; 22: 120–7.
55. Betz SJ. HPV-related papillary lesions of the oral mucosa: a review. *Head Neck Pathol* 2019; 13: 80–90.

56. Smith C, Pindborg JJ. *Histological Grading of Oral Epithelial Atypia by the Use of Photographic Standards*. Copenhagen: C. Hamburgers Bogtrykkeri A/S, 1969.
57. Wamakulasuriya S. Histological grading of oral epithelial dysplasia: revisited. *J Pathol* 2001; 194: 294–7.
58. Schepman KP, van der Meij EH, Smeele LE, van der Waal I. Malignant transformation of oral leukoplakia: a follow-up study of a hospital-based population of 166 patients with oral leukoplakia from The Netherlands. *Oral Oncol* 1998; 34: 270–5.
59. Reibel J, Gale N, Hille J, *et al*. Oral potentially malignant disorders and oral epithelial dysplasia. In: El-Naggar AK, Chan JKC, Grandis JR, Takata T, Slootweg PJ, editors. *WHO Classification of Head and Neck Tumours*. 4th ed. Lyon: IARC, 2017; 112–4.
60. McCluggage WG, Bharucha H, Caughley LM, *et al*. Interobserver variation in the reporting of cervical colposcopic biopsy specimens: comparison of grading systems. *J Clin Pathol* 1996; 49: 833–5.
61. Preti M, Mezzetti M, Robertson C, Sideri M. Inter-observer variation in histopathological diagnosis and grading of vulvar intraepithelial neoplasia: results of an European collaborative study. *Br J Obstet Gynaecol* 2000; 107: 594–9.
62. Montgomery E, Bronner MP, Goldblum JR, *et al*. Reproducibility of the diagnosis of dysplasia in Barrett esophagus: a reaffirmation. *Hum Pathol* 2001; 32: 368–78.
63. Silva BSD, Batista DCR, Roriz CFD, *et al*. Binary and WHO dysplasia grading systems for the prediction of malignant transformation of oral leukoplakia and erythroplakia: a systematic review and meta-analysis. *Clin Oral Invest* 2021; 25: 4329–40.
64. Shirata NK, Longatto A, Roteli-Martins C, Espoladore LMW, Pittoli JE, Syrjanen K. Applicability of liquid-based cytology to the assessment of DNA content in cervical lesions using static cytometry. *Anal Quant Cytol Histol* 2003; 25: 210–4.
65. Tabor MP, Braakhuis BJM, van der Wal JE, *et al*. Comparative molecular and histological grading of epithelial dysplasia of the oral cavity and the oropharynx. *J Pathol* 2003; 199: 354–60.
66. Rosin MP, Cheng X, Poh C, *et al*. Use of allelic loss to predict malignant risk for low-grade oral epithelial dysplasia. *Clin Cancer Res* 2000; 6: 357–62.
67. Partridge M, Pateromichelakis S, Phillips E, Emilion GG, A'Hern RP, Langdon JD. A case-control study confirms that microsatellite assay can identify patients at risk of developing oral squamous cell carcinoma within a field of cancerization. *Cancer Res* 2000; 60: 3893–8.
68. Abdelsalam M, Mayall BH, Chew K, Silverman S, Greenspan JS. Which oral white lesions will become malignant - an image cytometric study. *Oral Surg Oral Med Oral Pathol* 1990; 69: 345–50.
69. Abu Eid R, Landini G. Oral epithelial dysplasia: can quantifiable morphological features help in the grading dilemma? First Image J User and Developer Conference Proceedings, Luxembourg, 2006.
70. 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: 1934–40.
71. Mahmood H, Shaban M, Indave BI, Santos-Silva AR, Rajpoot N, Khurram SA. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: a systematic review. *Oral Oncol* 2020; 110: 104885.

Chapter 4 – Conventional Histological Feature Analysis

Chapter 4 – Conventional Histological Feature Analysis

The first part of this chapter has been reproduced from a peer-reviewed publication for which the candidate is first and primary contributing author. This article conducts a comprehensive analysis of conventional architectural and cytological histological features in OED to evaluate individual feature prevalence, observer agreement and statistical associations with malignant transformation and OED recurrence. Histological feature-specific prognostic models are developed and tested for OED outcome prediction.

The second part of this chapter takes the form of a publication-ready manuscript for which the candidate is first author. This study conducts validity testing of the proposed histological feature-specific prognostic models that are presented in the preceding study.

The candidate's contributions to these studies were:

i. obtaining ethical approval
ii. slide retrieval and sample preparation
iii. clinical data collection and histological re-grading
iv. working with pathologists to organise feature scoring
v. conducting histological feature scoring (for validity testing study)
vi. statistical analysis, model development and testing
vii. writing up the manuscripts

The study was conceptualised with guidance from supervisors.

Professor Syed Ali Khurram, Dr Nadim Islam and Dr Omar Kujan contributed to histological feature analysis and scoring.

Statistical support was provided by Mike Bradburn, senior statistician at the Clinical Trials Research Unit, University of Sheffield.

4.1 Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models

www.nature.com/modpathol

ARTICLE OPEN

Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models

Hanya Mahmood[1], Mike Bradburn[2], Nasir Rajpoot[3], Nadim Mohammed Islam[4], Omar Kujan[5] and Syed Ali Khurram[6]✉

Oral epithelial dysplasia (OED) is a precursor state usually preceding oral squamous cell carcinoma (OSCC). Histological grading is the current gold standard for OED prognostication but is subjective and variable with unreliable outcome prediction. We explore if individual OED histological features can be used to develop and evaluate prognostic models for malignant transformation and recurrence prediction. Digitised tissue slides for a cohort of 109 OED cases were reviewed by three expert pathologists, where the prevalence and agreement of architectural and cytological histological features was assessed and association with clinical outcomes analysed using Cox proportional hazards regression and Kaplan–Meier curves. Within the cohort, the most prevalent features were basal cell hyperplasia (72%) and irregular surface keratin (60%), and least common were verrucous surface (26%), loss of epithelial cohesion (30%), lymphocytic band and dyskeratosis (34%). Several features were significant for transformation ($p < 0.036$) and recurrence ($p < 0.015$) including bulbous rete pegs, hyperchromatism, loss of epithelial cohesion, loss of stratification, suprabasal mitoses and nuclear pleomorphism. This led us to propose two prognostic scoring systems including a '6-point model' using the six features showing a greater statistical association with transformation and recurrence (bulbous rete pegs, hyperchromatism, loss of epithelial cohesion, loss of stratification, suprabasal mitoses, nuclear pleomorphism) and a 'two-point model' using the two features with highest inter-pathologist agreement (loss of epithelial cohesion and bulbous rete pegs). Both the 'six point' and 'two point' models showed good predictive ability (AUROC ≥ 0.774 for transformation and 0.726 for recurrence) with further improvement when age, gender and histological grade were added. These results demonstrate a correlation between individual OED histological features and prognosis for the first time. The proposed models have the potential to simplify OED grading and aid patient management. Validation on larger multicentre cohorts with prospective analysis is needed to establish their usefulness in clinical practice.

Modern Pathology; https://doi.org/10.1038/s41379-022-01067-x

INTRODUCTION

Oral epithelial dysplasia (OED) is a chronic, progressive precursor epithelial disorder of the oral mucosa, characterised by abnormal maturation and stratification of the surface epithelium[1]. It is associated with a statistically increased risk of progression to oral squamous cell carcinoma (OSCC) which is among the topmost common cancers worldwide and has an increasing incidence and worsening prognosis[2,3]. Clinically, OED most commonly presents as a white patch/plaque (leukoplakia) with up to 50% of biopsied lesions showing dysplasia[4] and malignant transformation rate of 9.5% [99% CI 5.9–14.00%] or 1.56% per year[5]. OED can also be seen in other oral potentially malignant disorders (OPMD), a group of lesions and conditions characterised by an increased risk of malignant transformation, including oral submucous fibrosis, actinic keratosis, erythroplakia and erythroleukoplakia[6,7]. The presence of OED in these disorders increases their risk of malignant transformation[8].

At present, there are no biological or molecular markers proven to be prognostically significant (or in routine diagnostic use) for OED[4]. Histological grading remains the gold standard for predicting malignancy risk and is used to inform patient treatment and prognosis[9]. Over the years, OED grading systems have substantially evolved, and the current World Health Organisation (WHO) classification (2017) grades dysplasia based on the presence of sixteen different histological features[10]. The 'severity' of these features, both in terms of frequency and location in the epithelium, are used to classify lesions into 'low', 'moderate' and 'high' grades, representing an increasing risk for malignant transformation[9]. A recent meta-analysis showed moderate/severe OED to be associated with a greater risk of malignant transformation compared to mild OED with an odds ratio of 2.4 (99% CI 1.5–3.8)[5]. However, it remains unclear which lesions will progress, and which will recur, as the mechanisms for OED

[1]Academic Unit of Oral & Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield, 19 Claremont Crescent, Sheffield S10 2TA, UK. [2]Clinical Trials Research Unit, School of Health and Related Research, University of Sheffield, Sheffield, UK. [3]Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, Coventry, UK. [4]Department of Oral & Maxillofacial Diagnostic Sciences, College of Dentistry, University of Florida, Gainesville, FL, USA. [5]Oral Diagnostic and Surgical Sciences Division, UWA Dental School, The University of Western Australia, Perth, WA, Australia. [6]Unit of Oral & Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, Sheffield, UK. ✉email: s.a.khurram@sheffield.ac.uk

Received: 13 November 2021 Revised: 2 March 2022 Accepted: 3 March 2022
Published online: 31 March 2022

progression are poorly understood[9]. Furthermore, the histological features can individually be considered relatively non-specific, and some (or all) of the features may be seen in different grades of dysplasia, of which some lesions will transform, and others will not (irrespective of grade).

In addition to these issues, there are a number of other problems related to the current grading system[11]. Firstly, there is substantial subjectivity in histological interpretation between pathologists, which can result in wide inter- and intra-observer variability, with potential for an incorrect grade being assigned[12]. This variability can arise since individual features are ill-defined, and this is further complicated by division of the epithelium into 'thirds' which can be challenging. Secondly, grading does not reliably predict prognosis which means that lower grade lesions may progress to OSCC whereas higher grade lesions may remain static[4,10]. Thirdly, several of the established histologic features can also be seen in reactive lesions, such as the margins of ulcers or candida infections. It is accepted that a complex interaction exists between a combination of features including histological atypia, progressive molecular changes and chromosomal derangements to trigger cancer development, but the individual importance of these features in OED progression is not well established[13,14].

More recently, an alternative binary grading system (low/high grade) has been proposed[15]. This system grades dysplasia based on the overall number of cytological and architectural changes observed, and several studies have shown its improved reproducibility, inter-observer agreement and clinical utility as compared to the WHO system[15,16]. Despite these improvements though, neither systems consider the importance of individual histological features, or specify which of the features (in isolation or combination) are of greatest relevance for transformation and recurrence. Some older studies have compared OPMDs that did not transform to lesions that did[17], and others have linked certain histology features to a higher transformation risk[18]. However, conclusions from these studies should be treated with caution due to weaknesses in the proposed methodologies.

The aims of this study are twofold: first, to conduct a detailed histological assessment (and inter-observer agreement) of individual OED features to identify which were most prevalent and associated with a higher risk of malignant transformation and recurrence; second, to develop and propose feature-specific prognostic models for OED outcome prediction. To the best of our knowledge, this is the first study to explore histological feature-specific prognostic prediction of OED.

MATERIALS/SUBJECTS AND METHODS

Case selection, tissue preparation and conversion to digital images

A retrospective sample of sequential OED cases were retrieved between 2008 and 2013 from the Oral and Maxillofacial Pathology archive at the School of Clinical Dentistry (Sheffield, UK) using a local digital database (ethical approval: 18/WM/0335). To confirm cases which had progressed to OSCC at the same clinical site, a regional head and neck cancer (HNC) electronic records system was accessed which is a repository for HNC cases within South Yorkshire. Newly stained 4 µm Haematoxylin and Eosin (H&E) sections of the selected cases were obtained from formalin fixed paraffin embedded blocks and a digital slide scanner (Aperio CS2, Milton Keynes, UK) was used to obtain whole slide images (WSI) at x40 magnification.

Inclusion and exclusion criteria

The principal inclusion criteria were varying grades of OED retrieved from the Sheffield Oral and Maxillofacial Pathology archive with sufficient available tissue and availability of minimum five-year follow-up data. Where multiple biopsies had been taken over a period of follow-up, only the initial biopsy was selected for the study. The unit of Oral and Maxillofacial Pathology at Sheffield is a regional and national referral centre which receives referrals from a wide geographical area, however, following a confirmed tissue diagnosis any necessary treatment is provided by a local core Oral and Maxillofacial team and therefore cases treated outside this unit were by default excluded in this study. Additionally, cases were excluded if there was insufficient tissue for histological analysis, incomplete minimum follow up data or histological evidence of positive tissue margins on the subsequent excision (to avoid any bias in the recurrence data). The H&E slide and clinical records for all selected cases were reviewed by two authors (HM, SAK) to ensure the inclusion criteria was met.

Clinical data collection

Minimum five-year follow-up data was obtained from clinical notes and biopsy forms by HM. Data collection included patient demographics/characteristics (age, gender, intraoral site), histological OED grade and two main clinical outcomes of interest (time to transformation and recurrence). Transformation was defined as a dysplastic lesion which had progressed to OSCC at the same clinical site and within the follow-up period, and recurrence was defined as a dysplastic lesion which occurred again in the same clinical site following active treatment (i.e. surgical excision or laser treatment) within the follow-up period. All data was recorded by HM in a structured proforma using Microsoft Excel (2016) in an anonymised-linked format.

Histological evaluation and examiners

Three experienced oral and maxillofacial pathologists (NMI, OK, SAK) working in different international centres performed independent histological examination of the OED cohort. All pathologists were provided access to the WSIs via a cloud-based system. Each WSI was labelled with an anonymous-linked number, and all pathologists were blinded to the original diagnosis and clinical outcomes. The examiners were asked to independently assess the cases and identify which histological features amongst the WHO criteria were present and informed the diagnosis. They were also encouraged to specify any additional histological features which were considered important in influencing their diagnosis.

To determine which OED features were most prevalent, the examiners were asked to provide a binary score to record the presence (or absence) of individual features; a score of 1 was given if the feature was abundantly visible (and influenced diagnosis), and a score of 0 if the feature was absent or rare/focal. The topmost common histological OED features (as per consensus scoring) were further explored to determine feature-specific observer agreement and prognostic significance. To minimise examiner bias, no formal calibration exercises were attempted, although there was an informal discussion between the examiners to discuss their approach to this task. For consistency and to prevent double counting of similar appearing histological features, the pathologists agreed on general definitions for individual WHO features (as well as other commonly presenting features). For example, basal cell hyperplasia was considered if crowding/proliferation involved 1–2 layers of basal cells, whereas loss of epithelial stratification was considered if there was a disturbance in the organised 'stratified' layers of the epithelium and the layers were haphazardly organised or difficult to separate.

Finally, the original OED histological grades were independently reviewed by HM and where necessary, an updated grade was assigned. A standardised score sheet was designed in Microsoft Excel (2016) to record all examiner scoring and aid systematic analysis. All participating pathologists were clinical-academic pathologists with long-standing experience in the diagnosis of OED and OSCC.

Statistical evaluation

Statistical analyses were conducted using the Stata Statistical Software[19] (Version 17, 2021). The prevalence of OED features was calculated overall and for each examiner. Observer agreement was summarised as the percentage of patients for whom all three examiners agreed, and by two chance-corrected measures (Cohen's Kappa and Gwet's AC), where a value of 1 denotes perfect agreement and 0 relates to no agreement beyond chance alone.

Univariate associations between pathological features and clinical outcomes (transformation and recurrence) were visualised by Kaplan–Meier curves and analysed using a Cox proportional hazards regression model with Efron's correction for tied times. Thereafter, two prognostic models were developed in which the outcome of interest was event (transformation and recurrence) at any time. The prognostic performance of the two models were compared against each other as well as against patient/clinical characteristics (age, gender, intraoral site) and histological OED grade alone by generating the area under the receiver-operator characteristic curve (AUROC). All statistical tests were two-tailed and $p < 0.05$ were considered statistically significant.

RESULTS

Characteristics of the study cohort

151 previously diagnosed cases of OED were retrieved during the study period, of which 42 were excluded due to either insufficient tissue availability or incomplete minimum five-year clinical follow up data. Amongst the patient cohort, 67 (61%) were male and 42 (39%) were female with a median age of 67 years (IQR 57–77). Breakdown based on intraoral site were as follows: tongue 44 (40%), floor of mouth 23 (21%), buccal mucosa 17 (16%), gingivae 7 (6%), soft palate 6 (6%), hard palate 6 (6%) and lower lip 6 (6%). The clinical records showed that 34 (31%) of OED lesions were clinically monitored, 70 (64%) were surgically excised and 5 (5%) were treated with laser.

Prevalence and agreement of OED features

The final study cohort (Table 1) comprised 109 OED cases which were blindly re-evaluated to confirm 34 (31%) mild, 48 (44%) moderate and 27 (25%) severe dysplasia cases. Binary grading of these cases showed 73 (67%) to be low grade and 36 (33%) as high-grade lesions. Table 2 summarises the prevalence and observer agreement for the twelve most prominent OED features that were observed as per consensus scoring. The most common features were basal cell hyperplasia (72%) and irregular surface keratin (60%). The latter feature refers to any irregularity of the keratin layer, including a corrugated, shaggy or desquamative appearance. This feature was included since all pathologists highlighted it as a prominent feature in certain cases, and at present it is not on the list of WHO criteria. The least common were verrucous surface morphology (26%), loss of epithelial cohesion (30%), lymphocytic band (34%) and dyskeratosis (34%). All other features ranged between 36% and 57%.

Verrucous surface morphology had the highest agreement between pathologists (Kappa = 0.73, Gwet's AC1 = 0.83). Gwet's AC1 measurements were comparable for abrupt orthokeratosis (0.66), lymphocytic band (0.67) and loss of epithelial cohesion (0.69). Agreement for all other features was typically modest, with the worst agreement for hyperchromatism (Kappa and Gwet's AC1 both = 0.32) and suprabasal mitoses (Kappa and Gwet's AC1 both = 0.34) for which all three pathologists agreed for approximately half the patients.

Table 1. Characteristics of the study cohort.

Characteristic	Number (%) or median (IQR)
Age	67 (57–77)
Gender	
Female	42 (39%)
Male	67 (61%)
WHO grade	
Mild	34 (31%)
Moderate	48 (44%)
Severe	27 (25%)
Binary grade	
Low	73 (67%)
High	36 (33%)
Site of disease	
Tongue	44 (40%)
Floor of mouth	23 (21%)
Buccal mucosa	17 (16%)
Gingivae	7 (6%)
Hard palate	6 (6%)
Lower lip	6 (6%)
Soft palate	6 (6%)

Table 2. Observer agreement for OED feature analysis.

Prominent OED Features	Overall prevalence*	Complete agreement	Cohen's Kappa	Gwet's AC1
Basal cell hyperplasia	236 (72%)	63 (58%)	0.30	0.53
Bulbous/drop shaped rete pegs	187 (57%)	72 (66%)	0.54	0.56
Dyskeratosis	110 (34%)	68 (62%)	0.44	0.55
Hyperchromatism	176 (54%)	53 (49%)	0.32	0.32
Irregular surface keratin	196 (60%)	68 (62%)	0.48	0.52
Loss of epithelial cohesion	98 (30%)	80 (73%)	0.58	0.69
Loss of stratification	138 (42%)	61 (56%)	0.41	0.43
Suprabasal mitoses	148 (45%)	54 (50%)	0.34	0.34
Nuclear pleomorphism	118 (36%)	62 (57%)	0.38	0.47
Abrupt orthokeratosis	174 (53%)	81 (74%)	0.66	0.66
Lymphocytic band	112 (34%)	79 (72%)	0.60	0.67
Verrucous surface	85 (26%)	92 (84%)	0.73	0.83

*Denominator for overall prevalence is the number of assessments (327; 109 patients each with 3 assessments). Complete agreement is the percentage of patients (out of 109) where all three assessors agreed.

Table 3. Incidence of transformation and recurrence by OED feature.

Overall	Transformation 20 (18%)		Recurrence 27 (25%)	
Prominent OED Features	**Positive**	**Negative**	**Positive**	**Negative**
Basal cell hyperplasia	15 (18%)	5 (20%)	19 (23%)	8 (32%)
Bulbous/drop shaped rete pegs	18 (30%)	2 (4%)	20 (33%)	7 (14%)
Dyskeratosis	8 (24%)	12 (16%)	12 (36%)	15 (20%)
Hyperchromatism	15 (26%)	5 (10%)	20 (34%)	7 (14%)
Irregular surface keratin	10 (15%)	10 (24%)	16 (24%)	11 (26%)
Loss of epithelial cohesion	11 (35%)	9 (12%)	14 (45%)	13 (17%)
Loss of stratification	15 (34%)	5 (8%)	19 (43%)	8 (12%)
Suprabasal mitoses	14 (27%)	6 (10%)	19 (37%)	8 (14%)
Nuclear pleomorphism	13 (32%)	7 (10%)	17 (41%)	10 (15%)
Abrupt orthokeratosis	10 (17%)	10 (20%)	14 (23%)	13 (27%)
Lymphocytic band	9 (25%)	11 (15%)	12 (33%)	15 (21%)
Verrucous surface	6 (20%)	14 (18%)	8 (27%)	19 (24%)

For each feature, a consensus definition was used whereby the feature was assumed to be present if 2/3 observers rated it as being prominent, otherwise it was assumed absent.

Table 4. Hazard ratios and *p* values of individual OED features for their time to malignant transformation and recurrence.

Prominent OED features	Transformation		Recurrence	
	Hazard ratio	***p* value**	**Hazard ratio**	***p* value**
Basal cell hyperplasia	0.88 (95% CI 0.32, 2.42)	0.806	0.65 (95% CI 0.29, 1.49)	0.310
Bulbous rete pegs	8.27 (95% CI 1.92, 35.68)	0.005*	2.52 (95% CI 1.06, 5.96)	0.036*
Dyskeratosis	1.68 (95% CI 0.69, 4.11)	0.257	2.20 (95% CI 1.03, 4.70)	0.042*
Hyperchromatism	2.96 (95% CI 1.08, 8.15)	0.036*	2.90 (95% CI 1.23, 6.86)	0.015*
Irregular surface keratin	0.62 (95% CI 0.26, 1.49)	0.286	0.92 (95% CI 0.43, 1.99)	0.841
Loss of epithelial cohesion	3.78 (95% CI 1.57, 9.14)	0.003*	3.50 (95% CI 1.64, 7.46)	0.001*
Loss of stratification	5.35 (95% CI 1.94, 14.73)	0.001*	4.50 (95% CI 1.97, 10.30)	0.000*
Suprabasal mitoses	3.06 (95% CI 1.17, 7.96)	0.022*	3.17 (95% CI 1.39, 7.24)	0.006*
Nuclear pleomorphism	3.74 (95% CI 1.49, 9.38)	0.005*	3.45 (95% CI 1.58, 7.54)	0.002*
Abrupt orthokeratosis	0.78 (95% CI 0.32, 1.87)	0.572	0.85 (95% CI 0.40, 1.81)	0.680
Lymphocytic band	1.80 (95% CI 0.75, 4.35)	0.191	1.74 (95% CI 0.82, 3.73)	0.151
Verrucous surface	1.09 (95% CI 0.42, 2.85)	0.855	1.11 (95% CI 0.49, 2.53)	0.807

*Denotes a statistically significant finding.

OED feature-specific incidence of transformation and recurrence

Table 3 summarises feature-specific incidence for transformation and recurrence. Overall, 20 (18%) OED lesions transformed, and 27 (25%) lesions recurred following treatment. A higher incidence of transformation was seen when bulbous/drop shaped rete pegs (30%), loss of epithelial cohesion (35%), loss of stratification (34%) and nuclear pleomorphism (32%) were observed. The incidence of recurrence was also higher related to these same four features, as well as suprabasal mitoses (37%) and nuclear pleomorphism (41%).

Feature-specific correlation to clinical outcomes

Table 4 summarises the hazard ratios and *p* values of individual OED features for their time to the two clinical outcomes of interest (malignant transformation and recurrence). Six features were associated with a greater rate of transformation: bulbous/drop shaped rete pegs ($p = 0.005$) hyperchromatism ($p = 0.036$), loss of epithelial cohesion ($p = 0.003$), loss of stratification ($p = 0.001$), suprabasal mitoses ($p = 0.022$) and nuclear pleomorphism ($p = 0.005$).

These same six features (bulbous/drop shaped rete pegs $p = 0.036$, hyperchromatism $p = 0.015$, loss of epithelial cohesion $p = 0.001$, loss of stratification $p < 0.001$, suprabasal mitoses $p = 0.006$, nuclear pleomorphism $p = 0.002$), in addition to dyskeratosis ($p = 0.042$), were also positively associated with recurrence.

Proposed prognostic models for OED

Two prognostic models were explored to assess the potential for reliably predicting clinical outcomes of OED. In all cases, the number of covariates was minimised to limit the impact of overfitting.

Prognostic model 1: Six-point scoring system. The first scoring system allocated one point for the presence of each of the six OED features which were associated with a greater incidence of transformation and recurrence (bulbous/drop shaped rete pegs, hyperchromatism, loss of epithelial cohesion, loss of stratification, suprabasal mitoses, nuclear pleomorphism). Since the hazard ratios for these features (Table 4) are reasonably similar, each feature is allocated equal weight.

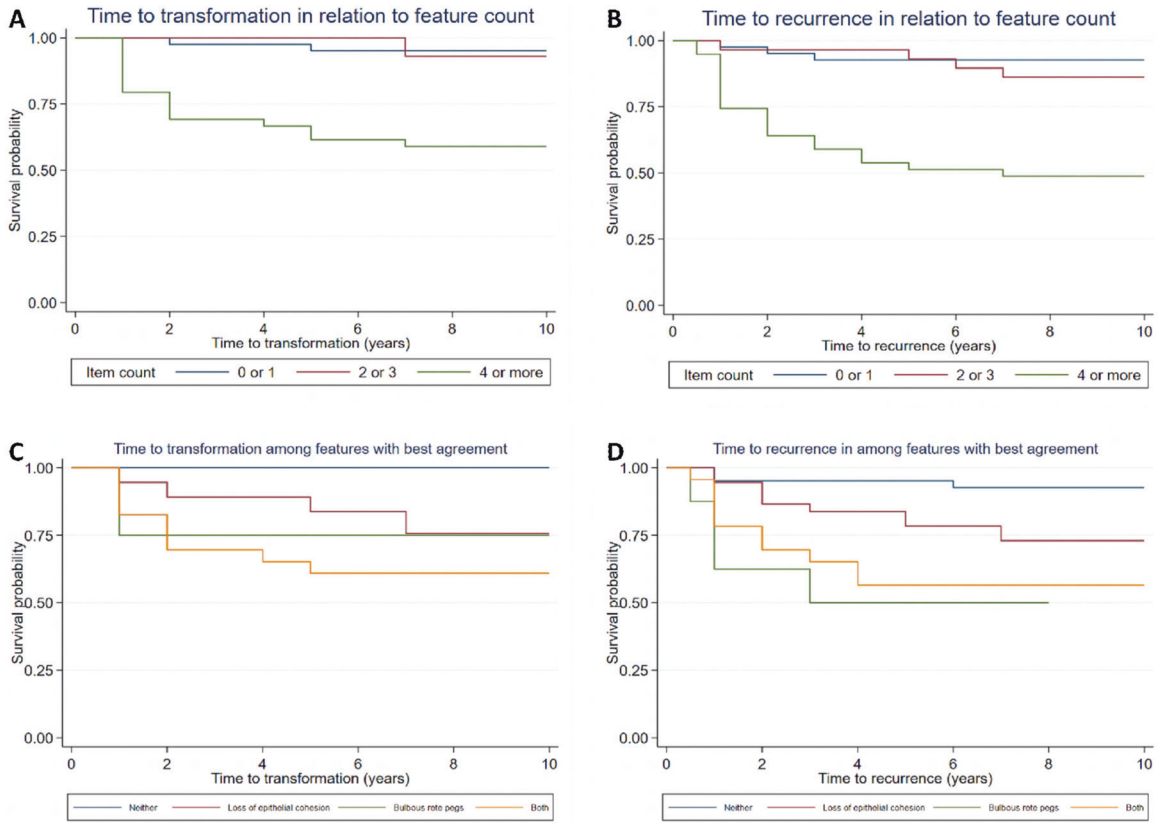

Fig. 1 Kaplan–Meier curves for time to transformation and recurrence for feature count based on the six-point scoring system (**A**, **B**) and the two-point scoring system (**C**, **D**).

Figure 1A and B (see Supplementary Material) present the Kaplan–Meier survival curves for time to transformation and time to recurrence in relation to the number of features present using the six-point scoring model. The predicted transformation rate at 2 years is estimated at 2% (95% CI 0–16%) for 0–1-point scoring, 0% for 2–3-point scoring and 31% (95% CI 19–48%) for 4–6-point scoring. At 5 years, these figures increase to 5% (95% CI 1–18%) for 0–1-point scoring and 38% (95% CI 25–55%) for 4–6-point scoring; there is no change in the rate for 2–3-point scoring (0%). For recurrence of OED, the respective predicted rates at two and five years were shown to be: 5% (95% CI 1–18%) and 7% (95% CI 2–21%) for 0–1-points; 3% (95% CI 0–22%) and 7% (95% CI 2–25%) for 2–3-points; 36% (95% CI 23–53%) and 49% (95% CI 34–65%) for 4–6 points. The lower recurrence and transformation rate seen for 2–3-point scoring compared to 0–1 points is unexpected but is likely to be related to the much lower number of cases in the 2–3 point category compared to the others. Validation on a more balanced larger cohort would be useful to determine the significance of these findings.

Few transformations and recurrences occurred more than five years post-baseline, and for simplicity the prognostic performance was assessed on the basis of whether the event happened rather than the time taken to occur. Figure 2 (see Supplementary Material) shows the receiver-operator characteristic curve (ROC) for these. The sensitivity and specificity appeared best balanced by using a cut off for either 4 or 5 points, with less events (for transformation and recurrence) when fewer features were present. The AUROCs for transformation and recurrence were 0.799 and 0.776, respectively.

Prognostic model 2: Reduced two-point scoring system. The second scoring system selected two features with the best inter-rater agreement, and which were also associated with transformation and recurrence (i.e. loss of epithelial cohesion and bulbous/drop shaped rete pegs). Figure 1C and D (see Supplementary Material) show Kaplan–Meier survival curves for time to transformation and recurrence based on the presence or absence of these two features. The combined presence of both features appeared to be associated with a higher risk of malignant transformation (39%, 95% CI 23–62%) at five years, in comparison to the presence of a single feature alone (loss of epithelial cohesion [16%, 95% CI 8–33%], bulbous/drop-shaped rete pegs [25%, 95% CI 7–69%]). However, the presence of bulbous/drop shaped rete pegs showed a higher risk of recurrence at five years (50%, 95% CI 23–85%) as compared to the presence of loss of epithelial cohesion (22%, 95% CI 11–39%) or when both of features were present in combination (43%, 95% CI 26–66%).

Effect of patient/clinical characteristics on prognostic models

The association between patient characteristics (age, gender, intraoral site), OED histological grade and clinical outcomes were also assessed. Overall, there was a modest association between patient characteristics and clinical outcomes. However, there was a trend for higher rates of transformation and recurrence amongst older patients compared to younger, and generally with higher graded lesions as well. Moderate OED lesions were associated with a marginally higher rate of malignancy and recurrence in comparison to severe OED lesions (31% vs 15%, 38% vs 26%, respectively, Table 5). The rates for intraoral clinical sites were, at best, modestly associated with dysplasia outcomes. None of the features had an AUROC as high as that achieved by the two scoring systems.

Table 6 illustrates the effect of adding the clinical characteristics (age, gender) and histological grade (WHO and binary) to each of the prognostic models, as represented by the AUROC. Adding age and gender into the models only marginally improved the

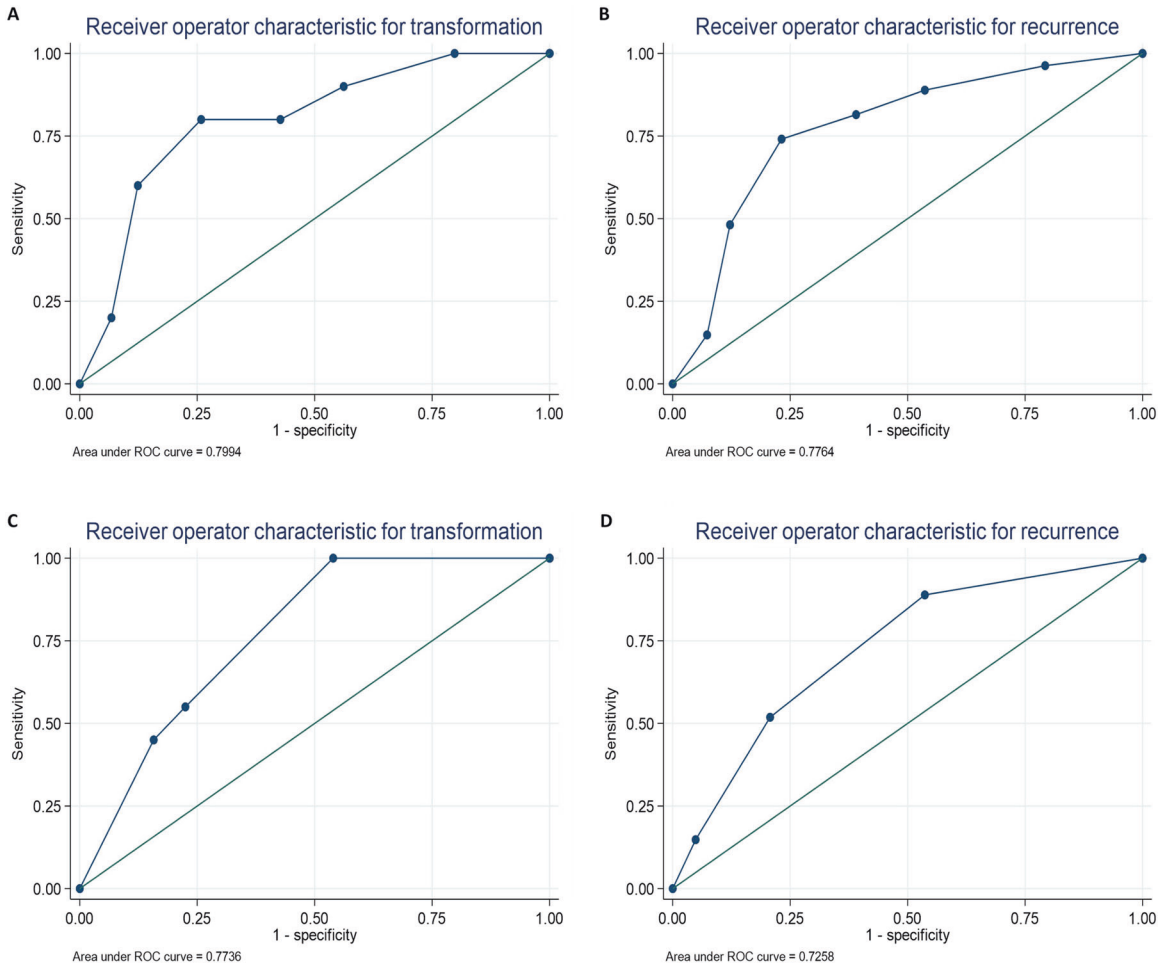

Fig. 2 ROC curves for transformation and recurrence for feature count based on the six-point scoring system (**A**, **B**) and the two-point scoring system (**C**, **D**).

predictive ability of the scoring (6-point model: 0.810 transformation, 0.804 recurrence; 2-point model: 0.810 transformation, 0.759 recurrence), reflecting the modest association of these characteristics with transformation and recurrence. Adding the histological grade improved the models further, particularly with the WHO grade compared to the binary grade (6-point model: 0.837 vs 0.812 for transformation, 0.812 vs 0.790 for recurrence; 2-point system: 0.843 vs 0.805 for transformation, 0.780 vs 0.755 for recurrence). The number of intraoral site categories and the relatively sparse number of patients for some sites meant it was not possible to jointly model this along with the proposed scoring approaches.

Comparison of proposed models to existing grading systems

The prognostic ability of the two proposed models were compared against the existing grading systems[20]. Both the 'six-point' and 'two-point' proposed models yielded a higher AUROC than achieved by either WHO or binary grading systems, although not all these differences were statistically significant. The more detailed six-point model demonstrated a statistically significantly higher AUROC than achieved by the WHO grading system for both transformation and recurrence, but a more marginal improvement over binary grading. The two-point model showed a significant improvement over WHO grading for transformation alone (Table 7).

Finally, the prognostic performance of the new models was calculated separately for each of the three raters, reflecting how the models are likely to be used in clinical practice. Both models showed reduced prognostic ability when used by a single rater, indicating a greater risk for misclassification compared to models that were based on consensus agreement. Of the 12 single-rater AUC measures derived from the proposed models, 11 remained higher than those derived from corresponding WHO or binary grade (Table 8). Nevertheless, this analysis indicates that significant improvements on existing grading requires greater levels of agreement by assessors.

DISCUSSION

This study reveals important and novel information about the prognostic significance of individual histological features of OED. We have demonstrated histological feature-specific correlation of OED to malignant transformation and recurrence, which has allowed us to propose two prognostic scoring models with a potential to simplify and aid OED diagnosis and grading in the future.

Overall, nine histological features were shown to be most prevalent amongst our OED cohort (Table 2). The top two most common features were basal cell hyperplasia (crowding) and irregular surface keratin; neither of which are currently part of the WHO criteria for OED diagnosis, although our study did not show them to be strongly linked to transformation or recurrence. The least prevalent features were verrucous surface morphology, lymphocytic band, loss of epithelial cohesion, dyskeratosis and nuclear pleomorphism. Interestingly, the latter three of these features were positively associated with clinical outcomes of interest; loss of epithelial cohesion (transformation $p = 0.003$, recurrence $p = 0.001$), nuclear pleomorphism (transformation $p =$

Table 5. Incidence for transformation and recurrence by patient characteristics and OED histological grade.

Model	*N*	Transformation	Recurrence
Age			
<55	23	3 (13%)	3 (13%)
55–64	20	4 (20%)	5 (25%)
65–74	31	5 (16%)	6 (19%)
>=75	35	8 (23%)	13 (37%)
AUROC		0.526	0.591
Gender			
Female	42	8 (19%)	10 (24%)
Male	67	12 (18%)	17 (25%)
AUROC		0.509	0.510
WHO grade			
Mild	34	1 (3%)	2 (6%)
Moderate	48	15 (31%)	18 (38%)
Severe	27	4 (15%)	7 (26%)
AUROC		0.601	0.624
Binary grade			
Low	73	8 (11%)	12 (16%)
High	36	12 (33%)	15 (42%)
AUROC		0.665	0.650
Site of disease			
Tongue	44	10 (23%)	14 (32%)
Floor of mouth	23	3 (13%)	3 (13%)
Buccal mucosa	17	3 (18%)	5 (29%)
Gingivae	7	3 (43%)	4 (57%)
Hard palate	6	0	0
Lower lip	6	1 (17%)	1 (17%)
Soft palate	6	0	0
AUROC		0.544	0.547

Table 6. AUROC for each model incorporating age, gender and grading.

Model	Transformation	Recurrence
6-point score only	0.799	0.776
6-point score + age + gender	0.810	0.804
6-point score + WHO grade	0.837	0.800
6-point score + binary grade	0.812	0.790
2-point score only	0.774	0.726
2-point score + age + gender	0.810	0.759
2-point score + WHO grade	0.843	0.780
2-point score + binary grade	0.805	0.755

0.005, recurrence $p = 0.002$) and dyskeratosis (recurrence $p = 0.042$) indicating that the presence of the features and not the frequency within the cohort was more important. It is evident that certain architectural features may be consistently easier to detect (even at lower magnification) as compared to other features at cellular or nuclear level. The use of immunohistochemical markers, such as Phosphorylated Histone H3 (PHH3) and Ki67 can be considered as adjuncts for the assessment of mitosis and cell proliferation[21], although more extensive evaluation of their usefulness as a prognostic indicator in OED is needed.

Our study showed observer agreement to be the highest for verrucous surface morphology, abrupt orthokeratosis, lymphocytic band and loss of epithelial cohesion, and worst for hyperchromatism and suprabasal mitoses, further highlighting the difficulty in objective analysis of certain features in clinical practice, particularly the more ambiguously defined cytological atypia. Several studies have investigated the variability in inter- and intra-observer agreement in the diagnosis and grading of OED, with substantially different outcomes ranging from poor to high observer agreement[22–25]. One of the challenges that arises in analysing inter-rater agreement is the variation that exists in pathologists' understanding and definitions of features due to their inherently subjective nature further complicated by the numerous changes to classifications and reporting definitions over the years. Although digital WSIs were used to mitigate the issue of variations in staining of glass slides for each pathologist, the experience of digital reporting/analysis may have caused some variation. In this study, apart from informal discussions there were no formal calibration exercises arranged prior to histological examination, as we had intended for grading and feature scoring to be most reflective of the real world and routine clinical practice. To overcome any deficiencies in feature prevalence and agreement, two chance-corrected measures were used, including bias adjusted Kappa and Gwet's AC1, as per statistical recommendation[26].

We found six histological features (bulbous/drop shaped rete pegs, hyperchromatism, loss of epithelial cohesion, loss of stratification, suprabasal mitoses, nuclear pleomorphism) to be associated with a greater incidence of transformation and recurrence. Although it is well acknowledged that atypical verrucous hyperplasia and/or keratoses are a subset of OPMD, and that proliferative verrucous leukoplakia has a high reported rate of malignant transformation[27,28], we did not find a statistical association between verrucous surface morphology and clinical outcomes in our study.

Although there was a modest association between patient characteristics and clinical outcomes, there is a statistical trend for higher rates of transformation and recurrence amongst older patients as well as higher graded lesions. This trend is well supported in the literature and is thought to be related to the aggregation of genetic alterations, immunosenescence and chronic exposure to environmental risk factors with advancing age[29,30]. Interestingly though, lesions graded as moderate dysplasia were associated with a marginally higher rate of malignancy and recurrence in comparison to severe dysplasia grades (31% vs 15%, 38% vs 26%, respectively, Table 5). These

Table 7. Comparison of AUROC between two-point and six-point models with existing grading systems.

	Transformation			Recurrence		
	AUC	*p* value v WHO grade	*p* value v binary grade	AUC	*p* value v WHO grade	*p* value v binary grade
WHO grade	0.601	-	-	0.624	-	-
Binary grade	0.665	-	-	0.650	-	-
Two-point model	0.774	<0.001	0.082	0.720	0.083	0.207
Six-point model	0.799	<0.001	0.082	0.776	0.003	0.050

Table 8. AUROC for two- and six-point models calculated separately for each rater.

	AUC for Malignant Transformation				AUC for OED Recurrence			
	Rater 1	Rater 2	Rater 3	Consensus	Rater 1	Rater 2	Rater 3	Consensus
Two-point model	0.785	0.741	0.692	0.774	0.784	0.691	0.633	0.720
Six-point model	0.760	0.733	0.692	0.799	0.753	0.750	0.639	0.776

findings could be explained by differences in treatments and clinical follow-up, particularly in relation to moderately graded OED lesions which are both challenging to diagnose/grade and treat. The lack of robust treatment guidelines means there is huge disparity in the management of such lesions between surgeons. Although our patient cohort was diagnosed at a single centre, differences in treatment regimens between regional hospitals, and medical/social risk factors are likely to have contributed to potential differences in their management. This further highlights the need for improved diagnostic methods which are independent of grade for more objective OED prognostication as well as more standardised treatment pathways.

We developed and assessed the potential of using two relatively simple point-based scoring systems, based on the presence or absence of certain histological features. Using the six-point model, patient scoring '4–6 points' were predicted to be at the highest risk of malignant transformation and recurrence at five years, estimated at 38% (95% CI 25–55%) and 49% (95% CI 34–65%), respectively. For the two-point model, predictions suggest that the presence of bulbous/drop shaped rete pegs alone have a greater predictive association with transformation (25%, 95% CI 7–69%) and recurrence (50%, 95% CI 23–85%) at five years, compared to the presence of loss of epithelial cohesion alone (transformation at five years: 16%, 95% CI 8–33% and recurrence at five years: 22%, 95% CI 11–39%).

Comparing the two systems, the six-point model had a greater discriminant performance with more separation of the survival and ROC curves (Figs. 1 and 2, see Supplementary Material). Although it is important to highlight that based on the modest agreement between pathologists seen in this study, it is inevitable that the performance of this system may be weakened if there was only a single assessor conducting the analysis. In contrast, the two-point model is a simplified approach that focusses only on the two features with the best inter-rater agreement (presence of loss of epithelial cohesion and/or bulbous/drop shaped rete pegs which are easier to identify). This model retained predictive ability contained in the groupings (especially for transformation) whilst being less susceptible to inter-rater disagreement.

The authors acknowledge a few limitations of this study. The first relates to the relatively small sample size which was obtained from a single centre. However, the department in question is a regional and national referral centre in the UK and therefore receives tissue samples from multiple hospitals covering a wide geographical area, thereby providing a sufficiently varied cohort for this pilot study. Furthermore, whilst the sample size may be considered small, it is larger than other studies which have explored OED analysis or proposed alternate OED grading classifications[12,16,21]. Nevertheless, application of these findings to substantially larger multicentre cohorts will allow more robust validation of the proposed potential prognostic models[31].

To the best of our knowledge, this is the first study to propose feature specific prognostic scoring models for OED. The proposed models have the potential to provide pathologists with greater insight into the risk of individual OED lesions based on feature-specific analysis, which will in turn aid clinical decision making with regards to treatment and follow-up. Larger validation of the models is required on multicentric cohorts, with prospective analysis to explore the impact of other clinical determinants such as medical/social risk factors as well as effects of treatment and frequency of monitoring. There is clearly potential for strengthening the predictive ability of the models by incorporating such measures.

Greater clarity on the definitions (and examples) for individual architectural and cytological features will greatly benefit pathologists with OED diagnosis/grading and help to improve intra-observer agreement. There is clearly a need for the development of a universal minimum dataset for the reporting of OED lesions, as well as benefit in double/consensus reporting by two pathologists to ensure accurate diagnosis and early treatment.

DATA AVAILABILITY

All data generated or analysed during this study are included in this published article.

REFERENCES

1. Lumerman H, Freedman P, Kerpel S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1:321–9 (1995).
2. Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral Oncol. 1:309–16 (2009).
3. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, 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).
4. Speight PM. Update on oral epithelial dysplasia and progression to cancer. Head Neck Pathol 1:61–6 (2007).
5. Locca O, Sollecito TP, Alawi F, Weinstein GS, Newman JG, De Virgilio A et al. Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. Head Neck 42:539–55 (2020).
6. Warnakulasuriya S, Kujan O, Aguirre-Urizar JM, Bagan JV, González-Moles MÁ, Kerr AR et al. Oral potentially malignant disorders: A consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer. Oral Dis 27:1862–80 (2021).
7. Reibel J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. Crit Rev Oral Biol Med 14:47–62 (2003).
8. Speight PM, Khurram SA, Kujan O. Oral potentially malignant disorders: risk of progression to malignancy. Oral Surg Oral Med Oral Pathol Oral Radiol. 125:612–27 (2018).
9. Odell E, Kujan O, Warnakulasuriya S, Sloan P. Oral epithelial dysplasia: recognition, grading and clinical significance. Oral Dis. 27:1947–76 (2021).
10. El-Naggar AK, Chan JK, Grandis JR, Takata T, Slootweg PJ. WHO Classification of Head and Neck Tumors (4th ed. Vol. 9) Tumours of the oral cavity and mobile tongue (p 112) (IARC, 2017).
11. Warnakulasuriya S. Histological grading of oral epithelial dysplasia: revisited. J. Pathol 194:294–7 (2001).
12. Kujan O, Khattab A, Oliver RJ, Roberts SA, 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 1:224–31 (2007).
13. Leemans CR, Snijders PJ, Brakenhoff RH. The molecular landscape of head and neck cancer. Nat Rev Cancer 18:269–82 (2018).
14. Califano J, Van Der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S et al. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer Res. 1:2488–92 (1996).
15. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. Oral Oncol 1:987–93 (2006).
16. Nankivell P, Williams H, Matthews P, Suortamo S, Snead D, McConkey C et al. The binary oral dysplasia grading system: validity testing and suggested improvement. Oral Surg Oral Med Oral Pathol Oral Radiol 1:87–94 (2013).
17. Kramer IR, Lucas RB, El-Labban N, Lister L. The use of discriminant analysis for examining the histological features of oral keratoses and lichen planus. Br. J. Cancer 24:673–83 (1970).

18. MacDonald DG, Saka SM. Structural indicators of the high risk lesion. In: Johnson NE, editor. Risk Markers for Oral Diseases. Oral Cancer: Detection of Patients and Lesions at Risk. C.U.P. (1991).
19. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC. StataCorp (2019).
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1:837–45 (1988).
21. Sudarshini N, Banavar SR, Nambiar SK, Augustine D, Haragannavar VC, Sowmya S et al. Immunohistochemical Stain-Phosphohistone H3: Most Specific Mitotic Marker. J. clin. diagn 1;12 (2018).
22. Pindborg JJ, Reibel J, Holmstrup P. Subjectivity in evaluating oral epithelial dysplasia, carcinoma in situ and initial carcinoma. J. Oral Pathol Med 14:698–708 (1985).
23. Karabulut A, Reibel J, Therkildsen MH, Praetorius F, Nielsen HW, Dabelsteen E. Observer variability in the histologic assessment of oral premalignant lesions. J. Oral Pathol Med 24:198–200 (1995).
24. DJ, Brothwell, Lewis DW, Leong I, Jordan RCK, and Leake J. L. Observer agreement in the grading of oral epithelial dysplasia. Community Dent. Oral Epidemiol 31:300–5 (2003).
25. Fischer DJ, Epstein JB, Morton Jr TH, Schwartz SM. Interobserver reliability in the histopathologic diagnosis of oral pre-malignant and malignant lesions. J. Oral Pathol Med 33:65–70 (2004).
26. Flight L, Julious SA. The disagreeable behaviour of the kappa statistic. Pharm. Stat 14:74–8 (2015).
27. Müller S. Oral epithelial dysplasia, atypical verrucous lesions and oral potentially malignant disorders: focus on histopathology. Oral Surg Oral Med Oral Pathol Oral Radiol 1:591–602 (2018).
28. Thompson LD, Fitzpatrick SG, Müller S, Eisenberg E, Upadhyaya JD, Lingen MW et al. Proliferative verrucous leukoplakia: an expert consensus guideline for standardized assessment and reporting. Head Neck Pathol 15:572–87 (2021).
29. Gupta PC, Mehta FS, Daftary DK, Pindborg JJ, Bhonsle RB, Jalnawalla PN et al. Incidence rates of oral cancer and natural history of oral precancerous lesions in a 10-year follow-up study of Indian villagers. Community Dent. Oral Epidemiol. 8:283–333 (1980).
30. Ranganathan K, Kavitha L. Oral epithelial dysplasia: Classifications and clinical relevance in risk assessment of oral potentially malignant disorders. J Oral Maxillofac Pathol 23:19 (2019).
31. Moons KG, Royston P, Vergouwe Y, Grobbee DE, Altman DG. Prognosis and prognostic research: what, why, and how? BMJ 23;338 (2009).

ACKNOWLEDGEMENTS

We thank Dr Hannah Walsh (Oral Pathology Specialist Registrar, School of Clinical Dentistry, Sheffield, UK) for her assistance with retrieval of the histology cases.

AUTHOR CONTRIBUTIONS

HM and SAK conceived and designed the study. HM retrieved and prepared the case cohort and re-analysed histological grading. NMI, OK and SAK undertook independent histological examination for feature scoring. MB provided statistical expertise. HM, MB, NMR, NMI, OK and SAK were involved in analysis of results. HM led on writing of the manuscript with contributions from MB, NMR, NMI, OK and SAK.

FUNDING

HM is funded by the National Institute for Health Research (Award ID: NIHR300904) and received a grant from Sheffield Hospitals Charity to support this work. NMR and SAK are partly funded by a Cancer Research UK Project grant.

COMPETING INTERESTS

The authors declare no competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval was granted by the West Midlands - Edgbaston Research Ethics Committee (reference: 18/WM/0335).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41379-022-01067-x.

Correspondence and requests for materials should be addressed to Syed Ali Khurram.

Reprints and permission information is available at http://www.nature.com/reprints

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

4.2 Validation of feature-based scoring and histological grading for the prediction of malignant transformation of oral epithelial dysplasia

Authors: **H. Mahmood[1],** M. Bradburn[2], N. M. Islam[3], S. A. Khurram[4]

[1] Academic Unit of Oral Surgery, School of Clinical Dentistry, University of Sheffield

[2] Clinical Trials Research Unit, Sheffield Centre for Health and Related Research, University of Sheffield

[3] Department of Oral & Maxillofacial Diagnostic Sciences, College of Dentistry, University of Florida, USA

[4] Unit of Oral & Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield

Abstract

Background: Oral epithelial dysplasia (OED) is a histopathological diagnosis given to lesions of the oral cavity that have an increased risk of progression to malignancy. Histological grading is used for prediction of malignant transformation risk of OED. However, grading is associated with significant subjectivity, observer variability, and inconsistency in prognosis prediction. Previously developed histological feature-specific models ('six-point' and 'two-point') demonstrated consistently better predictive performances compared to conventional grading systems. This study aims to validate these feature-based models on independent OED datasets.

Methods: Consecutive OED cases (dating 2012 – 2017) were acquired from four different centres. The six histological features used to formulate the models ('bulbous/drop shaped rete pegs', 'hyperchromatism', 'loss of epithelial cohesion', 'loss of stratification', 'suprabasal mitoses', 'nuclear pleomorphism') were independently scored by three observers. Histological feature prevalence and observer agreement were calculated. Malignant transformation risk was measured using AUROC curves, Kaplan Meier and Cox regression analysis.

Results: 102 OED were included (13 (13%) from Sheffield, 40 (39%) from Belfast, 30 (29%) from Birmingham and 19 (19%) from Brazil). 28 (27%) were graded as mild OED, 41 (40%) as moderate OED and 33 (32%) as severe OED. The 'six-point' system demonstrated superior performance for transformation prediction (AUROC of 0.81) compared to the 'two-point' system (AUROC=0.73, p=0.004), WHO grading (AUROC=0.71, p=0.03) and binary grading (AUROC=0.68, p=0.009). The predicted transformation rate for the 'six-point' model was 50% (95% CI 27 to 78%) when all 6 features were present compared to 35% (95% CI 23 to 53%) when 4-5 features were present and 14% (95% CI 5 to 32%) when 2-3 features were present.

Conclusion: This study validates the superior performance of the 'six-point' system for transformation risk on a multicentric sample of OED cases. Findings indicate that feature-specific models may be more reliable than existing histological grading systems for prognosis prediction and could help standardise treatment protocols and improve clinical outcomes for patients with OED.

Keywords: Oral epithelial dysplasia; malignant transformation; histological grading; prognosis

4.2.1 Background

Oral epithelial dysplasia (OED) is a premalignant histopathological condition given to lesions of the oral cavity that have an increased risk of progression to malignancy. Histological grading of OED presents significant challenges and uncertainties, particularly in relation to the prediction of malignant transformation risk. Grading is also associated with significant subjectivity and irreproducibility. However, despite the dubious reliability of this approach, it remains the mainstay for cancer risk prediction, influencing clinical treatment decisions for patients with OED. Another limitation of current grading systems, both WHO (2017) and the alternative binary system, is that they fail to ascribe prognostic value to individual histological features. In fact, until recently, there has been very little by means of research into histological predictors for OED progression.

In a recent study by Mahmood *et al.* (2022)[9], a correlation between individual OED histological features and prognosis was demonstrated, for the first time. The authors found six histological features ('bulbous/drop shaped rete pegs', 'hyperchromatism', 'loss of epithelial cohesion', 'loss of stratification', 'suprabasal mitoses', 'nuclear pleomorphism') to be associated with a greater risk of malignant transformation ($p < 0.036$) and OED recurrence ($p < 0.015$)[9]. These features were used to develop the 'six-point' and 'two-point' feature-based scoring models, in which a single point is allocated for the presence of each of the significant features. Both models demonstrated significantly better predictive ability (AUROC ≥ 0.774 for malignant transformation and 0.726 for recurrence) than the WHO (2017) and binary grading systems.

This study conducts validity testing of the 'six-point' and 'two-point' scoring models on independent external OED cases. There were three main objectives: first, to independently validate the prognostic ability of the 'six point' and 'two-point' scoring models for prediction of malignant transformation risk of OED; second, to evaluate the inter-observer variability of the 'six-point' and 'two-point' scoring models compared with the WHO (2017) and binary grading systems; third, to analyse the impact of clinical variables on the prognostic abilities of the 'six-point' and 'two-point' scoring models compared to existing clinical grading systems.

4.2.2 Methods

4.2.2.1 *Validation dataset and clinical data collection*

A multicentric retrospective sample of 104 OED cases were acquired in the form of digital whole slide images (WSI) from four different centres, as below.

i. School of Clinical Dentistry, University of Sheffield, UK (n=13)
ii. Precision Medicine Centre, Patrick G. Johnston Centre for Cancer Research, Queen's University Belfast, UK (n=40)
iii. Institute of Head and Neck Studies and Education (InHANSE), University of Birmingham, UK (n=30)
iv. Piracicaba Dental School, UNICAMP, Brazil (n=19)

Purposive sampling was used to acquire consecutive cases from these centres dating between 2012 and 2017. In the case of two samples from the Sheffield cohort, malignant transformation occurred within three months of OED diagnosis and due to the difficulty in reliably confirming whether underlying malignancy was already present, these cases were excluded from analysis. The resulting sample therefore comprised 102 OED cases for independent model validation. Further details of the validation cohort are provided in Table 3.

Prior to the inclusion of cases in this study, they were blindly reviewed by a Consultant Oral and Maxillofacial Pathologist to ensure tissue quality and to review the original histological grade. Where necessary, an updated grade using the WHO (2017) and binary systems were assigned. The inclusion criteria were varying grades of OED with sufficient epithelial tissue for analysis and availability of minimum three-year follow-up data. Cases demonstrating histological atypia secondary to inflammatory conditions (i.e. candida albicans infection or lichen planus) were excluded. OED lesions positive for the Human Papilloma Virus and lesions with verrucous surface morphology were also excluded as they are distinct entities with reportedly different behaviour that could introduce bias into the sample. Ethical approval was granted by the West Midlands - Edgbaston Research Ethics Committee (reference: 18/WM/0335) and the research was carried out in compliance with the Helsinki Declaration.

Clinical data collection included patient age, sex, biopsy site, original histological OED grade and the time to transformation in months (if applicable). As per the previous study, transformation was confirmed as an OED lesion which had progressed to OSCC at the same clinical site within the follow-up period. Due to difficulty in confirming OED recurrence status from collaborating centres, the authors decided to exclude this outcome measure for the presented study. Data from the four different centres was recorded in a structured proforma using Microsoft Excel (2016) in an anonymised-linked format.

The TRIPOD (transparent reporting of a multivariable prediction model for individual prognosis or diagnosis) guidelines[119] were used to strengthen the methodology and conclusions of this study.

4.2.2.2 *Histological Feature Assessment*

Three assessors, including two experienced clinical-academic Oral and Maxillofacial Pathologists (NI, SAK) and a clinician with extensive expertise and a specialist interest in OED analysis (HM) conducted independent histological feature examination. The assessors were given access to the WSI sample via a secure cloud-based system and were blinded to the original diagnosis and clinical outcomes.

The assessors independently scored the six histological features identified in the previous study as having adequate inter-rater agreement and being associated with malignant transformation risk (Figure 14). These six features are a subset of the original 12 histological features evaluated in the previous work:

i. Bulbous rete pegs
ii. Hyperchromatism
iii. Loss of epithelial cohesion
iv. Loss of stratification
v. Nuclear pleomorphism
vi. Suprabasal mitosis

Feature scoring was conducted using a similar approach to the initial study, whereby the assessors were asked to provide a binary score to record the presence or absence of individual features. A score of 1 was given if the feature was abundantly visible, and a score of 0 was given if the feature was absent or rare/focal. Scores were entered into a pre-developed Microsoft Excel (2016) spreadsheet.

Figure 14. *Histological features forming the 'six-point' and 'two-point' models.*

4.2.2.3 *Statistical Analysis & Outcome Measures*

Statistical analyses were conducted using the Stata Statistical Software (Version 17, 2021)[120]. All tests were two-tailed and $p < 0.05$ were considered statistically significant. Three outcomes were measured:

1) Histological feature prevalence

The prevalence of a feature was calculated separately for each assessor, and also by "consensus" which was defined as the number of patients for whom at least two assessors considered a feature "present".

2) Observer agreement

Agreement was summarised as the percentage of patients for whom all three raters agreed, and by two chance-corrected measures (Cohen's Kappa coefficient and Gwet's AC). For the latter two measures, a value of 1 denotes perfect agreement whilst 0 relates to no agreement beyond chance alone. The kappa scores were interpreted based on historical standards, with scores ≤0.2 representing "slight", 0.2-0.4 "fair", 0.4-06 "moderate", 0.6-0.8 "substantial", and 0.8-1.0 "near perfect" agreement[121]. Whilst the Kappa statistic is commonly used, it can be affected by prevalence, therefore the alternative Gwet AC1 measure was also used in this study[122].

3) Malignant transformation

Malignant transformation was analysed using two approaches. The first approach measured the incidence of transformation using the area under the receiver-operator characteristic (AUROC) curve to assess predictive ability. The second incorporated the follow up duration using Kaplan Meier to visualise associations between individual pathological features and time to transformation and used Cox regression to calculate Hazard Ratios with 95% confidence intervals.

4.2.3 Results

Characteristics of the Validation Sample

102 cases, each from a different participant, were used for independent model validation and analysis. The sample included 13 (13%) cases from Sheffield, 40 (39%) from Belfast, 30 (29%) from Birmingham and 19 (19%) from Brazil. Overall, 28 (27%) were graded as mild OED, 41 (40%) as moderate OED and 33 (32%) as severe OED. Binary grading confirmed 37 (36%) as low grade and 65 (64%) as high grade lesions. There was a slightly higher proportion of females (55, 54%) compared to males (47, 46%) with an average age of 58.9 years (IQR 12.7) (Table 3).

Table 3. *Characteristics of the study sample.*

Feature	N (%) or mean [SD]
Centre	
Sheffield	13 (13%)
Belfast	40 (39%)
Birmingham	30 (29%)
Brazil	19 (19%)
Age	58.9 [12.7]
Sex	
Female	55 (54%)
Male	47 (46%)
Grade (WHO)	
Mild	28 (27%)
Moderate	41 (40%)
Severe	33 (32%)
Grade (binary)	
Low	37 (36%)
High	65 (64%)

4.2.3.1 *Feature Prevalence*

With the exception of suprabasal mitoses, all six histological features were notably more prevalent in this study compared with the previous study which proposed the two new models (Mahmood *et al.* 2022)[9]. The most prevalent features in this study were 'hyperchromatism' (97%) and 'nuclear pleomorphism' (90%) and the least prevalent features were 'loss of epithelial cohesion' (45%) and 'suprabasal mitoses' (24%) (Table 4).

Table 4. *Feature-specific prevalence with comparison to previous stud.y*

*Feature prevalence presented as consensus (defined as two or more assessors recording a positive assessment) and separately for each rater. *Denominator is N=306, or 102 patient samples x 3 raters.*

	Present study				*Mahmood et al (2022)*
	Consensus*	Assessor 1	Assessor 2	Assessor 3	
Bulbous rete Pegs	72 (71%)	68 (67%)	67 (66%)	85 (83%)	*57%*
Hyperchromatism	99 (97%)	97 (95%)	98 (96%)	97 (95%)	*54%*
Loss of epithelial cohesion	46 (45%)	49 (48%)	28 (27%)	73 (72%)	*30%*
Loss of stratification	63 (62%)	70 (69%)	61 (60%)	46 (45%)	*42%*
Nuclear pleomorphism	92 (90%)	94 (92%)	92 (90%)	80 (78%)	*36%*
Suprabasal mitoses	24 (24%)	23 (23%)	21 (21%)	93 (91%)	*45%*

4.2.3.2 *Observer Agreement*

Interobserver agreement was typically modest for all six histological features, although the agreement of certain features, differed notably when compared to the previous study by Mahmood *et al.* (2022)[9] (Table 5). Two features in particular demonstrated greater interobserver variability in this study, namely 'loss of epithelial cohesion' (Gwet's AC1 0.29) and 'suprabasal mitosis' (Gwet's AC1 0.06).

Table 5. *Interobserver agreement with comparison to previous study*

	Complete agreement		Cohen's Kappa		Gwet's AC1	
	Present study	*Mahmood et al (2022)*	Present study	*Mahmood et al (2022)*	Present study	*Mahmood et al (2022)*
Bulbous rete Pegs	70 (69%)	*72 (66%)*	0.49	*0.54*	0.65	*0.56*
Hyperchromatism	93 (91%)	*53 (49%)*	0.33	*0.32*	0.94	*0.32*
Loss of epithelial cohesion	48 (47%)	*80 (73%)*	0.34	*0.58*	0.29	*0.69*
Loss of stratification	64 (63%)	*61 (56%)*	0.50	*0.41*	0.52	*0.43*
Nuclear pleomorphism	70 (75%)	*62 (57%)*	0.26	*0.38*	0.78	*0.47*
Suprabasal mitoses	29 (28%)	*54 (50%)*	0.21	*0.34*	0.06	*0.34*

4.2.3.3 *Malignant Transformation Incidence and Prediction*

The incidence of malignant transformation (based on the "consensus" definition) demonstrated that 'nuclear pleomorphism' and 'hyperchromatism' were commonly recorded both among transformed and non-transformed cases with little distinction between the two. But the remaining four features, whilst common, demonstrated more apparent separation between those that did and did not transform (Table 6).

Table 6. *Malignant transformation incidence in relation to individual histological features.*

Feature	Transformed (N=41)	Not transformed (N=61)
Bulbous rete Pegs	38 (93%)	34 (56%)
Hyperchromatism	40 (98%)	59 (97%)
Loss of epithelial cohesion	25 (61%)	21 (34%)
Loss of stratification	35 (85%)	28 (46%)
Nuclear pleomorphism	41 (100%)	51 (84%)
Suprabasal mitoses	19 (46%)	5 (8%)

As per the previous study, the six studied histological features were used to develop two scoring systems: a 'six-point' system in which each feature is given a single point if present, and a more simplified 'two-point' system based on the two features originally identified as having the best rater agreement ('loss of epithelial cohesion' and 'bulbous rete pegs'). It should be noted that, in this study, 'loss of epithelial cohesion' was among the least well agreed features.

Figure 15 demonstrates the incidence of malignant transformation in relation to the two scoring approaches, in addition to the WHO and binary grading systems. Whilst differences were apparent in relation to both, the most immediately obvious was the 'six-point' system which had an increasing proportion of cases that transformed as the score increased.

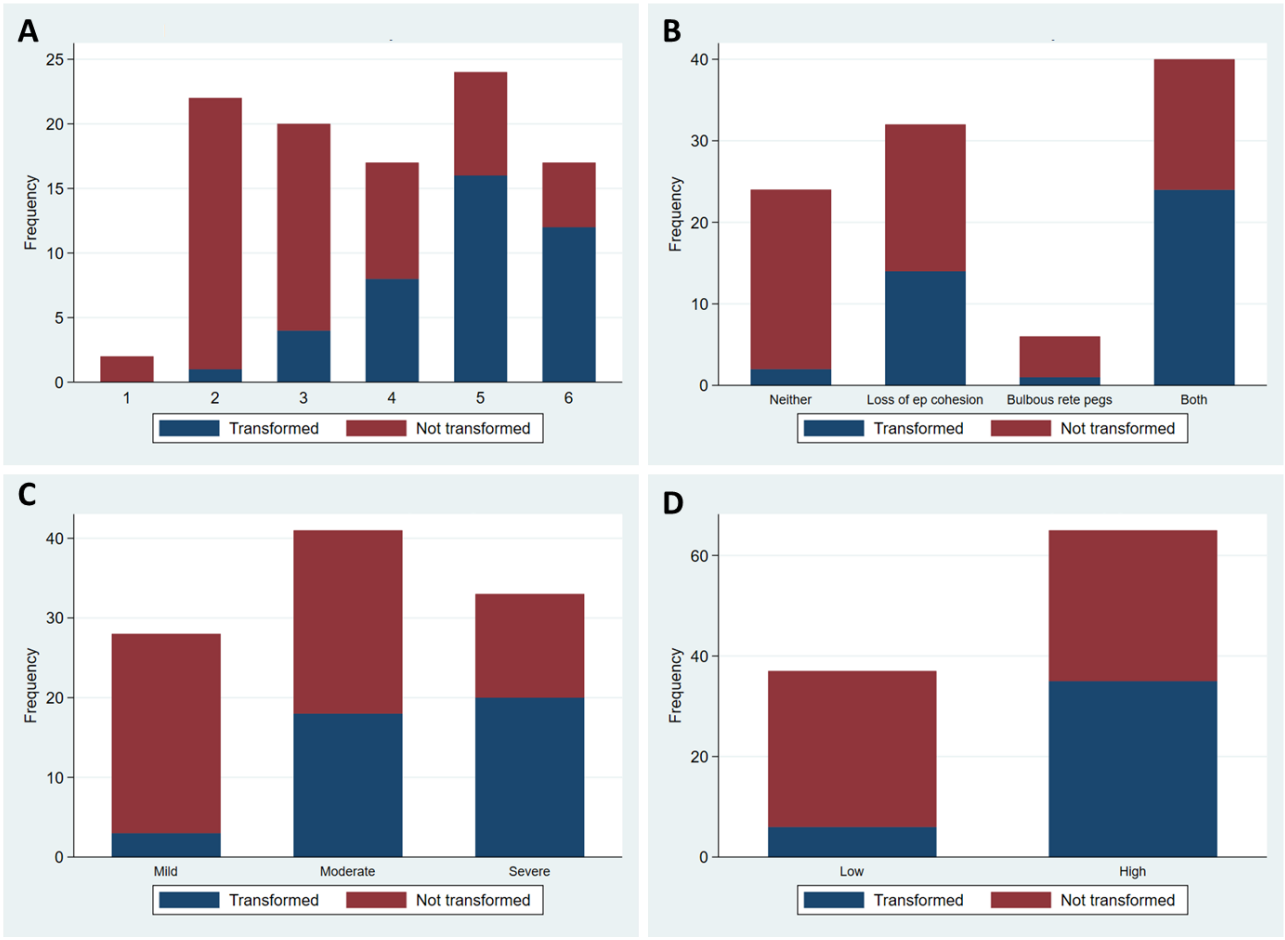

Figure 15. *Malignant transformation incidence in relation to scoring approaches.*

A) Incidence by count of features using six-point system; B) Incidence by count of features using two-point system; C) Incidence by WHO grading; D) Incidence by binary grading. Bar key: Navy blue = Transformed, Maroon = not transformed.

Figure 16 demonstrates the AUROC curves for the 'six-point' and 'two-point' scoring approaches compared with the WHO and binary grading systems. All four systems showed significantly better than chance discrimination ($p<0.0001$), but with varying degrees of prognostic strength. The 'six-point' system demonstrated the highest among these with an AUROC of 0.81 and was statistically significantly higher than the 'two-point' system (AUROC=0.73, p=0.004), WHO grading system (AUROC=0.71, p=0.03) and binary grading system (AUROC=0.68, p=0.009). On univariate analyses, 'bulbous rete pegs' alone had an AUROC of 0.68, with 'loss of stratification' (AUROC 0.70) and 'suprabasal mitosis' (AUROC 0.69) being similar.

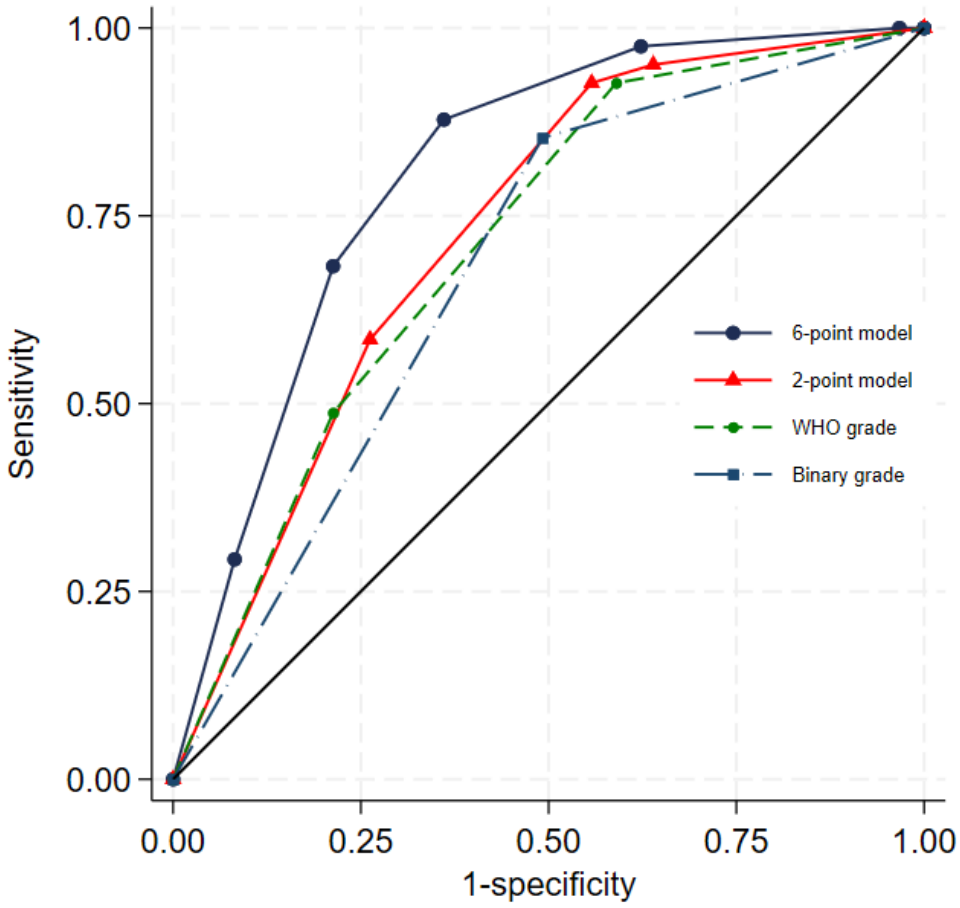

Figure 16. *AUROC for malignant transformation in relation to scoring approaches.*

All four systems showed significantly better than chance discrimination (p<0.0001), but with varying degrees of prognostic strength. The 'six-point' system demonstrated the highest among these with an AUROC of 0.81 and was statistically higher than the 'two-point' system (AUROC=0.73, p=0.004), WHO grading system (AUROC=0.71, p=0.03) and binary grading system (AUROC=0.68, p=0.009). Line Key: Navy blue = six-point model, Red = two-point model, Green = WHO grading, Pale Blue = Binary grading.

4.2.3.4 *Kaplan Meier Analysis for Time to Transformation*

Figure 17 shows the Kaplan Meier survival curves for time to malignant transformation for each of scoring approaches, in which the curves maintain the separation. Overall, an estimated 28% (95% CI 20 to 39%) had transformed within five years. For the 'six-point' model, the predicted transformation rate when all 6 features were present was 50% (95% CI 27 to 78%) compared to 35% (95% CI 23 to 53%) when 4-5 features were present and 14% (95% CI 5 to 32%) when 2-3 features were present. No transformations were recorded among patients with fewer than 2 features present.

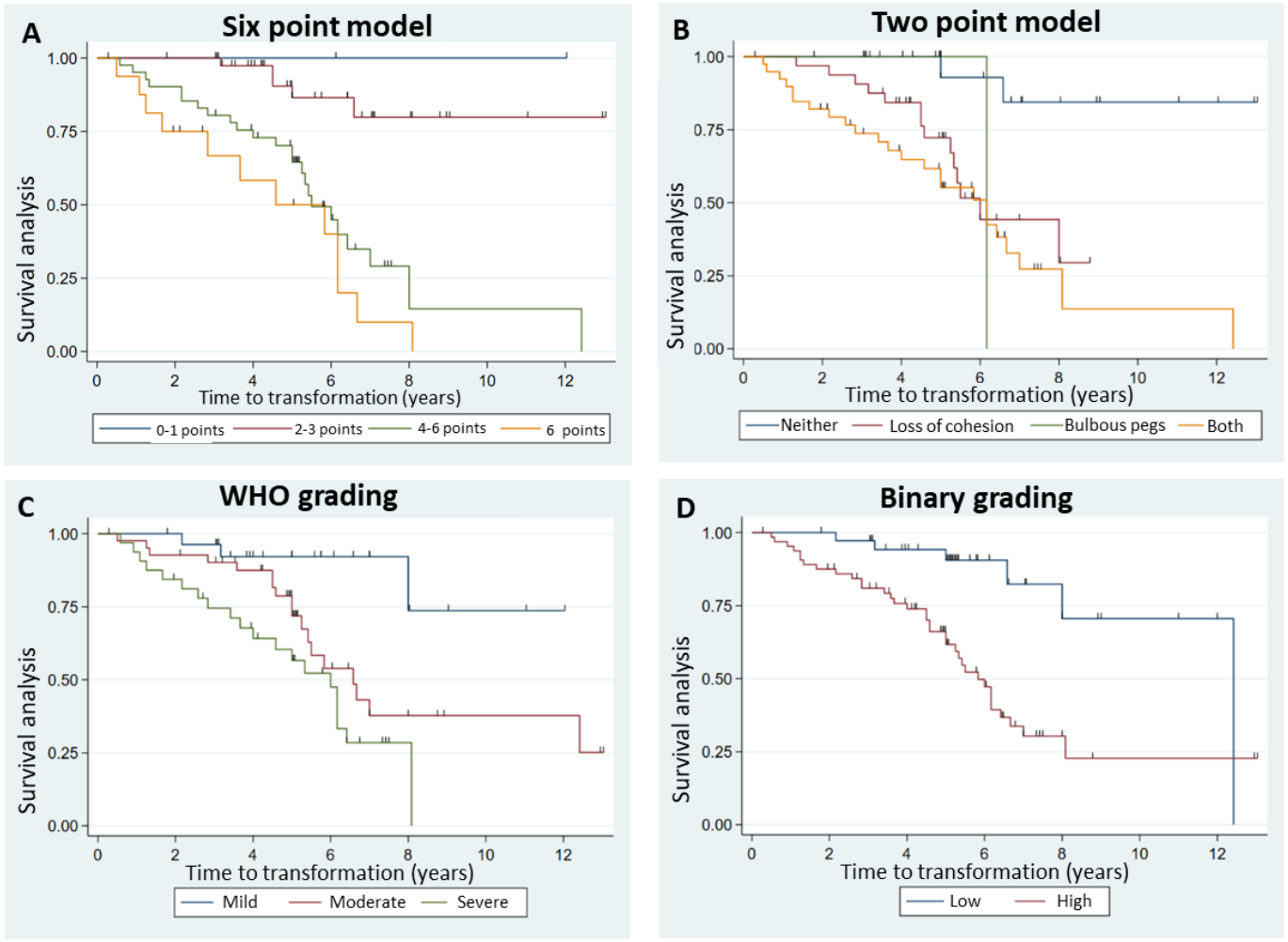

Figure 17. *Kaplan Meier curves for time to transformation for different scoring approaches.*

An estimated 28% (95% CI 20 to 39%) had transformed within five years. For the 'six-point' model, the predicted transformation rate when all 6 features were present was 50% (95% CI 27 to 78%) compared to 35% (95% CI 23 to 53%) when 4-5 features were present and 14% (95% CI 5 to 32%) when 2-3 features were present. No transformations were recorded among patients with fewer than 2 features present.

4.2.3.5 *Effect of Clinical Characteristics on Prognostic Models*

The models were also fitted with clinical characteristics (age, sex, histological grading) to see if any combination could improve predictive ability as characterised by the AUROC (Table 7). The 'six-point' model retained the best predictive ability compared to other models, with minor improvements when combined with other histological grading systems and clinical variables. The 'two-point' model along with the WHO and binary grading systems demonstrated modest improvements when combined with other clinical variables.

Table 7. *AUROC by scoring system with inclusion of additional variables.*

Models	AUROC
6-point model	0.807
6-point + age + sex	0.814
6-point + WHO grade	0.813
6-point + binary grade	0.820
2-point model	0.729
2-point + age + sex	0.742
2-point + WHO grade	0.763
2-point + binary grade	0.761
WHO grade	0.714
WHO grade + age + sex	0.740
Binary grade	0.681
Binary grade + age + sex	0.728

Figure 18 demonstrates AUROC and 95% confidence intervals for the different scoring approaches with comparisons between the presented study and the previous one by Mahmood *et al.* (2022)[9].

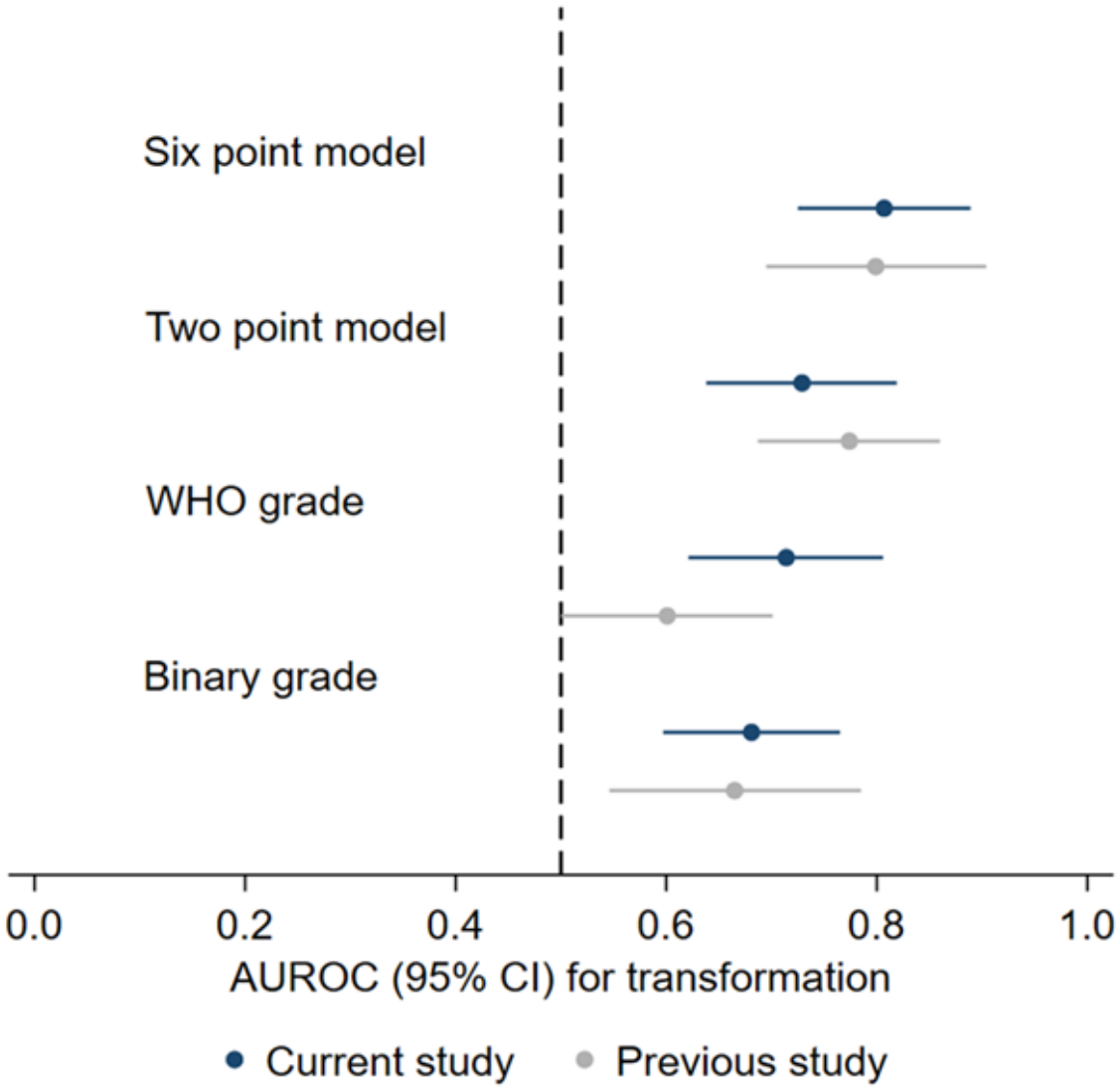

Figure 18. *Comparison of malignant transformation prediction between studies.*

4.2.3.6 *Univariate Associations of Individual Features*

Based on the assumption that each of the six histological features carry the same weight, the hazard ratios for the individual features are presented in Figure 19. Overall, the hazard ratios from the present study are similar to those observed in the previous study (Mahmood *et al.* 2022)[9] but do suggest that different features have different levels of association with malignant transformation risk. With regards to 'nuclear pleomorphism', since most cases had this feature present (92/102) and none of the remaining 10 cases transformed, it was not possible to reliably estimate a hazard ratio.

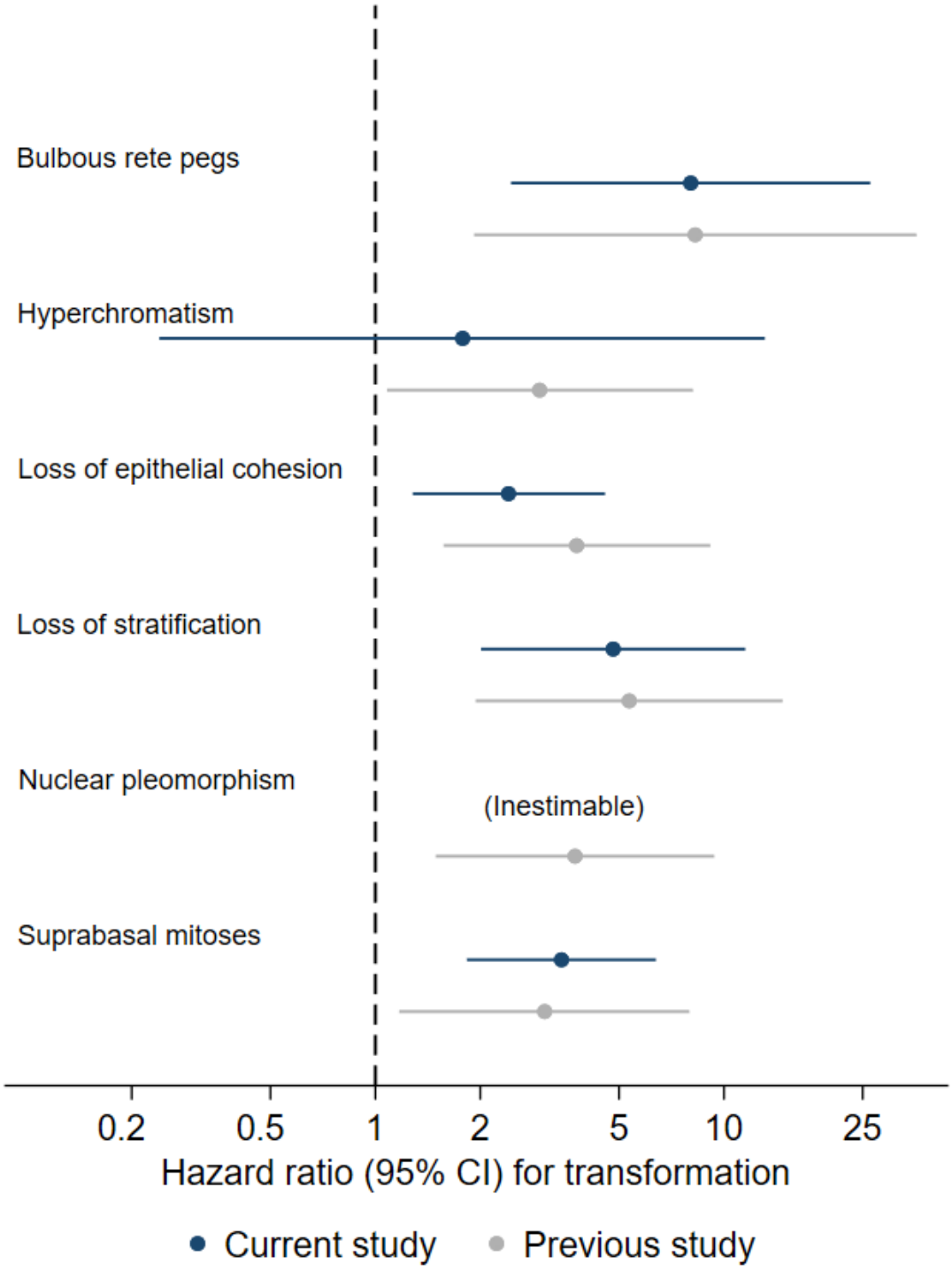

Figure 19. *Univariate association between time to transformation and individual features.*

Since the majority of cases (92/102) had nuclear pleomorphism, and none of the remaining 10 transformed, the hazard ratio could not be reliably estimated.

4.2.3.7 *Prognostic Performance by Assessors*

Similar to the previous study, the prognostic performance of the 'six-point' and 'two-point' models were calculated separately for each of the three raters, reflecting how the models are likely to be used in diagnostic practice (Table 8). The AUROC was almost identical for two of the three assessors, whilst the AUROC for the third rater was slightly lower. Overall, the 'six-point' model demonstrated better prognostic ability for the individual raters compared with the 'two-point' model.

Table 8. *Prognostic performance by individual assessor and overall.*

	Overall	Assessor 1	Assessor 2	Assessor 3
6-point model	0.81	0.81	0.80	0.75
2-point model	0.73	0.72	0.72	0.70

4.2.4 Discussion

This study conducts validity testing of the previously proposed 'six-point' and 'two-point' scoring models for malignant transformation prediction in OED. Findings demonstrate the 'six-point' model as being the most predictive system (AUROC 0.81) compared to the 'two-point' model, WHO and binary grading systems (AUROC 0.68-73). Whilst the 'two-point' model did perform better than the WHO and binary grading systems, this was to a lesser extent (AUROC 0.73). The addition of clinical variables further improved model performance for both the 'two'- and 'six-point' models, albeit marginally.

A key aspect of histological grading is to assist clinicians with the management of individual patients with OED. However, it is evident from this study that existing grading systems are less reliable at predicting prognosis, and there is a need for improvement. Currently, diagnosis and grading of OED using the 'gold standard' WHO (2017) classification relies on identification of a wide range of histological architectural and cytological features (28 in total) but this approach places little value on the strength of individual features. The 'six-point' model was first developed by Mahmood *et al.* (2022) using features identified as having adequate inter-rater agreement and a statistical association with clinical outcomes. This model demonstrates that, assessment of just six histological features with attributed strength, can significantly improve predictive ability, and may be more prognostically reliable than identification of a wider range of features.

Having said this, when using the 'six-point' model, the allocation of single point per feature implies that the six features each have the same importance, a choice that was made pragmatically on two grounds: to avoid overfitting to the data, and for ease of use of the eventual score. The assumption of equal weighting has been verified by evaluating the hazard ratio of each feature individually (Figure 19) demonstrating that different features have different levels of association with malignant

transformation. Therefore, any future improvements of this scoring model may consider giving different weights to the individual features, or even removing 'redundant' features entirely from the model. That said, the incremental benefit of doing so for the data presented in this study would have been relatively small. Re-fitting the model to allow the different weights for these features produced an AUROC of 0.82, compared to the 0.81 observed from allocating equal weight to the six individual features.

Interrater reproducibility is a potentially important feature of grading, however, reproducibility does not necessarily reflect accuracy, because it is possible to be reproducibly incorrect. In the present study, despite the considerable interobserver variability for certain histological features (i.e. 'loss of epithelial cohesion' - Gwet's AC1 0.29; 'suprabasal mitosis' - Gwet's AC1 0.06) (Table 4), the 'six-point' model still retained good prognostic strength, further supporting its robustness and credibility as a prognostic tool.

Another point of consideration for the newly proposed models is that they are based on the consensus of three raters; for instance, 'bulbous rete pegs' is defined as present if at least two of the three assessors considered it present, and otherwise absent. This may well overestimate how well the model works in practice where an assessment is made by a single pathologist; introducing measurement error will typically lessen associations. To address this, the 'two'- and 'six-point' models were calculated separately for each assessor (Table 8) and findings have shown that the 'six-point' model has better prognostic ability for the individual raters compared with the 'two-point' model.

The addition of clinical variables only marginally improved model performance (Table 7), supporting findings from the previous study by Mahmood *et al.* (2022)[9]. Due to the retrospective nature of the study, it was not possible to consistently and reliably obtain social history information, therefore the impact of smoking, tobacco use and alcohol on prognostic ability could not be investigated. Prospective validation of the proposed models should aim to consider evaluation of these variables, which are known to be important risk factors for OED and subsequent cancer development.

This study independently validates the prognostic ability of the 'six-point' model for OED progression on a multicentric sample of cases. Our findings highlight that the presence of a smaller number of features with high prognostic strength, may be potentially more important than the presence of a larger number of features of low to moderate prognostic strength. It is hoped that this approach will provide pathologists, clinicians and patients with more reliable information about the behaviour and progression of OED lesions and guide treatment decisions. Based on the findings from this study, the authors encourage the informal integration of the 'six-point' model alongside existing grading systems. Meanwhile, a larger scale multi-centre prospective study is required prior to establishing its formal use in clinical practice.

DATA AVAILABILITY

All data generated or analysed during this study are included in this published article.

ACKNOWLEDGEMENTS

The authors acknowledge all centres involved in this study for contribution of cases for model validation.

AUTHOR CONTRIBUTIONS

HM and SAK conceived and designed the study. HM prepared the validation cohort, collected, and organised the clinical data. NMI, SAK and HM undertook independent histological examination. MB provided statistical expertise. HM, MB, NMI and SAK were involved in analysis of results. HM led on writing of the manuscript with contributions from all authors.

FUNDING

HM is funded by the National Institute for Health Research (Award ID: NIHR300904). SAK acknowledges funding from Cancer Research UK (C63489/A29674).

COMPETING INTERESTS

SAK is a shareholder of Histofy, an AI start-up in computational pathology. The remaining authors do not declare any competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval was granted by the West Midlands - Edgbaston Research Ethics Committee (reference: 18/WM/0335).

Chapter 5 - Novel Morphometric Digital Feature Analysis

Chapter 5 – Novel Morphometric Digital Feature Analysis

This chapter has been reproduced from a peer-reviewed publication for which the candidate is first and primary contributing author. This study conducts a digital quantitative analysis of cellularity, nuclei morphometry, cell cytoplasm colour intensity and thickness/perimeter in OED epithelium and explores the prognostic value of these features to develop a predictive model for OED recurrence and malignant progression. Independent external validation of the proposed model was conducted using unseen datasets from three other national and international centres.

The candidate's contributions to this study were:

i. obtaining ethical approval
ii. slide retrieval, dataset preparation and clinical data collection
iii. digital quantitative and statistical analysis
iv. multivariable model development, testing and validation
v. writing up the manuscripts

The study was conceptualised with guidance from supervisors.

Datasets for model validation were acquired from external national and international centres including: 1) Precision Medicine Centre, Patrick G. Johnston Centre for Cancer Research, Queen's University Belfast, UK; 2) Institute of Head and Neck Studies and Education (InHANSE), Birmingham, UK and 3) Piracicaba Dental School, UNICAMP, Brazil.

Statistical support was provided by Mike Bradburn, senior statistician at the Clinical Trials Research Unit, University of Sheffield.

5.1 Development and validation of a multivariable model for prediction of malignant transformation and recurrence of oral epithelial dysplasia

ARTICLE OPEN

Check for updates

Development and validation of a multivariable model for prediction of malignant transformation and recurrence of oral epithelial dysplasia

Hanya Mahmood[1✉], Adam Shephard[2], Paul Hankinson[3], Mike Bradburn[4], Anna Luiza Damaceno Araujo[5], Alan Roger Santos-Silva[5], Marcio Ajudarte Lopes[5], Pablo Agustin Vargas[5], Kris D. McCombe[6], Stephanie G. Craig[6], Jacqueline James[6], Jill Brooks[7], Paul Nankivell[7], Hisham Mehanna[7], Nasir Rajpoot[2,8] and Syed Ali Khurram[3,8]

BACKGROUND: Oral epithelial dysplasia (OED) is the precursor to oral squamous cell carcinoma which is amongst the top ten cancers worldwide. Prognostic significance of conventional histological features in OED is not well established. Many additional histological abnormalities are seen in OED, but are insufficiently investigated, and have not been correlated to clinical outcomes.
METHODS: A digital quantitative analysis of epithelial cellularity, nuclear geometry, cytoplasm staining intensity and epithelial architecture/thickness is conducted on 75 OED whole-slide images (252 regions of interest) with feature-specific comparisons between grades and against non-dysplastic/control cases. Multivariable models were developed to evaluate prediction of OED recurrence and malignant transformation. The best performing models were externally validated on unseen cases pooled from four different centres ($n = 121$), of which 32% progressed to cancer, with an average transformation time of 45 months.
RESULTS: Grade-based differences were seen for cytoplasmic eosin, nuclear eccentricity, and circularity in basal epithelial cells of OED ($p < 0.05$). Nucleus circularity was associated with OED recurrence ($p = 0.018$) and epithelial perimeter associated with malignant transformation ($p = 0.03$). The developed model demonstrated superior predictive potential for malignant transformation (AUROC 0.77) and OED recurrence (AUROC 0.74) as compared with conventional WHO grading (AUROC 0.68 and 0.71, respectively). External validation supported the prognostic strength of this model.
CONCLUSIONS: This study supports a novel prognostic model which outperforms existing grading systems. Further studies are warranted to evaluate its significance for OED prognostication.

British Journal of Cancer; https://doi.org/10.1038/s41416-023-02438-0

BACKGROUND

Oral epithelial dysplasia (OED) is a 'pre-cancerous state' histologically characterised by cellular atypia with loss of normal maturation and stratification of stratified squamous epithelium [1, 2]. Its progression to malignancy (oral squamous cell carcinoma or OSCC) is a progressive multi-step process which can be initiated by chemical carcinogen exposure (such as tobacco) [3], genetic mutations [4–6] and in a small subset of cases, by high-risk human papilloma virus (HPV) [7]. The progression of OED to OSCC is variable (mild OED 1.7%; severe OED 3.57%, annually) [8, 9] and difficult to predict due to poor understanding of the disease pathway [10, 11].

Conventionally, a diagnosis of OED is reached following identification of a wide range of histological architectural (whole epithelium) and cytological (individual keratinocyte) abnormalities using the World Health Organisation (WHO) criteria [12]. This three-tier grading system (mild, moderate, severe) was recently updated from the fifth edition of the WHO classification to further expand the range of diagnostic features to twenty-seven in total (14). However, the prognostic strength of these features remains poorly understood [13, 11] and, individually, many of them are relatively non-specific [11] and evident in a host of other non-dysplastic conditions (such as reactive atypia in inflammatory and ulcerative conditions or fungal infections) [11, 14]. As such, conventional grading is an unreliable predictor of cancer risk, further complicated by inter and intra-observer inconsistencies [15], variations in interpretation of findings [11], and alternative proposed grading systems [16, 17]. Grading should, therefore, not be used as a sole indicator for treatment selection.

In addition to the 'conventional' OED features, there are also a range of other features seen in OED, which are not routinely quantified or analysed by the pathologist, nor known to be

[1]Academic Unit of Oral & Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield, Sheffield, UK. [2]Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, Warwick, UK. [3]Unit of Oral & Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, Sheffield, UK. [4]Clinical Trials Research Unit, School of Health and Related Research, University of Sheffield, Sheffield, UK. [5]Oral Diagnosis Department, Piracicaba Dental School, University of Campinas (UNICAMP), São Paulo, Brazil. [6]Precision Medicine Centre, Patrick G. Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, UK. [7]Institute of Head and Neck Studies and Education, Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK. [8]These authors jointly supervised this work: Nasir Rajpoot, Syed Ali Khurram.
✉email: h.mahmood@sheffield.ac.uk

Received: 15 April 2023 Revised: 2 September 2023 Accepted: 12 September 2023
Published online: 27 September 2023

correlated to clinical outcomes. Such features include alteration in cell numbers, differences in lesion architecture and thickness, variations in nuclei geometry and staining intensity of cell cytoplasm. The importance of these features in OED progression to malignancy has not been given much attention, perhaps due to difficulty in their visual assessment using conventional microscopy methods, and the time consuming and laborious nature of cellular level analysis.

Whilst several studies have focussed on the strength of grading alone, it is important to acknowledge that the 'global' grade is not always representative of feature severity, nor does it consider clinical variables (such as age, gender or clinical site). More recently, the 'six-point' and 'two-point' prognostic models were developed using cytological and architectures features associated with malignant transformation and recurrence with good inter-observer agreement [18]. The authors found that the strength of these models increased when combined with histological grading and clinical characteristics, outperforming conventional grading systems alone. This highlights the need to further explore prognostic associations of novel histological variables in a similar manner, through development and testing of multivariable models. With the advancement of digital pathology methods, it is now possible to conduct detailed quantitative histological analyses of digitised whole-slide images (or WSI) of Haematoxylin and Eosin (H&E) stained tissue sections using computer-assisted approaches [19]. WSIs contain large volumes of data which can be useful for exploration of prognostic markers. Digital image analysis allows automated detection of cell nuclei and subsequent quantification assessment of subcellular compartments [20] generating data in an objective and reproducible manner for downstream analysis.

This study consists of three parts. First, we conduct a digital quantitative analysis of cellularity, nuclei morphometry, cell cytoplasm colour intensity and thickness/perimeter in OED epithelium, to explore differences between dysplasia grades and non-dysplastic oral epithelium. Secondly, we explore the prognostic value of these features and develop a predictive model for OED recurrence and malignant progression. Finally, we conduct external validation of the proposed model using three independent datasets from other national and international centres.

Whilst the application of digital image analysis to study oral premalignant disorders is increasing, few studies have applied these methods for exploration of histological predictors in OED. This study provides novel insight into OED progression, identifying new and potentially important features for clinical outcome prediction.

MATERIALS AND METHODS

Training dataset and clinical data

A retrospective sample of 75 OED cases (one representative H&E slide per case) were used for quantitative histological feature analysis and development of the multivariate predictive models. Where feature-specific comparisons are made, a control sample of 25 non-dysplastic oral tissue sections (including hyperplasia and traumatic hyperkeratosis) were used. These slides were obtained from the local pathology archive (School of Clinical Dentistry, University of Sheffield, UK, dating 2008–2013). Purposive sampling was used to include equal numbers for mild, moderate and severe grades and controls ($n = 25$ each). A minimum of 5-year clinical follow-up was required. Cases were also re-graded using the binary OED grading system.

Prior to the inclusion of cases, slides were independently reviewed by two pathologists (SAK, PH) to ensure tissue sections were of suitable quality for analysis. HPV-related OED and verrucous lesions were excluded, based on morphological analysis, as they are distinct entities with reportedly different features and behaviour. Cases were also excluded if (1) there was no associated H&E slide, (2) there was insufficient epithelial tissue for analysis, (3) the slide was of poor staining quality, appeared distorted/blurred, had tissue artefacts/folds or (4) there was incomplete or irretrievable follow-up data. Ethical approval was obtained (18/WM/0335) and experimental methods were conducted in line with the Declaration of Helsinki.

Clinical data were obtained from patient case notes and various online hospital clinical systems. Collected data included demographic details and relevant diagnostic information including intra-oral biopsy site, original histological OED grade (WHO, 2017), treatment information, and whether the lesion had transformed or recurred. The clinicians abstracting this data were blinded to patient outcomes. We defined transformation as a dysplastic lesion which had progressed to OSCC at the same clinical site, and recurrence as a dysplastic lesion which had occurred at the same clinical site following surgical excision within the follow-up period.

Quantitative histological feature analysis

New 4 µm H&E sections of the selected cases were digitised to high-resolution WSIs using Aperio CS2 (Leica Biosystems, Germany) and Hamamatsu NanoZoomer 360 (Japan) scanners. QuPath (version 0.3.0) bioimage analysis software [21] was used for quantitative histological analysis. This platform was chosen due to its powerful annotation, visualisation and built-in cell and nuclear detection tools [21], in addition to its reported reproducibility in tissue-based biomarker studies [22].

Regions of interest (ROIs) corresponding to histologically representative areas were identified for each image and confirmed by several authors (HM, PH, SAK). Within each ROI, the full thickness of the epithelium was demarcated for localised quantitative analysis. For consistency, a minimum of three and a maximum of four ROIs were selected per whole-slide image, with a closely matched area ($300{,}000 \pm 400\ \mu m^{\wedge 2}$). A standardised cell detection threshold of 0.04 was set (at ×10 magnification) and other default parameters were kept consistent across all ROIs. The cell detection algorithm in QuPath was utilised to quantify cell numbers and extract a range of other features as outlined below:

i. Number of epithelial cells
ii. Nuclear geometry: nuclear circularity, nuclear eccentricity, nucleus to cell area ratio
iii. Staining intensity: nuclear haematoxylin optical density (OD), cytoplasm eosin OD
iv. Epithelium architecture/thickness: Perimeter (µm)

Due to the inherent nature of OED, ROI-level analysis was considered better than WSI-level analysis, as dysplasia is not always visualised throughout the tissue section. In total, 325 ROIs were generated for analysis and model development (mild OED $n = 84$, moderate OED $n = 83$, severe OED $n = 85$, control $n = 73$). Extracted data were systematically recorded in an Excel spreadsheet (Microsoft Corporation, 2018).

Measuring cell detection accuracy

To assess the reliability of QuPath's cell detection algorithm, we conducted a measure of accuracy based on 10 ROIs (Fig. 1). Ground-truth nuclear segmentations within these 10 ROIs were generated semi-automatically. An expert oral pathologist (SAK) clicked on each nucleus in the 10 ROIs. These nuclear 'clicks' were then passed through NuClick [23], a deep learning model that takes 'click' inputs as a guiding signal to generate nuclear boundaries. These segmentations were then further refined manually, where necessary, to ensure accurate nuclear segmentations. On comparison of the QuPath nuclear segmentations to the ground-truth segmentations, the results were promising, producing a raw Dice score of 0.73 (nuclei vs. background), a detection quality score of 0.67, a segmentation quality score of 0.72 and a panoptic quality of 0.49. These scores are in line with more recent nuclear instance segmentation papers such as HoVer-Net [24], HoVer-Net+ [25] and PanNuke [26], where Dice score and segmentation quality measure the performance of nuclei segmentation and detection quality, panoptic quality and aggregated jaccard index give measures of the individual nuclear detections.

Development and validation of prognostic models

Multivariable logistic regression analysis was conducted to model the different histological features for prediction of malignant transformation and OED recurrence. The models were developed using different combinations of parameters, and for clinical interest we did not restrict these combinations to only features that were statistically significant. Model performance was visualised by measuring the Area Under the Receiver Operating Characteristic (AUROC) curve.

External validation was performed using further OED cases pooled from four different centres ($n = 121$, 287 ROI: 79 mild, 106 moderate and 112 severe) (Table 1); these cases were not part of the original model

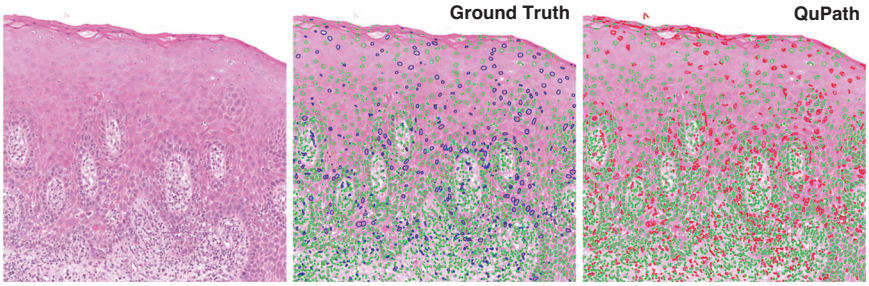

Fig. 1 A comparison of the ground truth (middle) vs. QuPath's cell detection tool (right) for nuclear detection and segmentation. The raw image is displayed on the left. Colour key: green nuclei = true positives; blue nuclei = false negatives; red nuclei = false positives.

Table 1. Cohorts for model development and validation with respective grade (WHO, 2017) breakdown and clinical outcomes.

	N	Transformation	Recurrence
DEVELOPMENT COHORT			
School of Clinical Dentistry, Sheffield, UK			
Mild	25	1 (1%)	2 (3%)
Moderate	25	9 (12%	11 (15%)
Severe	25	9 (12%)	13 (17%)
Total	*75*	*19 (25%)*	*26 (35%)*
Overall Total	***75***		
VALIDATION COHORT			
School of Clinical Dentistry, Sheffield, UK			
Mild	4	0 (0%)	0 (0%)
Moderate	6	2 (13%)	2 (13%)
Severe	5	3 (20%)	2 (13%)
Total	*15*	*5 (33%)*	*4 (27%)*
Piracicaba Dental School, UNICAMP, Brazil			
Mild	8	2 (11%)	1 (5%)
Moderate	7	1 (5%)	1 (5%)
Severe	4	1 (5%)	1 (5%)
Total	*19*	*4 (21%)*	*3 (15%)*
Queen's University Belfast, UK			
Mild	1	1 (2%)	Not recorded
Moderate	16	10 (22%)	
Severe	28	14 (31%)	
Total	*45*	*25 (55%)*	
Institute of Head and Neck Studies and Education (InHANSE), Birmingham, UK			
Mild	20	0 (0%)	Not recorded
Moderate	14	3 (7%)	
Severe	8	2 (5%)	
Total	*42*	*5 (12%)*	
Overall Total	***121***		

The total number in the development and validation cohort are in bold.

development. 60 patients were female (49.5%) and 61 were male (50.5%), with a mean age of 59 years (S.D 12.55). Intra-oral sites included ventral/lateral tongue (60%, $n = 72$), buccal mucosa (14%, $n = 17$), floor of mouth (16%, $n = 20$), palate (6%, $n = 7$) and alveolar ridge (4%, $n = 5$). Independent grading confirmed 27% mild ($n = 33$), 36% moderate ($n = 43$) and 37% severe ($n = 45$) OED lesions. Binary grading confirmed 43 low grade (36%) and 78 high grade (64%) lesions. 39 lesions progressed to OSCC (32%), with an average time to transformation of 45 months (median 42, S.D 33, IQR 54) (Table 1). Amongst the 121 cases, only 34 cases were used for validation of the recurrence model; the remainder could not be used to reliably predict recurrence status due to incomplete follow-up data. The average time to recurrence was 19 months (median 12, S.D 17, IQR 28).

Statistical analysis

GraphPad Prism® statistical software (version 9.3.1) was used for analysis. Descriptive statistics were performed for all histological variables. Continuous data were tested for normality using Shapiro–Wilk or D'Agostino & Pearson tests. Where normal distribution was assumed, an unpaired two-tailed *T*-test or one-way ANOVA with an applicable post-hoc analysis (Tukey's or Dunnett's) was performed for pairwise comparisons. Individual histological feature associations with clinical outcomes were determined using binary logistic regression, and multivariate regression analysis for the development and testing of the models. Prognostic discrimination was visualised by AUROC curves with a 95% confidence interval. All tests were two-tailed, and *p* values adjusted for multiple comparisons testing.

RESULTS

Quantitative analysis and model development

Amongst the OED dataset used for quantitative feature analysis and model development ($n = 75$), independent grading assessment confirmed 25 each for mild, moderate and severe WHO grades (Table 1). Binary grading revealed 33 low grade and 42 high grade lesions. 47 patients (63%) were male and 28 (37%) females with a median age of 65 years (IQR 21). Intra-oral sites included the floor of mouth ($n = 15$, 20%), buccal/labial mucosa ($n = 13$, 17%), tongue ($n = 39$, 52%), gingivae ($n = 4$, 5%) and hard/soft palate ($n = 4$, 5%). 25% of lesions ($n = 19$) progressed to OSCC, with an average transformation time of 31 months (median 24, S.D 26.14, IQR 48). 35% of lesions recurred ($n = 26$) with an average recurrence time of 35 months (median 24, S.D 35.42, IQR 24) (Table 1).

Quantitative analysis of cellularity

Increased epithelial cellularity was seen in OED compared to control (mean cell number: mild OED 1773 [95% CI 1541–2005], moderate OED 1776 [95% CI 1637–1915], severe OED 1909 [95% CI 1600–2076] vs. control 1508 [95% CI 1302–1728]), though these differences were not significant (Fig. 2). In contrast, the cellularity in OED was reduced in the basal epithelial layer of OED compared to control (mild OED + moderate OED + severe OED vs. control $p = 0.02$; low grade OED + high grade OED vs. control $p = 0.01$). Further differences were observed between individual grades: moderate OED vs. control ($p = 0.02$, 95% CI 5.66–90.99) and high grade OED vs. control ($p = 0.007$, 95% CI 11.75–82.87) (Fig. 2). There was no statistical association between epithelial cellularity (in the full thickness of the epithelium or basal epithelial layer) with clinical outcomes (Table 2).

Quantitative analysis of nuclear and cytoplasmic features

Grade-related differences in cytoplasmic eosin OD were seen in OED epithelium (mild OED vs. moderate OED vs. severe OED $p = 0.03$ and low grade vs. high grade OED $p = 0.02$) with higher

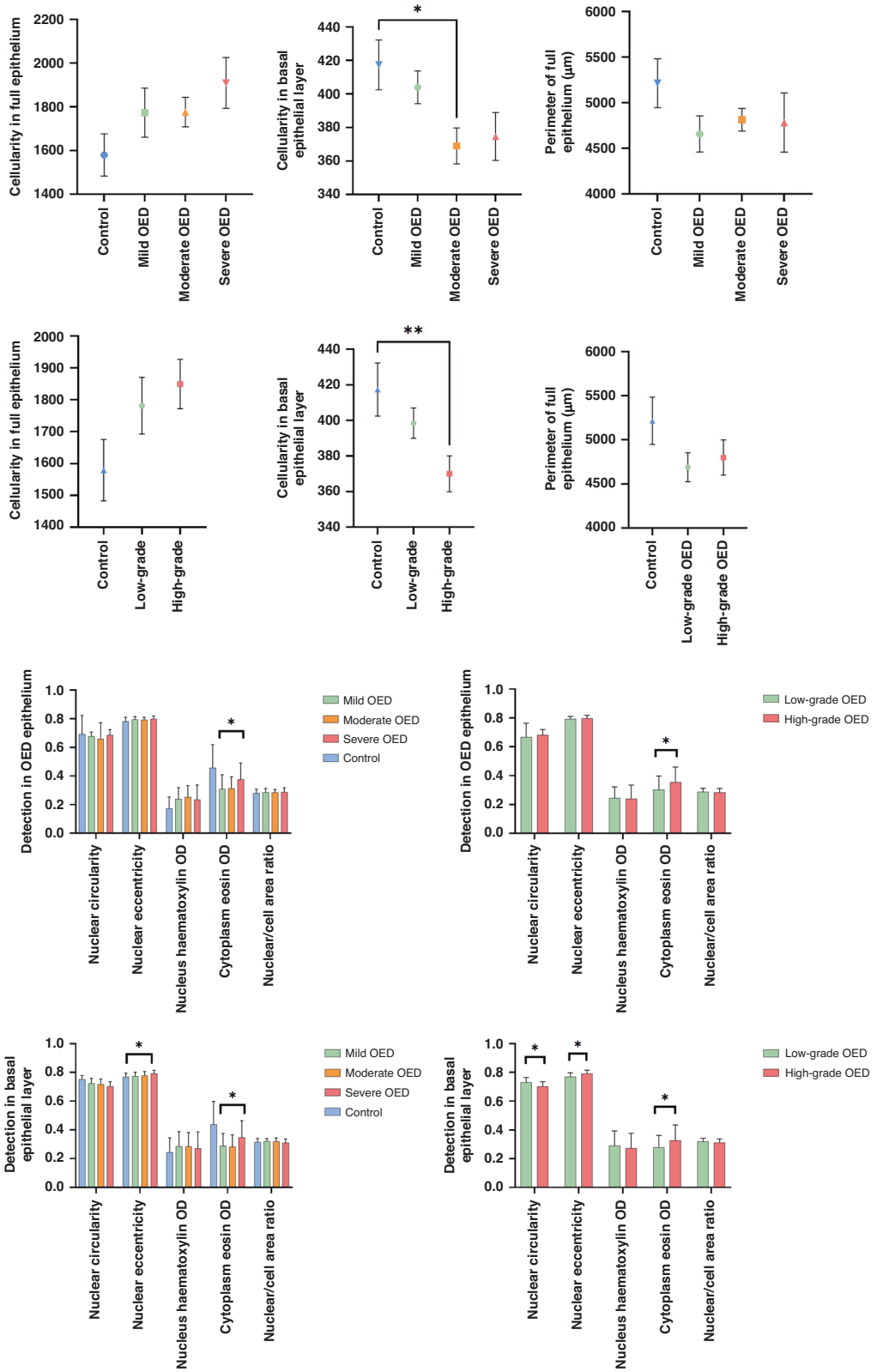

Fig. 2 Grade-wise analysis of cellularity, nuclear and cytoplasmic features and perimeter in OED with comparison to control. Displayed values represent mean values ± standard error. Asterisk denotes statistically significant result (* $p \leq 0.05$, ** $p \leq 0.01$).

detection levels in the more severe lesions (Fig. 2). No significant grade-related differences were seen for nuclear circularity, nuclear eccentricity, nuclear haematoxylin OD or nucleus/cell area ratio (Fig. 2). Basal epithelial layer analysis demonstrated significant differences between WHO grades for nuclear eccentricity ($p = 0.02$) and cytoplasm eosin ($p = 0.04$). These two features were also significant between binary grades ($p = 0.0004$ and $p = 0.04$, respectively) in addition to nuclear circularity

Table 2. Statistical association of individual histological features with clinical outcomes for the full thickness of the epithelium and basal epithelial layer ($n = 75$).

Event	Full epithelium		Full epithelium		Basal epithelial layer		Basal epithelial layer	
	MT		R		MT		R	
	AUROC	*p* value	AUROC	*p* value	AUROC	*p* value	AUROC	*p* value
Cellularity	0.5075	0.7877	0.5188	0.841	0.5241	0.7824	0.5137	0.6818
Perimeter	0.6118	0.0336*	0.5683	0.1012	*not assessed*			
Nucleus circularity	0.6471	0.0645	0.6903	0.0180*	0.535	0.7659	0.5534	0.6668
Nucleus eccentricity	0.6405	0.0639	0.5687	0.2975	0.5145	0.9563	0.5573	0.2392
Nucleus haematoxylin OD	0.5348	0.8844	0.5553	0.5495	0.5464	0.6162	0.5522	0.5059
Cytoplasm eosin OD	0.5531	0.4115	0.5856	0.1747	0.5559	0.5787	0.5895	0.1915
Nuclear/cell area ratio	0.5428	0.2621	0.5114	0.6103	0.5423	0.3326	0.5353	0.3967

Asterisk denotes statistically significant result (* $p \le 0.05$, ** $p \le 0.01$).
MT malignant transformation, *R* OED recurrence *AUROC* area under receiver operating characteristic.

($p = 0.0005$) (Fig. 2). There was no statistical association between individual nuclear and cytoplasmic features (for either full epithelium or basal epithelial layer) and malignant progression. However, nuclear circularity in the full epithelium was associated with OED recurrence ($p = 0.02$, AUROC 0.69, 95% CI 0.56–0.82) (Table 2).

Quantitative analysis of thickness/perimeter of OED epithelium

There were no significant differences in the perimeter/thickness of the full epithelium between OED grades (WHO or binary) or comparison to controls (Fig. 2). However, there was a statistical association between the perimeter of OED epithelium and malignant transformation ($p = 0.02$, AUROC 0.61 with 95% CI 0.47–0.75) (Table 2).

Prediction of malignant transformation

Comparisons between the various models demonstrated that increasing the number of histological variables strengthened the models' predictive performance, both for malignant transformation and OED recurrence (Table 3). Model 6 ("epithelial cellularity" + "nuclear circularity" + "nuclear eccentricity" + "nucleus haematoxylin OD mean" + "cytoplasm eosin OD mean" + " nuclear/cell area ratio" + "perimeter of epithelium") showed good prognostic value for prediction of malignant transformation (AUROC 0.77, 95% CI 0.65–0.90, $p = 0.0004$) compared to models with fewer histological variables (Table 3, Fig. 3). This model demonstrated a negative predictive power of 81.25% and a positive predictive power of 63.64%. The odds ratios for individual features formulating this model are presented in Supplementary Table 1. The strength of this model was further improved by incorporating histological grading systems. The addition of "WHO grading" (model 7) increased AUROC to 0.85 (95% CI 0.74–0.96, $p < 0.0001$), and the addition of "Binary grading" (model 8) produced an AUROC of 0.86 (95% CI 0.77–0.96, $p < 0.0001$) (Table 3, Fig. 3). These models performed better than WHO grading (AUROC 0.68, 95% CI 0.56–0.81, $p = 0.014$) and binary grading systems alone (AUROC 0.72, 95% CI 0.60–0.84, $p = 0.003$) for prediction of malignancy.

External validation was conducted on 121 additional cases using the same features as in model 6. Doing so showed a similar AUROC of 0.76 (95% CI 0.68–0.85, $p < 0.0001$) for malignant transformation (Table 3).

Prediction of OED recurrence

Model 6 ("epithelial cellularity" + "nuclear circularity" + "nuclear eccentricity" + "nucleus haematoxylin OD mean" + "cytoplasm eosin OD mean" + "nuclear/cell area ratio" + "perimeter of epithelium") also demonstrated good performance for prediction of OED recurrence with an AUROC of 0.74 (95% CI 0.62–0.87, $p = 0.0006$) in comparison to models 1–5 where AUROC ranged between 0.69 and 0.72 (Table 3, Fig. 3). This model demonstrated a negative predictive power of 74.58% and a positive predictive power of 68.75%. The odds ratio for individual features formulating this model are presented in Supplementary Table 2. Similar to the malignant transformation models, the incorporation of current grading systems increased model performance. The addition of "Binary grading" (model 8) increased AUROC to 0.81 (95% CI 0.72–0.91, $p < 0.0001$), and addition of "WHO grading" (model 7) optimised the performance further, yielding an AUROC of 0.82 (95% CI 0.72–0.91, $p < 0.0001$) (Table 3, Fig. 3).

External validation was limited to 34 individuals with confirmed recurrence status. Model 6 retained its superior performance with an AUROC of 0.93 (95% CI 0.81–1.00, $p = 0.0005$) for OED recurrence (Table 3).

DISCUSSION

In this study, we use digital image analysis to explore and extract quantitative data for several histological features (cell number, nuclear and cytoplasm geometric and intensity features, lesion thickness/perimeter) in OED epithelium to determine their diagnostic importance and relationship with clinical outcomes. We focussed the analysis on the full thickness of the epithelium, as opposed to conventional epithelial 'thirds' used with WHO grading, to remove layer restriction and subjectivity.

The unique aspect of this study is the development of multivariate models for outcome prediction using the digitally quantified histological data. Our findings demonstrated that the combination of all the major digital histological features (Model 6: "epithelial cellularity" + "nuclear circularity" + "nuclear eccentricity" + "nucleus haematoxylin OD mean" + "cytoplasm eosin OD mean" + "nuclear/cell area ratio" + "perimeter of epithelium") was associated with greater predictive performance for both malignant transformation and OED recurrence in comparison to conventional histological grading systems (Table 3, Fig. 3). Model 6 yielded good predictive performance for malignant transformation (AUROC of 0.77, 95% CI 0.64–0.90, $p = 0.0004$) and OED recurrence (AUROC of 0.74, 95% CI 0.61–0.86, $p = 0.0006$) which exceeded that of WHO grading (AUROC 0.69, $p = 0.01$) and binary grading (AUROC 0.72, $p = 0.0037$) alone. External validation of model 6 supported its superior performance (AUROC of 0.76, 95% CI 0.68 –0.85, $p < 0.0001$ for malignant transformation and AUROC of 0.93, 95% CI 0.81–1.00, $p = 0.0005$ for OED recurrence) (Table 3). The

Table 3. Multivariate histological models for prediction of malignant transformation (MT) and OED recurrence (R).

Model	Histological Variables	Malignant Transformation			OED Recurrence		
		AUROC	95% CI	*p* value	AUROC	95% CI	*p* value
	"WHO Grade"	0.688	0.5643 to 0.8116	0.0148[a]	0.7159	0.5993 to 0.8324	0.0022[a]
	"Binary Grade"	0.7242	0.6022 to 0.8461	0.0037[a]	0.719	0.6001 to 0.8379	0.0019[a]
	"Epithelial cellularity" "nuclear circularity"	0.6419	0.5048 to 0.7791	0.0658	0.6931	0.5653 to 0.8209	0.0062[a]
1		0.7117	0.6131 to 0.8103	0.0002[a]	0.7778	0.5883 to 0.9672	0.0253[a]
	"Epithelial cellularity" "nuclear circularity" "nuclear eccentricity"	0.6617	0.5197 to 0.8036	0.0361[a]	0.6939	0.5655 to 0.8223	0.0060[a]
2		0.7367	0.6446 to 0.8289	<0.0001[a]	0.836	0.6422 to 1.000	0.0068[a]
	"Epithelial cellularity" "nuclear circularity" "nuclear eccentricity" "nuclear haematoxylin OD"	0.6607	0.5178 to 0.8037	0.0372[a]	0.6923	0.5634 to 0.8212	0.0064[a]
3		0.7473	0.6579 to 0.8367	<0.0001[a]	0.9048	0.7912 to 1.000	0.0011[a]
	"Epithelial cellularity" "nuclear circularity" "nuclear eccentricity" "nuclear haematoxylin OD mean" "cytoplasm eosin OD mean"	0.6983	0.5486 to 0.8481	0.0102[a]	0.7143	0.5849 to 0.8437	0.0024[a]
4		0.7467	0.6574 to 0.8360	<0.0001[a]	0.8995	0.7833 to 1.000	0.0013[a]
	"Epithelial cellularity" "nuclear circularity" "nuclear eccentricity" "nuclear haematoxylin OD" "cytoplasm eosin OD mean" "nuclear/cell area ratio"	0.7491	0.6134 to 0.8847	0.0012[a]	0.7292	0.6037 to 0.8547	0.0012[a]
5		0.7692	0.6832 to 0.8553	<0.0001[a]	0.9153	0.8045 to 1.000	0.0008[a]
	"Epithelial cellularity" "nuclear circularity" "nuclear eccentricity" "nuclear haematoxylin OD" "cytoplasm eosin OD mean" "nuclear/cell area ratio" "Perimeter of epithelium"	0.7744	0.6463 to 0.9026	0.0004[a]	0.741	0.6164 to 0.8656	0.0006[a]
6		0.7692	0.6833 to 0.8552	<0.0001[a]	0.9312	0.8103 to 1.000	0.0005[a]
	"Epithelial cellularity" "nuclear circularity" "nuclear eccentricity" "nuclear haematoxylin OD" "cytoplasm eosin OD mean" "nuclear/cell area ratio" "Perimeter of epithelium" "WHO Grade"	0.8506	0.7428 to 0.9583	<0.0001[a]	0.8155	0.7184 to 0.9127	<0.0001[a]
7		0.788	0.7085 to 0.8762	<0.0001[a]	0.963	0.9011 to 1.000	0.0002[a]
	"Epithelial cellularity" "nuclear circularity" "nuclear eccentricity" "nuclear haematoxylin OD" "cytoplasm eosin OD mean" "nuclear/cell area ratio" "Perimeter of epithelium" "Binary Grade"	0.8647	0.7666 to 0.9628	<0.0001[a]	0.814	0.7174 to 0.9105	<0.0001[a]
8		0.7946	0.7035 to 0.8724	<0.0001[a]	0.9418	0.8465 to 1.000	0.0004[a]

Each model is developed using 75 OED slides (252 ROI; 84 mild, 83 moderate, 85 severe). The first two rows indicate the prognostic strength of individual grading systems without additional variables, for reference. The rows highlighted in grey indicate the external validation results for model prediction (MT, $n = 121$; R, $n = 34$).

[a]Denotes a statistically significant finding.

performance of this model was further enhanced by adding individual grading systems (Models 7 and 8, Table 3), which yielded even better predictive potential than each alone, highlighting the potentially valid contribution of grading to clinical outcome prediction. More extensive validation on larger datasets is needed to establish the clinical utility of these features and models.

With regards to epithelial cellularity, there was an increased cell number in OED epithelium (compared to control) with more pronounced cellularity in severe/high grade OED lesions (Fig. 2). In contrast, a reduced basal cellularity was seen in OED epithelium compared to control ($p = 0.02$) with similar differences between moderate OED vs. control ($p = 0.02$) and high grade OED vs. control ($p = 0.007$). This is contrary to what we had expected, since basal cell crowding is thought to be associated with dysplasia severity. This finding can be explained by the increased level of cellular disarrangement and pleomorphism seen in more dysplastic regions, which in turn, may have resulted in fewer cells being detected. Whilst there were no significant prognostic correlations for cellularity in our study, further investigation of its diagnostic importance is worth exploring, considering the quantitative differences observed against non-dysplastic lesions.

Epithelial thickness was quantitatively evaluated by measuring the perimeter (length/distance) of the lesion margin/periphery, as an indirect measure of rete process/ridge morphology, a common feature of OED. Findings demonstrated the perimeter of OED epithelium to be particularly associated with malignant transformation ($p = 0.03$, AUROC 0.61 with 95% CI 0.47–0.75) (Table 2) indicating that epithelial thickness (or indirectly, curvature) may be a potentially important predictor of OED progression.

Whilst some studies have explored the effect of histomorphological characteristics, such as lesion thickness and cellularity in the diagnosis of premalignant lesions, few have studied this specifically in OED. One study showed differences in maximum lesion thickness and cellularity in high grade cervical squamous intraepithelial lesions compared to p16-positive cervical tissue biopsies [27]. In another study, microscopic analysis of oesophageal squamous dysplasia showed increased cellularity, disordered cell arrangement and loss of polarity in the basal layer [28]. Further analysis of keratin thickness, pattern and morphology in OED would be interesting to explore, particularly as abrupt orthokeratosis and verrucous surface architectures are frequently seen in oral potentially malignant disorders.

Analysis of nuclear and cytoplasmic features highlighted certain features to be more pronounced in OED with some differences between grades (Fig. 2). Relevant features include cytoplasm eosin OD (full epithelium, $p = 0.025$–0.035; basal layer, $p = 0.037$–0.039), nucleus eccentricity (basal layer, $p = 0.0004$ for binary grades, $p = 0.016$ for WHO grades) and nucleus circularity (basal layer, $p = 0.0005$ for binary grades). Eosin is a common synthetic dye used in H&E tissue analysis. It is a negatively charged dye which stains basic components of a cell, mainly positively charged proteins (or acidophilic) structures such as amino groups in the cytoplasm a bright pink colour, which contrasts with blue haematoxylin staining [29]. There is a lack of published research to explain the clinical relevance of increased cytoplasm eosin levels, particularly concerning OED diagnosis. However, our findings may be explained by the altered nuclear morphometry in dysplastic cells, which in turn, could affect eosin amount and representation. For example, an increase in cell size (cellular pleomorphism) may relate to increased cytoplasm eosin content. Furthermore, the presence of dyskeratosis and premature/individual cell keratinisation may contribute, giving the cytoplasm a more eosinophilic cytoplasm. Other potentially relevant clinical features relate to nuclear eccentricity (the displacement of the

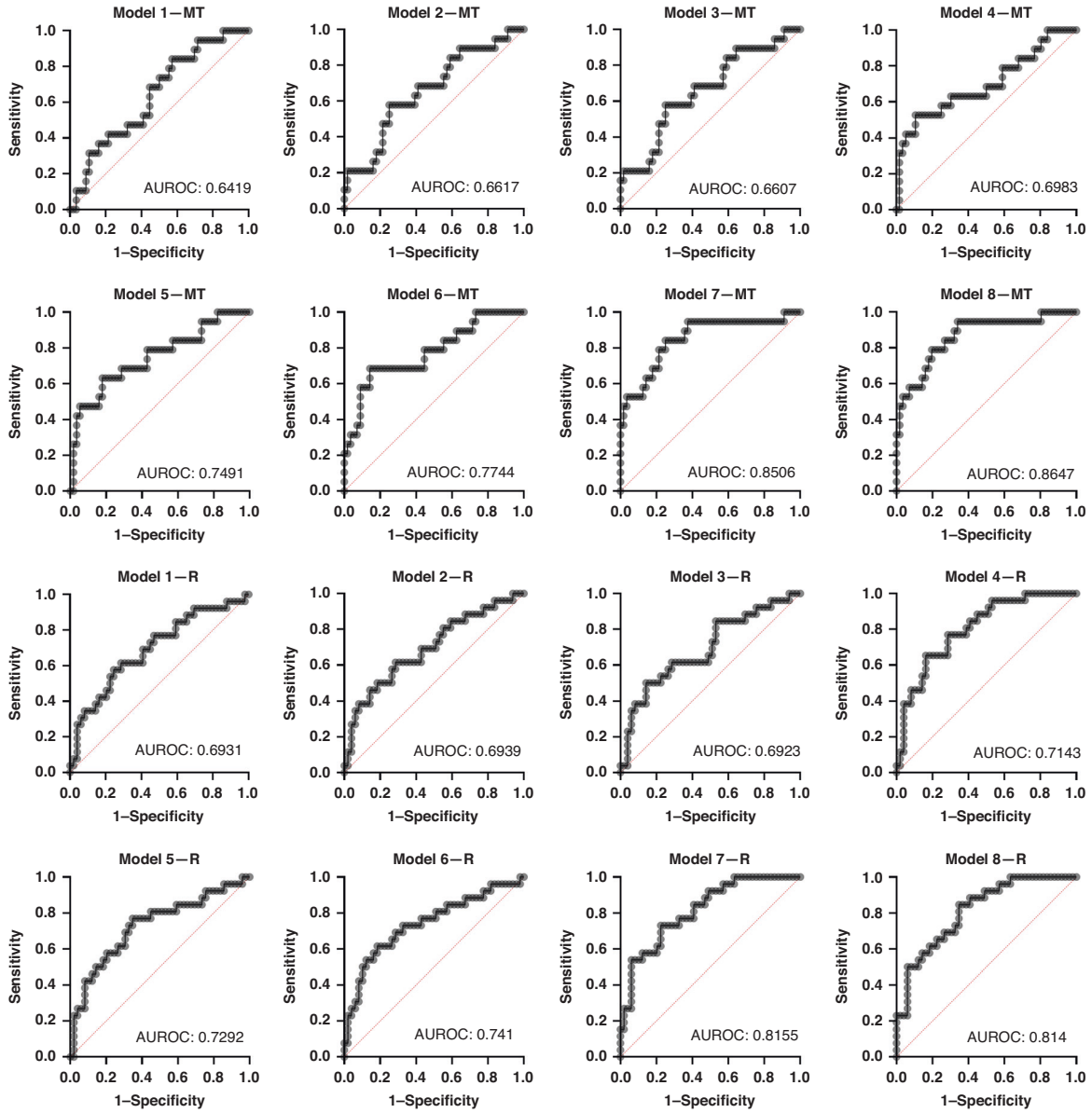

Fig. 3 ROC curves for multivariate models for malignant transformation (MT) and OED recurrence (R).

nucleus from the centre of the cell) and nuclear circularity (the degree to which the nucleus has deviated from circularity). Both these features are of diagnostic relevance in dysplastic lesions, in addition to other nuclear morphological features such as circularity, compactness, density and nuclear-cell ratio [30, 31]. In our study, nuclear circularity was also found to be associated with OED recurrence ($p = 0.02$, AUROC 0.693 with 95% CI 0.56–0.82) (Table 2).

The authors acknowledge some limitations of the current study. The first relates to the training sample used for quantitative analysis and model development, which includes cases from a single centre. However, the department in question is a UK national referral centre receiving OED cases from a wide geographical region, thereby incorporating varied samples from different patient groups. Whilst the sample size may be considered small, it must be highlighted that the digital quantitative evaluation has been conducted on 252 OED ROIs which is considered sufficient to draw initial conclusions. The second limitation relates to the selection of cases through purposive sampling, which may introduce an element of bias. However, we have tried to mitigate this risk, as our models have been tested on WSIs from four different national and international centres. These cases include a variable mix of dysplasia grades, a large proportion of transformed cases and have been scanned using different scanners. As such, the sample has increased biological and technical diversity which improves the robustness and generalisability of the developed models. Another limitation relates to potential staining variations between WSI cohorts. This has been somewhat overcome by using a training sample from a single centre, which follows consistent staining protocols. Whilst stain variation may introduce problems when machine learning algorithms are being developed, it is less problematic in purely quantitative digital image analysis. Slide stains are also not 'normalised' in routine diagnostic practice, so it is a more representative reflection of real-world approaches.

As we progress towards a digital pathology workflow, this study highlights the potential of digital image analysis tools for the study and prognostication of complex oral diseases. Our findings provide novel insights into OED progression, by

identifying new histological predictors that are not routinely considered in diagnostic practice. To the best of the authors' knowledge, this is the first study to quantify and correlate cellularity, geometric, colour and perimeter features in OED to clinical outcomes. Further analysis on larger multicentric cohorts is needed to validate and refine our findings to determine the full prognostic value of the studied features for wider clinical application.

DATA AVAILABILITY

All data analysed during this study are included in this published paper. Correspondence and requests for materials should be addressed to Dr HM.

REFERENCES

1. Smith J, Rattay T, McConkey C, Helliwell T, Mehanna H. Biomarkers in dysplasia of the oral cavity: a systematic review. Oral Oncol. 2009;45:647–53.
2. Pindborg JJ, Reichart PA, Smith CJ, Van der Waal I. Histological typing of cancer and precancer of the oral mucosa: In collaboration with LH Sobin and Pathologists in 9 Countries. Springer Sci Rev. (1997). In 'Precancerous Lesions (Histological Classification), page 24–6.
3. Tilakaratne WM, Jayasooriya PR, Jayasuriya NS, De Silva RK. Oral epithelial dysplasia: Causes, quantification, prognosis, and management challenges. Periodontol 2000 2019;80:126–47.
4. Gómez-Armayones S, Chimenos-Küstner E, Arranz C, Tous S, Marquez S, Penín RM, et al. Risk factors for oral epithelial dysplasias to become malignant: clinical implications. Int J. Oral Maxillofac Surg. 2022;51:473–80.
5. Califano J, Van Der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, et al. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer Res. 1996;56:2488–92.
6. Barnes L, Eveson JW, Sidransky D, Reichart P. Pathology and genetics of head and neck tumours. IARC (2005). In Chapter 4, 'Epithelial precursor lesions', page 177–9.
7. De la Cour CD, Sperling CD, Belmonte F, Syrjanen S, Verdoodt F, Kjaer SK. Prevalence of human papillomavirus in oral epithelial dysplasia: Systematic review and meta-analysis. Head Neck. 2020;42:2975–84.
8. Iocca O, Sollecito TP, Alawi F, Weinstein GS, Newman JG, De Virgilio A, et al. Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. Head Neck. 2020;42:539–55.
9. Hankinson PM, Mohammed-Ali RI, Smith AT, Khurram SA. Malignant transformation in a cohort of patients with oral epithelial dysplasia. Br J Oral Maxillofac Surg. 2021;59:1099–101.
10. Van der Waal I. Potentially malignant disorders of the oral and oropharyngeal mucosa; terminology, classification and present concepts of management. Oral Oncol. 2009;45:317–23.
11. Odell E, Kujan O, Warnakulasuriya S, Sloan P. Oral epithelial dysplasia: Recognition, grading and clinical significance. Oral Dis. 2021;27:1947–76.
12. El-Naggar AK, Chan JK, Grandis JR. WHO classification of head and neck tumours (2017). In Chapter 4, 'Oral potentially malignant disorders and oral epithelial dysplasia'. page 112–3.
13. Reibel J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. Crit Rev Oral Biol Med. 2003;14:47–62.
14. Muller S, Tilakaratne WM. Update from the 5th Edition of the World Health Organization Classification of Head and Neck Tumors: Tumours of the Oral Cavity and Mobile Tongue. Head Neck Pathol. 2022;16:54–62.
15. Kujan O, Khattab A, Oliver RJ, Roberts SA, 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. 2007;43:224–31.
16. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. Oral Oncol. 2006;42:987–93.
17. Nankivell P, Williams H, Matthews P, Suortamo S, Snead D, McConkey C, et al. The binary oral dysplasia grading system: validity testing and suggested improvement. Oral Surg Oral Med Oral Pathol Oral Radiol. 2013;115:87–94.
18. Mahmood H, Bradburn M, Rajpoot N, Islam NM, Kujan O, Khurram SA. Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models. Mod Pathol. 2022;35:1151–9.
19. Madabhushi A. Digital pathology image analysis: opportunities and challenges. Imaging Med. 2009;1:7.
20. Aeffner F, Zarella MD, Buchbinder N, Bui MM, Goodman MR, Hartman DJ, et al. Introduction to digital image analysis in whole-slide imaging: a white paper from the digital pathology association. J Pathol Inform. 2019;10:9.
21. Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD, et al. QuPath: Open-source software for digital pathology image analysis. Sci Rep. 2017;7:1–7.
22. Loughrey MB, Bankhead P, Coleman HG, Hagan RS, Craig S, McCorry AM, et al. Validation of the systematic scoring of immunohistochemically stained tumour tissue microarrays using QuPath digital image analysis. Histopathology 2018;73:327–38.
23. Koohbanani NA, Jahanifar M, Tajadin NZ, Rajpoot N. NuClick: a deep learning framework for interactive segmentation of microscopic images. Med Image Anal. 2020;65:101771.
24. Graham S, Vu QD, Raza SE, Azam A, Tsang YW, Kwak JT, et al. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal. 2019;58:101563.
25. Shephard AJ, Graham S, Bashir S, Jahanifar M, Mahmood H, Khurram A, et al. Simultaneous nuclear instance and layer segmentation in oral epithelial dysplasia. In Proceedings of the IEEE/CVF International Conference on Computer Vision 2021. p. 552–561.
26. Gamper J, Koohbanani NA, Benes K, Graham S, Jahanifar M, Khurram SA, et al. Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv*:2003.10778.
27. Gutnik H, Kastelic P, Valenčak AO, Poljak M, Fležar MS. Histomorphologic assessment and distribution of high-risk human papillomavirus (HPV) types in cervical high grade squamous intraepithelial lesions with unusual histomorphologic features. Virchows Arch. 2020;476:251–60.
28. Hou WH, Shen MQ, Hou WD, Zhang XL, Jin ML. Clinicopathological features of differentiated-type dysplasia of the esophagus. Zhonghua Bing li xue za zhi= Chin. J Pathol. 2021;50:1353–9.
29. Chan JK. The wonderful colors of the hematoxylin–eosin stain in diagnostic surgical pathology. Int J Surg Pathol. 2014;22:12–32.
30. Bag S, Conjeti S, Das RK, Pal M, Anura A, Paul RR et al. Computational analysis of p63+ nuclei distribution pattern by graph theoretic approach in an oral precancer (sub-mucous fibrosis). J Pathol Inform. 2013;1;4:35.
31. Gadiwan M, Madhushankari GS, Mandana DD, Praveen SB, Selvamani MS, Pradeep DS. Nuclear features in different grades of epithelial dysplasia in leukoplakia: A computer assisted microscopic study. J Oral Maxillofac Pathol. 2014;18:194.

ACKNOWLEDGEMENTS

The authors thank Dr Hannah Walsh (Consultant in Oral and Maxillofacial Pathology School of Clinical Dentistry, Sheffield, UK) for assistance with case retrieval and Dr Amir Zaki Abdullah Zubir (Postdoctoral Research Associate, School of Clinical Dentistry, Sheffield, UK) for assistance with slide scanning.

AUTHOR CONTRIBUTIONS

HM and SAK conceptualised the study. HM prepared slides, collected clinical data, conducted digital quantitative analysis and wrote the primary paper. PH, ALDA, JB, ARS, MAL, PAV, KDM, SGC, JJ, PN, Professor HM and SAK were involved in histological analysis and collection of follow-up data. HM, AS, MB, NR and SAK contributed to results analysis, and MB provided statistical expertise. All authors read and approved the final paper. NR and SAK provided supervisory guidance.

FUNDING

HM received funding from the National Institute for Health Research for this study, which forms part of her Doctoral Research Fellowship (NIHR300904). AS, NR and SAK acknowledge funding from Cancer Research UK (C63489/A29674). Prof HM reports grants from UK National Institute of Health research, Cancer Research UK, the UK Medical Research Council, and AstraZeneca; advisory board fees from AstraZeneca, MSD, Merck, Nanobiotix, and Seagen; and is Director of Warwickshire Head Neck clinic and Docpsert Health. He is also an NIHR Senior Investigator. The views expressed in this paper are those of the author(s) and not necessarily those of the NIHR, or the Department of Health and Social Care.

COMPETING INTERESTS

NR is the Director and CSO of Histofy, an AI start-up in the area of computational pathology. SAK is a shareholder of Histofy. The remaining authors do not declare any competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval was obtained from the West Midlands Research Ethics Committee (18/WM/0335).

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41416-023-02438-0.

Correspondence and requests for materials should be addressed to Hanya Mahmood.

Reprints and permission information is available at http://www.nature.com/reprints

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

Chapter 6 – Mitotic Feature Analysis

Chapter 6 – Mitotic Feature Analysis

This chapter has been reproduced from a peer-reviewed publication[11] for which the candidate shares joint first authorship with Hrishikesh Sathyamoorthy, a previous MSc student at University of Sheffield, for whom the candidate was a co-supervisor.

The study evaluates mitotic activity in OED assessed on H&E whole-slide images and immunohistochemistry for PHH3. Whilst the general study methods in this article are similar to that which Hrishikesh Sathyamoorthy submitted as part of his MSc project, the final analyses, statistical modelling and results are different.

The candidate's contributions to this study were:

i. co-design of study with final author
ii. obtaining ethical approval
iii. slide retrieval, dataset preparation and clinical data collection
iv. statistical analysis and modelling of quantitative data
v. multivariable model development and testing
vi. writing of the manuscript

Dr Amir Zaki Abdullah Zubir prepared the digital images.

Dr Paul Hankinson and Professor Syed Ali Khurram performed histological grading analysis.

Dr Hrishikesh Sathyamoorthy performed the immunohistochemistry and manual mitosis counting. Professor Syed Ali Khurram executed manual recounting.

6.1 Prognostic importance of mitosis quantification and PHH3 expression in oral epithelial dysplasia

Virchows Archiv
https://doi.org/10.1007/s00428-023-03668-6

ORIGINAL ARTICLE

Prognostic importance of mitosis quantification and PHH3 expression in oral epithelial dysplasia

Hrishikesh Sathyamoorthy[1] · Hanya Mahmood[2] · Amir Zaki Abdullah Zubir[1] · Paul Hankinson[1] · Syed Ali Khurram[1]

Received: 3 July 2023 / Revised: 28 September 2023 / Accepted: 29 September 2023

Abstract

Oral epithelial dysplasia (OED) is diagnosed and graded using a range of histological features, making grading subjective and challenging. Mitotic counting and phosphohistone-H3 (PHH3) staining have been used for the prognostication of various malignancies; however, their importance in OED remains unexplored. This study conducts a quantitative analysis of mitotic activity in OED using both haematoxylin and eosin (H&E)-stained slides and immunohistochemical (IHC) staining for PHH3. Specifically, the diagnostic and prognostic importance of mitotic number, mitotic type and intra-epithelial location is evaluated. Whole slide images (WSI) of OED ($n=60$) and non-dysplastic tissue ($n=8$) were prepared for analysis. Five-year follow-up data was collected. The total number of mitosis (TNOM), mitosis type and intra-epithelial location was manually evaluated on H&E images and a digital mitotic count performed on PHH3-stained WSI. Statistical associations between these features and OED grade, malignant transformation and OED recurrence were determined. Mitosis count increased with grade severity (H&E: $p<0.005$; IHC: $p<0.05$), and grade-based differences were seen for mitosis type and location ($p<0.05$). The ratio of normal-to-abnormal mitoses was higher in OED (1.61) than control (1.25) and reduced with grade severity. TNOM, type and location were better predictors when combined with histological grading, with the most prognostic models demonstrating an AUROC of 0.81 for transformation and 0.78 for recurrence, exceeding conventional grading. Mitosis quantification and PHH3 staining can be an adjunct to conventional H&E assessment and grading for the prediction of OED prognosis. Validation on larger multicentre cohorts is needed to establish these findings.

Keywords PHH3; Phosphohistone-H3 · Oral epithelial dysplasia · Mitosis · Malignant transformation · Recurrence · Oral squamous cell carcinoma

Hrishikesh Sathyamoorthy and Hanya Mahmood joint first authorship.

✉ Syed Ali Khurram
s.a.khurram@sheffield.ac.uk

Hrishikesh Sathyamoorthy
hrishikeshsukesh@gmail.com

Hanya Mahmood
h.mahmood@sheffield.ac.uk

Amir Zaki Abdullah Zubir
a.abdullahzubir@sheffield.ac.uk

Paul Hankinson
p.hankinson@sheffield.ac.uk

[1] Unit of Oral and Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield, 19 Claremont Crescent, Sheffield S10 2TA, UK

[2] Academic Unit of Oral & Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield, 19 Claremont Crescent, Sheffield S10 2TA, UK

Abbreviations

OED	Oral epithelial dysplasia
OSCC	Oral squamous cell carcinoma
HPV	Human papillomavirus
WHO	World Health Organization
H&E	Haematoxylin and eosin
PHH3	Phosphohistone-H3
IHC	Immunohistochemistry
WSI	Whole slide image
ROI	Region of interest
DPX	Dibutyl phthalate polystyrene xylene
ANOVA	Analysis of variance
AUROC	Area under receiver operator characteristic
TNOM	Total number of mitoses
Ki-67	Kiel-67

Published online: 26 October 2023

Introduction

Oral epithelial dysplasia (OED) describes a spectrum of histologically identified architectural and cytological disturbances involving the oral epithelium [1]. These lesions may progress to oral squamous cell carcinoma (OSCC) [2]. Higher grade lesions have higher risk of transformation, highlighting the need for an early and accurate diagnosis [1]. OSCC is the most common malignant neoplasm of the oral cavity associated with a myriad of environmental aetiologies and genetic alterations [3–5].

Because of the direct relationship between OED and malignant transformation, the dysplasia grade is considered the most important prognosticator for malignant transformation [5]. However, the current grading system (WHO, 2017) is associated with poor reproducibility, which can result in an inconsistent and unreliable diagnosis [6]. Suggestions to mitigate these shortcomings include the use of clinical determinants and molecular markers [7]. The binary grading system is an alternative criteria proposed to improve observer reproducibility by quantifying the minimum number of cytological and architectural features required for a diagnosis [8]. However, this classification uses the same histological features listed in the WHO Classification, and there remains a lack of high-quality evidence to support the prognostic importance of many of these features [2]. The recent update from the 5th Edition of the WHO Classification includes additional features, such as apoptotic mitoses and single cell keratinisation. However, the clinical relevance for inclusion of these features is unclear [9]. A recent study explored histological feature-specific associations in OED with clinical outcomes. The predictive performance of the proposed models for OED progression exceeded conventional grading [10]. However, a more detailed and prospective analysis of individual histological features is still needed to establish a more objective predictive/grading system.

Mitotic figure counting is used for diagnosis and prognostication of various malignancies [11–14] including breast, gastric and neuroendocrine carcinomas [13, 15–17]. However, its importance in precancer diagnosis and progression is yet to be explored. The main limitation of mitosis counting is the tediousness of the manual approach, in addition to interpretation differences due to variations in chromatin arrangements in the different mitotic stages, and the resemblance of apoptotic bodies and pyknotic nuclei with mitotic bodies (Fig. 1) [18]. Many of these limitations can now be overcome by the increasing number of digital/computational tools which allow for automated quantification, providing more objective, efficient and reliable outputs [19]. However, in the case of mitotic cell counting, attention also needs to be given to the presence of abnormal mitotic forms, characterised by mitotic asymmetry or an abnormal segregation of chromosomes [20].

Various biomarkers have been implicated in OED progression, but the evidence to support their routine use is still lacking [21]. Phosphohistone-H3 (PHH3) is a specific protein phosphorylated during chromatin condensation in mitosis [22]. It stains positively during the late G2 phase and M phase. Phosphorylation of the histone H3 starts to occur just before prophase which is not identifiable on haematoxylin and eosin (H&E) examination [18], lending to the role of PHH3 a useful marker.

The aims of this study were threefold: first, to conduct a quantitative analysis of mitotic activity in OED (including number, type and intra-epithelial location of mitoses) using digitised H&E sections and immunohistochemical (IHC)-stained tissue with PHH3; second, to evaluate changes in mitotic activity relative to OED progression; and third, to develop and explore multivariable models using mitotic features for prediction of OED recurrence and malignant transformation, with comparison to conventional grading.

Material and methods

Case selection and tissue processing

Following ethical approval (reference 18/WM/0335), a retrospective sample of 68 H&E-stained tissue sections were retrieved from the department archive. The sample comprised OED sections ($n=60$) of varying grades (mild, moderate, severe) with 5-year post-diagnosis data, in addition to non-dysplastic control samples ($n=8$) which included cases of benign hyperplasia, scar tissue and inflammatory oral lichen planus. Verrucous and HPV-related OED lesions were excluded based on morphological features, as they are distinct entities with reportedly different behaviours.

Prior to the inclusion, cases were independently reviewed by a consultant oral and maxillofacial pathologist (SAK) to ensure there was sufficient epithelial tissue for analysis. Cases with insufficient tissue, gross artefact or tangentially cut sections were excluded. All cases were then blindly re-evaluated by SAK, HM (clinician with extensive expertise and specialist interest in OED analysis) and PH (trainee oral and maxillofacial pathologist) to confirm the original diagnosis and where necessary assign an updated OED grade (using WHO and binary systems). Grading variability was measured by a Cohen's kappa score, which resulted in a value of 0.900, demonstrating good interobserver agreement.

New 5-µm-thick formalin-fixed paraffin-embedded sections of the selected cases were obtained for H&E and IHC staining. The sections were scanned at 40× magnification using an

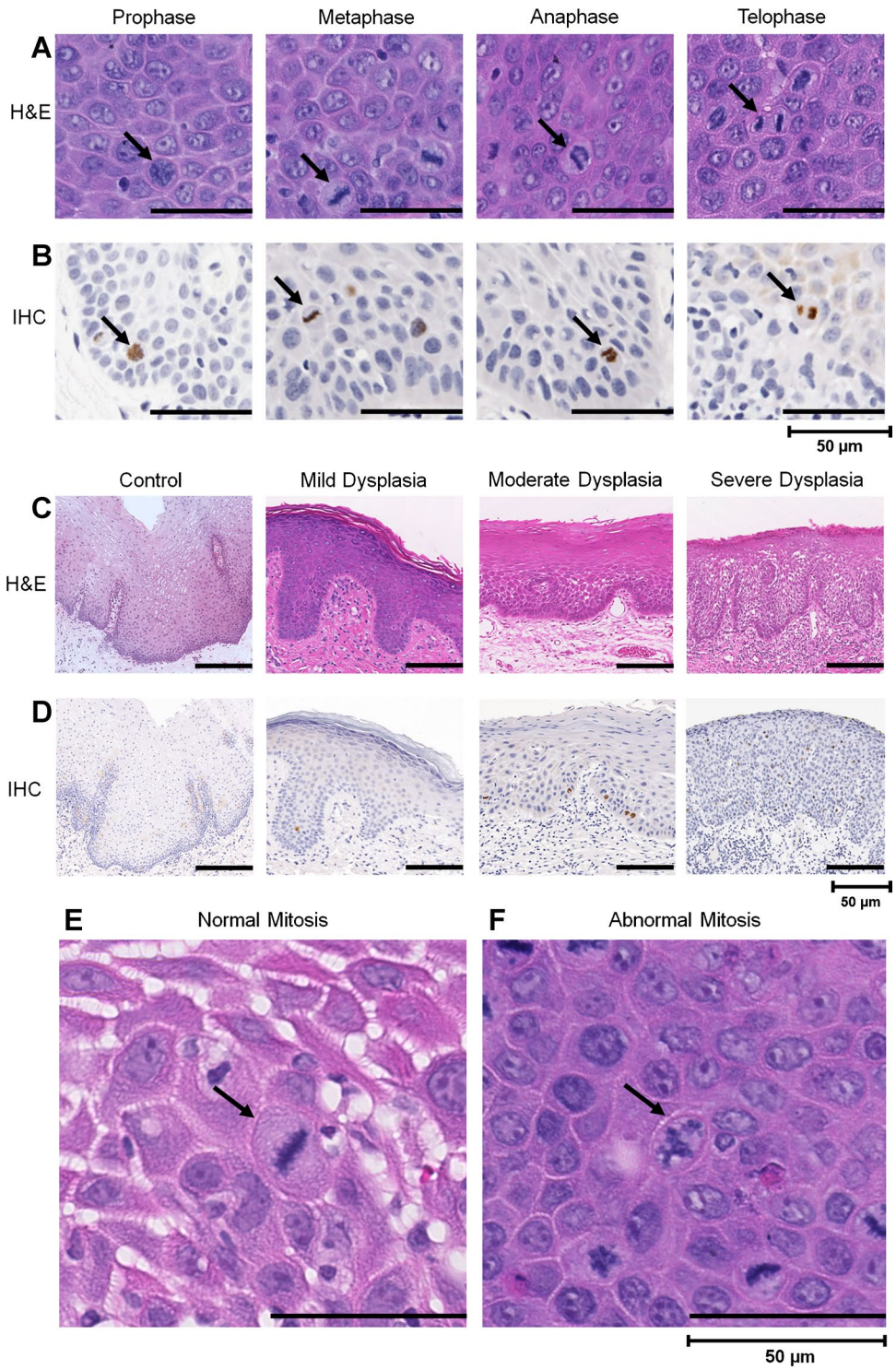

Fig. 1 Photomicrographs (40×) demonstrating the different mitotic stages observed in OED (black arrows) based on H&E (**A**) and PHH3-IHC staining (**B**). Photomicrographs (20×) demonstrating the different OED grades (WHO, 2017) on H&E (**C**) and PHH3-IHC-stained images (**D**). H&E photomicrographs (40×) demonstrating 'normal' appearance of mitosis (**E**) and 'abnormal' appearance of mitosis (**F**) highlighted by black arrow

Aperio-CS2 scanner (Leica Biosystems, Milton Keynes, UK) to obtain high-resolution whole slide images (WSI) producing 68 H&E slides and 67 IHC slides for analysis. The IHC sample had one less case due to technical scanning/imaging difficulties, resulting in its exclusion at the final stage.

Clinical data collection included patient age at diagnosis, sex, biopsy site, original histological grade (WHO, 2017), status of malignant transformation and recurrence (lesion that progressed to OSCC or recurred at the same clinical site following treatment within 5 years).

Immunohistochemical staining for PHH3

IHC staining was carried out for the mitosis marker PHH3 (Ser10) using a previously described protocol [23]. A primary rabbit anti-human PHH3 polyclonal antibody (#9701; Cell Signalling Technology, 1:100 dilution) and a secondary goat anti-rabbit antibody was used. Following IHC, counterstaining with haematoxylin and mounting in DPX was done for further analysis.

Analysis of mitosis activity in OED

QuPath software (v.0.3.2) was used for identification of regions of interest (ROI) and subsequent mitotic feature analysis [24]. For all slides, five rectangular-shaped ROIs of a consistent size (area≈165,000 mm^2) corresponding to representative dysplastic and non-dysplastic regions were selected at 20× magnification and verified by two experienced clinicians (HM, SAK).

For the H&E sample ($n=68$), two observers (HS, SAK), blinded to clinical outcomes, were asked to independently count and record (i) the total number of mitoses (TNOM), (ii) the number of 'normal' and 'abnormal' mitoses and (iii) the intra-epithelial mitosis location ('basal' or 'suprabasal') in each field. An agreement between the observers was made on how to qualify a 'normal' and 'abnormal' mitosis. An equational bipartition of the chromosomal material was used as standard for 'normal' mitosis [25], whereas the presence of abnormalities like binucleation, pyknotic nuclei, micronuclei and broken-egg appearances qualified the mitoses to be 'abnormal' [26]. A kappa score of 0.646 was obtained between the two observers for independent mitosis counting. In cases of wide disagreement, a consensus score was agreed/used for the downstream analyses. The means and standard deviation for the mitosis variables (TNOM, type and location) from the five ROIs were recorded and an average obtained for each case.

For the PHH3-IHC sample ($n=67$), QuPath's inbuilt 'positive cell detection' algorithm was applied for automated quantification of positively stained mitoses, and intra-epithelial mitosis location recorded through manual assessment (by HS, SAK). Due to the nature of the automated detection, the mitosis type could not be confirmed in the IHC sample. All data were exported onto a pre-structured spreadsheet in Microsoft Excel® (v.2206).

Statistical analyses

Statistical analyses were conducted in GraphPad Prism (v9) and IBM SPSS Statistics (v29.0.1.0). Data was tested for normality following which appropriate statistical tests were selected. Unpaired Student's *t*-tests and one-way ANOVA were performed to compare differences in the TNOM, mitosis type and intra-epithelial location between OED grades and relative to control. Where relevant, an appropriate post hoc analysis (Tukey's/Dunnett's) was performed for pairwise comparisons. For the H&E analysis, the mean mitosis number and ratio of normal-to-abnormal mitoses were measured and compared between grades. Paired sample *t*-tests were conducted to compare the number of normal and abnormal mitoses across OED grades.

Multivariable logistic regression models were explored separately for H&E and PHH3-IHC samples, to assess statistical relationships between individual and combined mitotic variables (TNOM, mitosis type, intra-epithelial location) with clinical outcomes (malignant transformation and OED recurrence). The effect of adding clinical variables (age, sex, intraoral site) and histological grade (WHO, binary) on model performance was assessed. The area under the receiver operator characteristic (ROC) curve was used to assess model accuracy and visualise performance. A p value of <0.05 was considered statistically significant. Figure 2 depicts the workflow methodology for this study.

Results

Characteristics of the OED cohort

Amongst the 60 OED cases, 39 (65%) were male, and 21 (35%) were female, with a mean age of 61.73 years (IQR 18.5). The clinical intraoral site distribution was the tongue $n=28$ (46.67%), floor of mouth $n=15$ (25%), buccal mucosa $n=8$ (13.33%), gingivae $n=5$ (8.33%) and palate $n=4$ (6.67%). The WHO histological grade distribution (following blind re-analysis) was mild OED = 20 (33.33%), moderate OED = 17 (28.33%) and severe OED = 23 (38.33%). Binary grade distribution was low-grade OED = 25 (41.7%) and high-grade OED = 35 (58.3%). A total of 14 cases (23.33%) transformed to OSCC, amongst which 8 (57.1%) were moderately dysplastic and 6 (42.9%) were severely dysplastic. Of the 19 (31.67%) cases that recurred after treatment, 8 were moderately dysplastic (42.1%), and 11 were severely dysplastic (57.9%).

Analysis of H&E and PHH3 mitotic count

Both the H&E and IHC analyses yielded a statistically significant difference in the TNOM between WHO grades (H&E: $p=0.0005$; IHC: $p=0.0073$) and binary OED grades (H&E: $p=0.0012$; IHC: $p=0.0403$) (Fig. 3). A significant difference was also seen when comparing TNOM between the following groups: mild OED vs severe OED (H&E: $p=0.0006$; IHC: $p=0.0197$), moderate OED vs severe OED (H&E: $p=0.0113$; IHC: $p=0.0181$), severe OED vs control (H&E: $p=0.0004$; IHC: $p=0.0009$) and high-grade OED

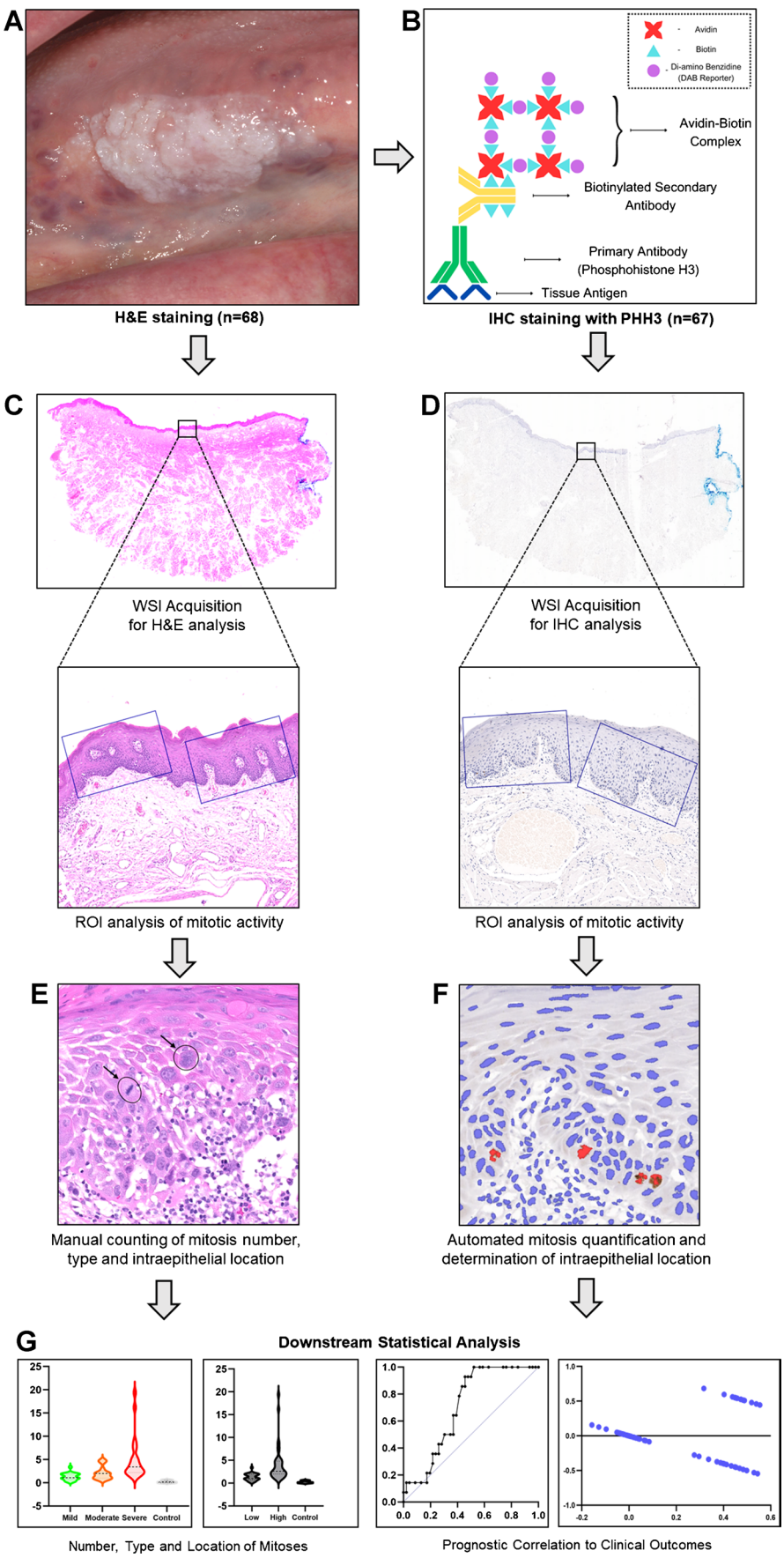

Fig. 2 Overall workflow methodology of the study. **A** Identification, retrieval and preparation of H&E sample ($n=68$). **B** Preparation of PHH3-IHC sample ($n=67$). Conversion of tissue sections to digital WSI and identification of ROI for H&E (**C**) and PHH3-IHC analysis (**D**). **E** Manual assessment of mitosis activity (number, type, location) on H&E. **F** Automated mitosis quantification for PHH3-IHC sample. **G** Statistical analysis to assess mitotic activity in OED with correlation to clinical outcomes

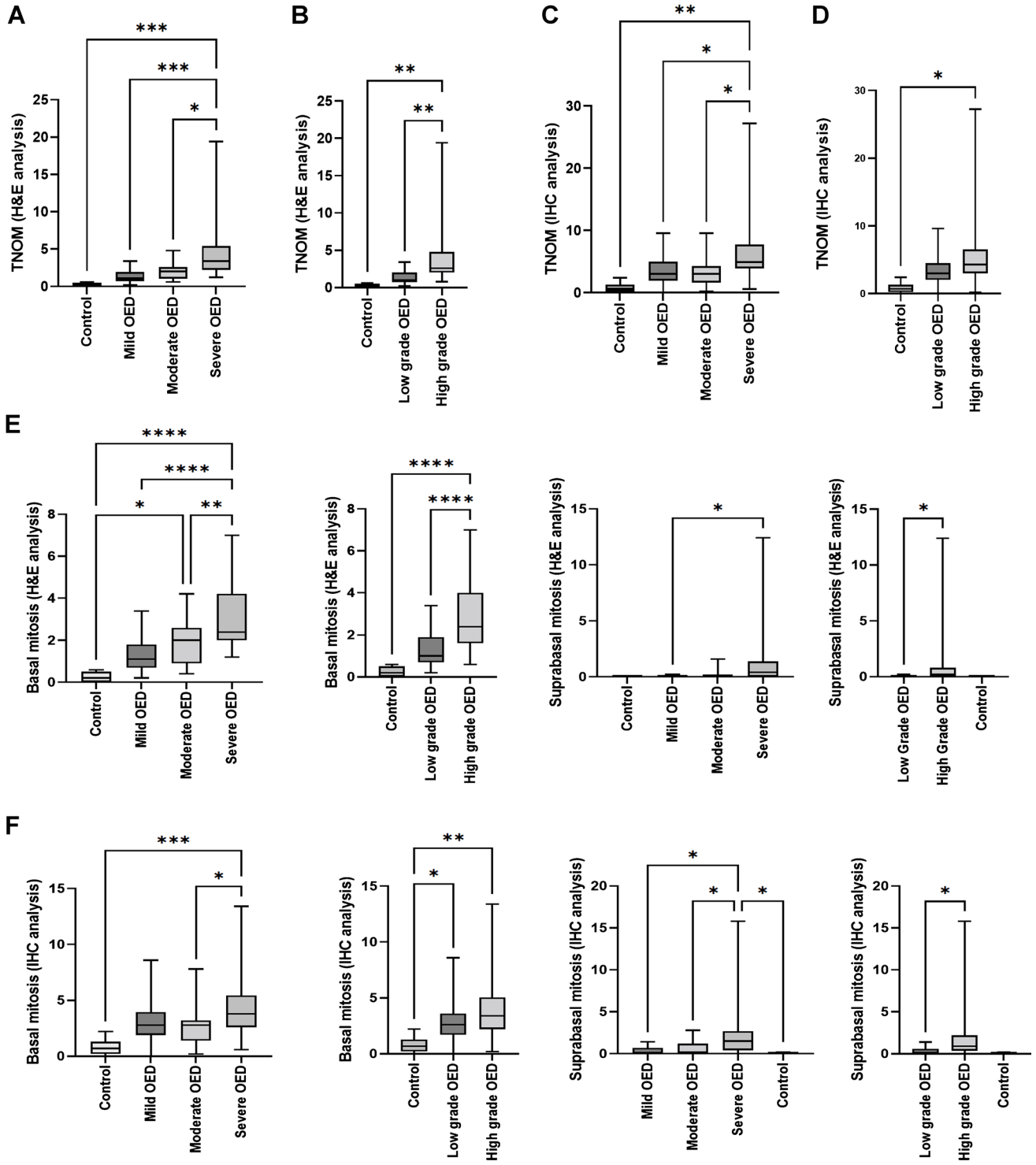

Fig. 3 Analysis of the TNOM based on H&E sections (**A**, **B**) and PHH3-IHC sections (**C**, **D**) with comparisons between histological grades and relative to control. Analysis of intra-epithelial mitosis location based on H&E sections (**E**) and PHH3-IHC sections (**F**) with comparisons between histological grade and relative to control. Asterisk indicates a statistically significant finding (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, **** $p \leq 0.0001$)

vs control (H&E: $p = 0.0022$; IHC: $p = 0.0064$) (Fig. 3).The remaining pairwise comparisons (mild OED vs moderate OED, mild OED vs control, moderate OED vs control and low-grade OED vs control) were not statistically significant. The mean mitosis number increased with grade severity (H&E: mild OED 1.32, moderate OED 2.09, severe OED

4.93, low-grade OED 1.32, high-grade OED 4.07) and relative to control (0.20). A similar trend was seen in IHC analysis (mild OED 3.47, moderate OED 3.26, severe OED 7.16, low-grade OED 3.36, high-grade OED 5.84, control 0.825).

H&E analysis of mitosis type

Normal mitotic figures

There was a significant difference in the average number of 'normal' mitoses between WHO grades ($p = 0.0016$) and binary grades ($p = 0.0040$) (Table 1). Significant differences were also seen between the following groups: control vs severe OED ($p = 0.0004$), control vs high-grade OED ($p = 0.0023$), mild OED vs severe OED ($p = 0.0026$) and moderate OED vs severe OED ($p = 0.0143$) (Table 1).

Abnormal mitotic figures

Similar trends were seen for the presence of 'abnormal' mitoses between WHO grades ($p = 0.0010$) and binary grades ($p = 0.0016$) (Table 1) in addition to comparisons between control vs severe OED ($p = 0.0032$), control vs high-grade OED ($p = 0.0116$), mild OED vs severe OED ($p = 0.0010$) and moderate OED vs severe OED ($p = 0.0322$) (Table 1).

Normal-to-abnormal mitosis ratio

The ratio of normal-to-abnormal mitoses was higher in OED (1.61) compared to control (1.25). This ratio was found to reduce with increasing grade severity. The ratios for mild, moderate and severe grades were 3.26, 1.49 and 1.43, and for low and high grades, 2.75 and 1.44, respectively. Statistically significant differences were observed when comparing the ratio of normal/abnormal mitoses across different grades ($p = 0.0001$ mild OED, $p = 0.0289$ moderate OED, $p = 0.0470$ severe OED, $p < 0.0001$ low-grade OED, $p = 0.0137$ high-grade OED).

Analysis of H&E and PHH3 mitosis location

Basal mitoses

A higher number of basal mitoses were observed with increasing grade severity, for WHO (mild OED = 1.3, moderate OED = 1.905882353, severe OED = 3.269565217) and binary grading (low-grade OED = 1.296, high-grade OED = 2.89) and relative to control (0.2) on H&E assessment ($p < 0.0001$). A similar trend was also seen on PHH3-IHC assessment between WHO grades and relative to control ($p = 0.0287$) (Fig. 3). Further comparisons demonstrated significance differences between mild OED vs. severe OED (H&E: $p < 0.0001$), moderate OED vs. severe OED (H&E: $p = 0.0076$; IHC: $p = 0.0383$), moderate OED vs. control (H&E: $p = 0.0163$), severe OED vs. control (H&E: $p < 0.0001$; IHC: $p = 0.0005$), low-grade OED vs. control (IHC: $p = 0.0495$) and high-grade OED vs. control (H&E: $p < 0.0001$; IHC: $p = 0.0024$) (Fig. 3). The remaining pairwise comparisons were not statistically significant.

Suprabasal mitoses

An increasing number of suprabasal mitoses were also observed with grade severity. Significant differences were shown between WHO grades (H&E: $p = 0.0174$; IHC: $p = 0.0076$) and binary grades (H&E: $p = 0.0364$; IHC: $p = 0.0202$) as well as between the following groups: mild OED vs. severe OED (H&E: $p = 0.0302$; IHC: $p = 0.0123$), moderate OED vs. severe OED (only IHC: $p = 0.0446$) and severe OED vs. control (only IHC: $p = 0.0435$) (Fig. 3). The remaining pairwise comparisons were not statistically significant.

Multivariable model development exploration

The association between mitosis variables, clinical characteristics, histological grades and clinical outcomes was assessed (for H&E and PHH3-IHC analysis) using multiple

Table 1 H&E analysis of mitosis type (measured by the presence and number of 'normal' and 'abnormal' mitoses) between individual grades of dysplasia (WHO and binary) and relative to control

Test parameters	Normal mitoses		Abnormal mitoses	
	p value	95% CI	*p* value	95% CI
Mild vs moderate OED	0.9051	− 1.655 to 1.158	0.5612	− 1.674 to 0.6705
Mild vs severe OED	0.0026*	− 3.198 to − 0.5906	0.0010*	− 2.811 to − 0.6383
Moderate vs severe OED	0.0143*	− 3.009 to − 0.2817	0.0322*	− 2.360 to − 0.08647
Control vs mild OED	0.3911	− 2.527 to 0.7569	0.9543	− 1.579 to 1.159
Control vs moderate OED	0.2359	− 2.817 to 0.5490	0.4349	− 2.114 to 0.6909
Control vs severe OED	0.0004*	− 4.390 to − 1.168	0.0032*	− 3.278 to − 0.5920
Control vs low-grade OED	0.3361	− 2.391 to 0.7048	0.8349	− 1.535 to 1.031
Control vs high-grade OED	0.0023*	− 3.774 to − 0.7875	0.0116*	− 2.806 to − 0.3312

Asterisk indicates statistical significance

logistic regression. For comparative purposes, the prognostic strength of conventional grading systems (WHO and binary) was also evaluated (Tables 2 and 3).

Prognostic potential of TNOM on H&E and PHH3-IHC sections

The TNOM alone had a modest association with malignant transformation (H&E: AUROC 0.5753; IHC: 0.5468) and OED recurrence (H&E: AUROC 0.6297; IHC: 0.5197), though the strength of association increased when combined with WHO grading (H&E: AUROC 0.7065 for transformation, 0.7401 for recurrence; IHC: AUROC 0.7460 for transformation, 0.7783 for recurrence) and binary grading (H&E: AUROC 0.722 for transformation, AUROC 0.6926 for recurrence; IHC: AUROC 0.7484 for transformation, AUROC 0.7184 for recurrence). The addition of clinical variables to TNOM had little or no effect on model performance (Table 2).

Prognostic potential of mitosis location on H&E and IHC-PHH3 sections

'Basal' mitosis was modestly associated with malignant transformation (H&E: AUROC 0.5815; IHC: AUROC 0.6381) and recurrence (H&E: AUROC 0.6175; IHC: 0.5411). In comparison, 'suprabasal' mitosis had a marginally weaker prognostic association (H&E: AUROC 0.5388 for transformation, AUROC 0.5854 for recurrence; IHC: 0.6278 for transformation, AUROC 0.6217 for recurrence). Whilst the addition of clinical variables had little overall effect on the prognostic strength of mitosis location, the incorporation of histological grading improved predictive strength, particularly for 'suprabasal' mitoses on H&E ('suprabasal mitoses' + 'WHO grade' = AUROC of 0.736 for transformation and 0.7458 for recurrence) (Table 2).

Prognostic potential of mitosis type on H&E sections

'Abnormal' mitoses alone had a greater predictive strength than 'normal' mitoses on H&E for transformation (AUROC 0.6856 vs 0.5016, respectively) and recurrence (AUROC 0.7022 vs 0.5552, respectively). However, incorporation of histological grading improved the predictive strength for 'normal' mitoses to a greater extent than for 'abnormal' mitoses ('normal mitoses' + 'WHO grade' = AUROC 0.7469, $p = 0.0055$ vs 'abnormal mitoses' + 'WHO grade' = AUROC 0.6537, $p = 0.0836$). The addition of clinical variables had little or no effect on model performance (Table 3).

Prognostic models using combined mitosis features

Combining the different mitosis variables with histological grading produced the most predictive models. The most superior model for prediction of transformation ('abnormal mitoses' + 'suprabasal mitoses' + 'TNOM' + 'WHO grade') produced an AUROC of 0.8113 ($p = 0.0005$, 95% CI 0.6987 to 0.9239), and the most superior model for prediction of recurrence ('abnormal mitosis' + 'basal mitoses' + 'TNOM' + 'WHO grade') achieved an AUROC of 0.7895 ($p = 0.0003$, 95% CI 0.6777 to 0.9013). Both these models outperformed conventional grading systems (Table 3).

Discussion

This study highlights the potential importance of mitosis assessment and quantification in OED diagnosis and prognostication. Mitosis counting has been effectively implemented in the diagnosis of various malignancies [13, 17, 27–29], but its diagnostic importance in oral precancers remains largely unexplored. Due to the limitations of manual mitotic figure counting, PHH3 was explored to evaluate its role as a diagnostic and prognostic adjunct to conventional H&E assessment.

The role of various oncogenes in OED progression to cancer still remains unvalidated [30]. Ki-67 being a cell cycle marker, rather than a specific marker of mitosis, has shown conflicting results. In one study, the value of PHH3 and Ki-67 for measuring mitotic activity in OSCC demonstrated a significant association between expression of PHH3 ($p = 0.016$) and mitotic activity ($p = 0.031$) with survival time; however, no similar relationship was found with Ki-67 ($p = 0.295$) [31]. In another study, the presence, location and pattern of Ki-67 positivity demonstrated variable results for differentiation between normal tissue, OED and OSCC [32]. The unreliability of Ki-67 [32, 33] and the successful use of PHH3 as an independent biomarker in various different malignancies [13, 15, 17, 22, 34] led us to explore this marker further.

The TNOM was shown to increase proportionally with grade severity on both H&E and PHH3-IHC analyses, supporting findings in the existing literature [35–38. This could be explained by the increased stem cell turnover and quantity of abnormal mutations [39]. Overall, PHH3 mitotic count was greater than H&E, likely due to the inclusion of early prophase stage, which cannot be reliably distinguished on H&E-stained sections. In a previous study, a comparison in mitotic count between H&E and crystal violet-stained sections demonstrated significant differences between non-dysplastic oral mucosa, OED and OSCC [39]. Whilst our findings revealed a greater

Table 2 Exploration of multivariate prognostic models based on the TNOM, mitosis location, clinical variables and histological grading systems (H&E $n=68$, PHH3-IHC $n=67-5$ ROI per WSI)

Model features	H&E models						PHH3-IHC models					
	Malignant transformation			OED recurrence			Malignant transformation			OED recurrence		
	AUROC	*p* value	95% CI	AUROC	*p* value	95% CI	AUROC	*p* value	95% CI	AUROC	*p* value	95% CI
WHO grading	0.6537	0.0836	0.5163 to 0.7911	0.7202	0.0064*	0.5950 to 0.8453	0.6635	0.0665	0.5266 to 0.8004	0.7316	0.0043*	0.6074 to 0.8557
Binary grading	0.6786	0.0444*	0.5289 to 0.8282	0.6893	0.0191*	0.5501 to 0.8286	0.6841	0.0388*	0.5347 to 0.8336	0.6961	0.0156*	0.5569 to 0.8352
TNOM	0.5753	0.3966	0.4197 to 0.7309	0.6297	0.1085	0.4798 to 0.7795	0.5468	0.5992	0.3837 to 0.7099	0.5197	0.8077	0.3606 to 0.6766
TNOM + WHO grading	0.7065	0.0201*	0.5766 to 0.8364	0.7401	0.0030*	0.6180 to 0.8621	0.746	0.0058*	0.6074 to 0.8847	**0.7783**	**0.0006***	**0.6635 to 0.8930**
TNOM + binary grading	0.722	0.0124*	0.5817 to 0.8624	0.6926	0.0171*	0.5585 to 0.8266	0.7484	0.0053*	0.5992 to 0.8976	0.7184	0.0071*	0.5821 to 0.8548
TNOM + age	0.5776	0.3822	0.4203 to 0.7350	0.6412	0.0805	0.4989 to 0.7835	0.6063	0.2326	0.4346 to 0.7781	0.6276	0.1156	0.4820 to 0.7733
TNOM + sex	0.5233	0.7932	0.3643 to 0.6823	0.586	0.287	0.4327 to 0.7393	0.596	0.2811	0.4286 to 0.7634	0.5395	0.6265	0.3819 to 0.6971
TNOM + age + sex	0.6685	0.0579	0.4908 to 0.8462	0.629	0.1103	0.4767 to 0.7814	0.6587	0.0748	0.4728 to 0.8446	0.6395	0.0855	0.4850 to 0.7940
TNOM + clinical site	0.5652	0.4629	0.4089 to 0.7215	0.5757	0.3485	0.4305 to 0.7210	0.5786	0.3778	0.4212 to 0.7360	0.5237	0.7703	0.3678 to 0.6796
Basal mitoses	0.5815	0.3588	0.4256 to 0.7375	0.6175	0.1459	0.4674 to 0.7676	0.6381	0.1211	0.4744 to 0.8018	0.5411	0.5868	0.3825 to 0.7057
Basal mitoses + WHO grading	0.6793	0.0435*	0.5421 to 0.8166	0.7246	0.0054*	0.5997 to 0.8496	0.7643	0.003*	0.6141 to 0.945	0.7737	0.0007*	0.6540 to 0.8934
Basal mitoses + binary grading	0.7057	0.0206*	0.5607 to 0.8508	0.697	0.0147*	0.5638 to 0.8303	**0.7714**	**0.0023***	**0.6260 to 0.9169**	0.7401	0.0031*	0.6111 to 0.8691
Basal mitoses + age	0.5885	0.3191	0.4286 to 0.7485	0.6483	0.0664	0.5058 to 0.7907	0.6627	0.0678	0.5008 to 0.8246	0.6362	0.0932	0.4915 to 0.7809
Basal mitoses + sex	0.5753	0.3966	0.4113 to 0.7393	0.6085	0.1793	0.4583 to 0.7587	0.6381	0.1211	0.4751 to 0.8011	0.5539	0.506	0.3928 to 0.7151
Basal mitoses + age + sex	0.6157	0.1929	0.4342 to 0.7972	0.6406	0.0818	0.4923 to 0.7888	0.681	0.0422*	0.5118 to 0.8501	0.6526	0.0599	0.5001 to 0.8051
Basal mitoses + clinical site	0.5854	0.3364	0.4297 to 0.7411	0.6085	0.1793	0.4632 to 0.7537	0.6254	0.1593	0.4610 to 0.7898	0.5401	0.6208	0.3804 to 0.6998
Suprabasal mitoses	0.5388	0.6622	0.3642 to 0.7135	0.5854	0.2906	0.4261 to 0.7446	0.6278	0.1515	0.4727 to 0.7828	0.6217	0.1335	0.4686 to 0.7748
Suprabasal mitoses + WHO grading	0.736	0.0079*	0.6130 to 0.8591	**0.7458**	**0.0023***	**0.6249 to 0.8667**	0.6794	0.0441*	0.5462 to 0.8125	0.7533	0.0018*	0.6331 to 0.8735
Suprabasal mitoses + binary grading	**0.7562**	**0.0039***	**0.6192 to 0.8932**	0.6945	0.0161*	0.5521 to 0.8369	0.6746	0.05	0.5124 to 0.8368	0.6816	0.0252*	0.5355 to 0.8276
Suprabasal mitoses + age	0.5877	0.3234	0.4285 to 0.7470	0.6168	0.1481	0.4734 to 0.7602	0.5532	0.5506	0.3862 to 0.7201	0.6625	0.0451*	0.5245 to 0.8005
Suprabasal mitoses + sex	0.5839	0.3453	0.4315 to 0.7362	0.5122	0.88	0.3611 to 0.6633	0.5103	0.9078	0.3427 to 0.6779	0.5414	0.6094	0.3872 to 0.6957
Suprabasal mitoses + age + sex	0.6584	0.0746	0.4846 to 0.8322	0.6175	0.1459	0.4639 to 0.7710	0.6389	0.119	0.4547 to 0.8231	0.6447	0.0744	0.4959 to 0.7936
Suprabasal mitoses + clinical site	0.5963	0.2785	0.4249 to 0.7676	0.5398	0.6223	0.3875 to 0.6920	0.5492	0.5807	0.3911 to 0.7073	0.5355	0.6614	0.3812 to 0.6898

The first two rows indicate the prognostic values for existing grading systems for comparative purposes. Highlighted rows indicate the most predictive models overall. Asterisk indicates a statistically significant finding. *AUROC* area under receiver operating characteristic

Text in bold indicate the most significant values/models

Table 3 Exploration of multivariate prognostic models based on the type of mitoses, clinical variables and histological grading systems on H&E assessment ($n = 68 - 5$ ROI per WSI)

Model features	H&E models					
	Malignant transformation			OED recurrence		
	AUROC	*p* value	95% CI	AUROC	*p* value	95% CI
WHO grading	*0.6537*	*0.0836*	*0.5163 to 0.7911*	*0.7202*	*0.0064**	*0.5950 to 0.8453*
Binary grading	*0.6786*	*0.0444**	*0.5289 to 0.8282*	*0.6893*	*0.0191**	*0.5501 to 0.8286*
Normal mitoses	0.5016	0.9861	0.3383 to 0.6648	0.5552	0.4944	0.3983 to 0.7121
Normal mitoses + WHO grading	0.7469	0.0055*	0.6229 to 0.8709	0.7548	0.0016*	0.6358 to 0.8738
Normal mitoses + binary grading	0.7663	0.0027*	0.6479 to 0.8847	0.7298	0.0044*	0.6053 to 0.8543
Normal mitoses + age	0.5901	0.3107	0.4386 to 0.7415	0.6354	0.0936	0.4897 to 0.7812
Normal mitoses + sex	0.58	0.3681	0.4306 to 0.7294	0.5019	0.981	0.3478 to 0.6561
Normal mitoses + age + sex	0.6693	0.0568	0.4979 to 0.8406	0.6316	0.1033	0.4804 to 0.7827
Normal mitoses + clinical site	0.5613	0.49	0.4082 to 0.7145	0.552	0.5198	0.4041 to 0.6999
Abnormal mitoses	0.6856	0.0367*	0.5441 to 0.8270	0.7022	0.0123*	0.5648 to 0.8396
Abnormal mitoses + WHO grading	0.6537	0.0836	0.5169 to 0.7906	0.7163	0.0074*	0.5901 to 0.8425
Abnormal mitoses + binary grading	0.6475	0.0968	0.4777 to 0.8173	0.6444	0.0738	0.4942 to 0.7947
Abnormal mitoses + age	0.6071	0.2278	0.4532 to 0.7611	0.6515	0.0608	0.5139 to 0.7890
Abnormal mitoses + sex	0.632	0.1374	0.4760 to 0.7880	0.6354	0.0936	0.4874 to 0.7834
Abnormal mitoses + age + sex	0.6281	0.1493	0.4502 to 0.8060	0.6393	0.0847	0.4912 to 0.7874
Abnormal mitoses + clinical site	0.6328	0.1351	0.4802 to 0.7853	0.6207	0.1352	0.4809 to 0.7605
Abnormal mitoses + suprabasal mitoses	0.7616	0.0032*	0.6162 to 0.9071	0.6823	0.024*	0.5311 to 0.8335
Abnormal mitoses + suprabasal mitoses + WHO	**0.7888**	**0.0012***	**0.6762 to 0.9014**	**0.7715**	**0.0008***	**0.6554 to 0.8876**
Abnormal mitoses + suprabasal mitoses + TNOM	0.7803	0.0016*	0.6331 to 0.9275	0.6643	0.0419*	0.5114 to 0.8172
Abnormal mitoses + suprabasal mitoses + TNOM + WHO	**0.8113**	**0.0005***	**0.6987 to 0.9239**	**0.7747**	**0.0007***	**0.6598 to 0.8896**
Abnormal mitoses + basal mitoses	0.6157	0.1929	0.4636 to 0.7677	0.5796	0.3245	0.4234 to 0.7358
Abnormal mitoses + basal mitoses + WHO	0.6778	0.0454*	0.5407 to 0.8149	0.7208	0.0063*	0.5950 to 0.8466
Abnormal mitosis + basal mitoses + TNOM	0.7143	0.0159*	0.5500 to 0.8786	0.6457	0.0713	0.4893 to 0.8021
Abnormal mitosis + basal mitoses + TNOM + WHO	0.764	0.003*	0.6406 to 0.8873	**0.7895**	**0.0003***	**0.6777 to 0.9013**

The first two rows indicate the prognostic values for existing grading systems for comparative purposes. Highlighted rows indicate the top most predictive models overall. Asterisk indicates a statistically significant finding. *AUROC* area under receiver operating characteristic

Text in bold indicate the most significant values/models

difference between mild and severe OED, control and high-grade/severe OED, promising differences were also observed between the more 'demanding' groups (moderate vs severe OED) in terms of mitosis number, mitosis type and mitosis location.

H&E analysis of mitosis type demonstrated a higher ratio of normal-to-abnormal mitoses in OED than control, which decreased with grade severity. Mitosis location assessment on H&E and IHC analysis demonstrated significant differences in the number of 'basal' and 'suprabasal' mitoses between grades. 'Suprabasal' mitoses were shown to be more predictive than 'basal' mitoses on PHH3-IHC. A study on meningioma demonstrated that PHH3 mitotic counts had a better interobserver correlation than H&E mitotic counts ($R_m = 0.83$ vs 0.77, respectively) [40], with good discrimination between grades (AUROC 0.91). Our study suggested similar findings, with better generally performance for PHH3-IHC models than H&E models, particularly for TNOM and mitosis location (Table 2). This is likely to be related to greater objectivity of mitosis assessment with PHH3 staining.

Prognostic models combining TNOM, mitosis type, location and histological grading showed better prediction for transformation and recurrence. Generally, the addition of clinical variables had minimal impact on model performances, whereas histological grading boosted predictive potential. Such a trend was also observed in a study by Mahmood et al. where inclusion of grades improved prognostic strength of histological OED models [10].

The most predictive H&E models for malignant transformation ('abnormal mitoses' + 'suprabasal mitoses' + 'TNOM' + 'WHO grade' = AUROC 0.8113) and OED recurrence ('abnormal mitosis' + 'basal mitoses' + 'TNOM' + 'WHO grade' = AUROC 0.7895) (AUROC 0.65) incorporated multiple mitotic features and outperformed conventional WHO grading on its own. In

the case of PHH3-IHC models, the most superior models utilised fewer mitotic features for prediction of transformation ('basal mitoses' + 'binary grading' = AUROC 0.7714) and recurrence ('TNOM' + 'WHO grading' = AUROC 0.7783). These findings indicate that PHH3-IHC may be important for prognostication of OED, complementing H&E analysis.

The authors acknowledge a few limitations. First, the follow-up period comprised 5 years. Whilst transformation may occur later [41], a number of studies have shown transformation incidence to be highest during the first 5 years. [5, 41–44] A study by Hankinson et al. (2021) reported a median transformation time of 22 months (IQR 46.0) for a cohort of OED cases ($n = 150$) retrieved from the same centre as that used for this study [45]. Second, cases were from a single-centre, and the sample size could be regarded as small [46, 47]. However, the unit in question is a national tertiary centre providing service to a large geographical region, thereby increasing the biological diversity of the sample. Furthermore, the sample has an equitable distribution of dysplasia grades with inclusion of transformed and non-transformed cases. For an early exploratory study that serves as a basis for future work, our sample is similar to many other studies [31, 48, 49] of this kind. The control cases were included for clinical interest and early comparative analysis, hence the small numbers. They did not contribute to the prognostic work, which was the important and novel aspect of this study.

In conclusion, we report increased mitotic activity with OED progression. Mitotic quantification using PHH3-IHC is potentially more reliable than H&E analysis, with typically greater predictive strength, even with inclusion of fewer variables. The addition of histological grading further improved performance of PHH3-IHC models, more so than the H&E models. To the best of our knowledge, this is one of the first studies to utilise mitosis quantification and compare H&E with PHH3-IHC for OED analysis and prognosis prediction. The promising results call for further exploration of H&E and IHC markers to contribute to a more objective grading of OED and reliable prognosis prediction. Further studies with larger multicentre cohorts are required for clinical validation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00428-023-03668-6.

Author contribution H.M. and S.A.K. designed the study. H.M. retrieved and prepared the case cohort with clinical data collection. H.M. and P.H. re-analysed the histological grading with S.A.K. H.S. carried out immunohistochemistry with guidance from A.Z.A.Z. A.Z.A.Z. prepared the digital images. H.S. carried out ROI analysis and manual and automated mitosis quantification, and S.A.K. executed the manual recounting. H.S. and H.M. performed statistical analysis. H.S., H.M. and S.A.K. were involved in the analysis and interpretation of results. H.S. and H.M. led on manuscript writing with contributions from A.Z.A.Z., P.H. and S.A.K. All authors approved the final version of the manuscript.

Funding HM is funded by the National Institute for Health Research (reference NIHR300904). PH is partially funded by a pre-doctoral bursary from Cancer Research UK and the Pathological Society for Great Britain and Ireland (reference RCCPSB-Nov21\100001). AZAZ is funded by a British Council Pakistan Grant (reference 20-ICRG-46/RGM/HEC/2020). SAK is partially funded by a Cancer Research UK Project Grant (reference C63489/A29674).

Data Availability All the data derived from this study are included in the manuscript. Further information for reasonable requests (if needed) can be made available, by contacting the last author (s.a.khurram@sheffield.ac.uk).

Declarations

Ethics approval Ethical approval was granted by the West Midlands-Edgbaston Research Ethics Committee (reference 18/WM/0335).

Conflict of interest The authors declare no competing interests.

References

1. Warnakulasuriya S, Reibel J, Bouquot J, Dabelsteen E (2008) Oral epithelial dysplasia classification systems: predictive value, utility, weaknesses and scope for improvement. J Oral Pathol Med 37(3):127–133. https://doi.org/10.1111/j.1600-0714.2007.00584.x
2. Odell E, Kujan O, Warnakulasuriya S, Sloan P (2021) Oral epithelial dysplasia: recognition, grading and clinical significance. Oral Dis 27(8):1947–1976. https://doi.org/10.1111/odi.13993
3. Pilborough AE, Lambert DW, Khurram SA (2019) Extranodal extension in oral cancer: a role for the nodal microenvironment? J Oral Pathol Med 48(10):863–870. https://doi.org/10.1111/jop.12870
4. Pires FR, Ramos AB, Coutinho de Oliveira JB, Tavares AS, Ribeiro da Luz PS, Bartholomeu R, dos Santos TC (2013) Oral squamous cell carcinoma: clinicopathological features from 346 cases from a single oral pathology service during an 8-year period. J Appl Oral Sci 21(5):460–467. https://doi.org/10.1590/1679-775720130317
5. Speight PM, Khurram SA, Kujan O (2018) Oral potentially malignant disorders: risk of progression to malignancy. Oral Surg Oral Med Oral Pathol Oral Radiol 125(6):612–627. https://doi.org/10.1016/j.oooo.2017.12.011
6. Holmstrup P, Vedtofte P, Reibel J, Stoltze K (2007) Oral premalignant lesions: is a biopsy reliable? J Oral Pathol Med 36(5):262–266. https://doi.org/10.1111/j.1600-0714.2007.00513.x

7. Ranganathan K, Kavitha L (2019) Oral epithelial dysplasia: classifications and clinical relevance in risk assessment of oral potentially malignant disorders. J Oral Maxillofac Pathol 23(1):19–27. https://doi.org/10.4103/jomfp.JOMFP_13_19
8. Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P (2006) Evaluation of a new binary system of grading oral epithelia dysplasia for prediction of malignant transformation. Oral Oncol 42(10):987–993. https://doi.org/10.1016/j.oraloncology.2005.12.014
9. Muller S, Tilakaratne WM (2022) Update from the 5th edition of the World Health Organization classification of head and neck tumors: tumours of the oral cavity and mobile tongue. Head Neck Pathol 16(1):54–62
10. Mahmood H, Bradburn M, Rajpoot N, Islam NM, Kujan O, Khurram SA (2022) Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models. Mod Pathol 35(9):1151–1159. https://doi.org/10.1038/s41379-022-01067-x
11. Tamori Y, Suzuki E, Deng W-M (2016) Epithelial tumors originate in tumor hotspots, a tissue-intrinsic microenvironment. PLoS Biol 14(9):e1002537. https://doi.org/10.1371/journal.pbio.1002537
12. Dessauvagie BF, Thomas C, Robinson C, Frost FA, Harvey J, Sterrett GF (2015) Validation of mitosis counting by automated phosphohistone H3 (PHH3) digital image analysis in a breast carcinoma tissue microarray. Pathology 47(4):329–334. https://doi.org/10.1097/pat.0000000000000248
13. Skaland I, Janssen EAM, Gudlaugsson E et al (2007) Phosphohistone H3 expression has much stronger prognostic value than classical prognosticators in invasive lymph node-negative breast cancer patients less than 55 years of age. Mod Pathol 20(12):1307–1315. https://doi.org/10.1038/modpathol.3800972
14. Stalhammar G, Robertson S, Wedlund L et al (2018) Digital image analysis of Ki67 in hot spots is superior to both manual Ki67 and mitotic counts in breast cancer. Histopathology 72(6):974–989. https://doi.org/10.1111/his.13452
15. Takahashi H, Murai Y, Tsuneyama K et al (2006) Overexpression of phosphorylated histone H3 is an indicator of poor prognosis in gastric adenocarcinoma patients. Appl Immunohistochem Mol Morphol 14(3):296–302. https://doi.org/10.1097/00129039-200609000-00007
16. Colman H, Giannini C, Huang L et al (2006) Assessment and prognostic significance of mitotic index using the mitosis marker phospho-histone H3 in low and intermediate-grade infiltrating astrocytomas. Am J Surg Pathol 30(5):657–664. https://doi.org/10.1097/01.pas.0000202048.28203.25
17. Tsuta K, Liu DC, Kalhor N, Wistuba II, Moran CA (2011) Using the mitosis-specific marker anti-phosphohistone H3 to assess mitosis in pulmonary neuroendocrine carcinomas. Am J Clin Pathol 136(2):252–259. https://doi.org/10.1309/ajcpdxfopxgef0rp
18. Veras E, Malpica A, Deavers MT, Silva EG (2009) Mitosis-specific marker phospho-histone H3 in the assessment of mitotic index in uterine smooth muscle tumors: a pilot study. Int J Gynecol Pathol 28(4):316–321. https://doi.org/10.1097/PGP.0b013e318193df97
19. Veta M, van Diest PJ, Jiwa M, Al-Janabi S, Pluim JPW (2016) Mitosis counting in breast cancer: object-level interobserver agreement and comparison to an automatic method. PLoS One 11(8):e0161286. https://doi.org/10.1371/journal.pone.0161286
20. Donovan TA, Moore FM, Bertram CA et al (2021) Mitotic figures-normal, atypical, and imposters: a guide to identification. Vet Pathol 58(2):243–257. https://doi.org/10.1177/0300985820980049
21. Smith J, Rattay T, McConkey C, Helliwell T, Mehanna H (2009) Biomarkers in dysplasia of the oral cavity: a systematic review. Oral Oncol 45(8):647–653. https://doi.org/10.1016/j.oraloncology.2009.02.006
22. Sawicka A, Seiser C (2012) Histone H3 phosphorylation - a versatile chromatin modification for different occasions. Biochimie 94(11):2193–2201. https://doi.org/10.1016/j.biochi.2012.04.018
23. Abdullah Zubir AZ, Whawell SA, Wong TS, Khurram SA (2020) The chemokine lymphotactin and its recombinant variants in oral cancer cell regulation. Oral Dis 26(8):1668–1676. https://doi.org/10.1111/odi.13500
24. Bankhead P, Loughrey MB, Fernandez JA et al (2017) QuPath: open source software for digital pathology image analysis. Sci Rep 7(1):16878. https://doi.org/10.1038/s41598-017-17204-5
25. Steinbeck R (2001) Pathologic mitoses and pathology of mitosis in tumorigenesis. Eur J Histochem 45(4):311–318
26. Ankle MR, Kale AD, Charantimath S (2007) Comparison of staining of mitotic figures by haematoxylin and eosin-and crystal violet stains, in oral epithelial dysplasia and squamous cell carcinoma. Indian J Dent Res 18(3):101–105
27. Pidhorecky I, Cheney RT, Kraybill WG, Gibbs JF (2000) Gastrointestinal stromal tumors: current diagnosis, biologic behavior, and management. Ann Surg Oncol 7:705–712
28. Diaconescu D, Diaconescu S, Chesca A, Toma S (2011) The value of mitotic counting in prostate carcinomas. Int J Math Models Methods Appl Sci 2:379–386
29. Goodarzi M, Correa AM, Ajani JA et al (2009) Anti-phosphorylated histone H3 expression in Barrett's esophagus, low-grade dysplasia, high-grade dysplasia, and adenocarcinoma. Mod Pathol 22(12):1612–1621
30. Pitiyage G, Tilakaratne WM, Tavassoli M, Warnakulasuriya S (2009) Molecular markers in oral epithelial dysplasia: review. J Oral Pathol Med 38(10):737–752. https://doi.org/10.1111/j.1600-0714.2009.00804.x
31. Tancredi-Cueto N, Vigil-Bastitta G, Bologna-Molina R, Beovide-Cortegoso V (2022) The value of phosphohistone H3 as a cell proliferation marker in oral squamous cell carcinoma. A comparative study with Ki-67 and the mitotic activity index. Med Oral Patol Oral Cir Bucal. 27(5):e444
32. Birajdar SS, Radhika M, Paremala K, Sudhakara M, Soumya M, Gadivan M (2014) Expression of Ki-67 in normal oral epithelium, leukoplakic oral epithelium and oral squamous cell carcinoma. J Oral Maxillofac Pathol 18(2):169–176. https://doi.org/10.4103/0973-029x.140729
33. Scholzen T, Gerdes J (2000) The Ki-67 protein: from the known and the unknown. J Cell Physiol 182(3):311–322. https://doi.org/10.1002/(sici)1097-4652(200003)182:3%3c311::aid-jcp1%3e3.0.co;2-9
34. Zhu P, Zhang C-B, Yang P et al (2016) Phosphohistone H3 (pHH3) is a prognostic and epithelial to mesenchymal transition marker in diffuse gliomas. Oncotarget 7(29):45005–45014. https://doi.org/10.18632/oncotarget.7154
35. Steinbeck RG (2004) Dysplasia in view of the cell cycle. Eur J Histochem 48(3):203–211
36. Kovesi G, Szende B (2003) Changes in apoptosis and mitotic index, p53 and Ki67 expression in various types of oral leukoplakia. Oncology 65(4):331–336. https://doi.org/10.1159/000074646
37. Pilati SFM, Bianco BC, Vieira DSC, Modolo F (2017) Histopathologic features in actinic cheilitis by the comparison of grading dysplasia systems. Oral Dis 23(2):219–224. https://doi.org/10.1111/odi.12597
38. Camara PR, Dutra SN, Takahama A, Fontes K, Azevedo RS (2016) A comparative study using WHO and binary oral epithelial dysplasia grading systems in actinic cheilitis. Oral Dis 22(6):523–529. https://doi.org/10.1111/odi.12484
39. Tandon A, Singh NN, Brave VR, Sreedhar G (2016) Image analysis assisted study of mitotic figures in oral epithelial dysplasia and squamous cell carcinoma using differential stains. J Oral Biol Craniofac Res 6(Suppl 1):S18–S23. https://doi.org/10.1016/j.jobcr.2016.09.003

40. Duregon E, Cassenti A, Pittaro A et al (2015) Better see to better agree: phosphohistone H3 increases interobserver agreement in mitotic count for meningioma grading and imposes new specific thresholds. Neuro Oncol 17(5):663–669. https://doi.org/10.1093/neuonc/nov002
41. Nevanpaa TT, Terava AE, Laine HK, Rautava J (2022) Malignant transformation of oral epithelial dysplasia in Southwest Finland. Sci Rep 12(1):8261. https://doi.org/10.1038/s41598-022-12441-9
42. Ho MW, Risk JM, Woolgar JA et al (2012) The clinical determinants of malignant transformation in oral epithelial dysplasia. Oral Oncol 48(10):969–976. https://doi.org/10.1016/j.oraloncology.2012.04.002
43. Lumerman H, Freedman P, Kerpel S (1995) Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. Oral Surg Oral Med Oral Pathol Oral Radiol Endodontol 79(3):321–329
44. Liu W, Bao Z-X, Shi L-J, Tang G-Y, Zhou Z-T (2011) Malignant transformation of oral epithelial dysplasia: clinicopathological risk factors and outcome analysis in a retrospective cohort of 138 cases. Histopathology 59(4):733–740. https://doi.org/10.1111/j.1365-2559.2011.03938.x
45. Hankinson PM, Mohammed-Ali RI, Smith AT, Khurram SA (2021) Malignant transformation in a cohort of patients with oral epithelial dysplasia. Br J Oral Maxillofac Surg 59(9):1099–1101. https://doi.org/10.1016/j.bjoms.2021.02.019
46. Jaber MA, Elameen EM (2021) Long-term follow-up of oral epithelial dysplasia: a hospital based cross-sectional study. J Dent Sci 16(1):304–310. https://doi.org/10.1016/j.jds.2020.04.003
47. Liu W, Bao ZX, Shi LJ, Tang GY, Zhou ZT (2011) Malignant transformation of oral epithelial dysplasia: clinicopathological risk factors and outcome analysis in a retrospective cohort of 138 cases. Histopathology 59(4):733–740
48. Ivina AA, Semkin VA, Khabadze ZS, Babichenko II (2019) Immunohistochemical study of Ki-67, PHH3, and CK15 protein expression in oral epithelial malignancy. Arkhiv Patologii 81(5):30–34. https://doi.org/10.17116/patol20198105130
49. Dragomir L, Simionescu C, Mărgăritescu C, Stepan A, Dragomir IM, Popescu M (2012) P53, p16 and Ki67 immunoexpression in oral squamous carcinomas. Rom J Morphol Embryol 53(1):89–93

Chapter 7 – Machine Learning Models

Chapter 7 – Machine Learning Models

This chapter is formed by a publication-ready manuscript for which the candidate is first author. This study explores supervised ML techniques for classification of dysplastic, immune, and stromal cells in OED. The study also aims to explore spatial arrangements of cells in OED and conducts a digital quantitative analysis of stromal cellularity in OED and evaluates the prognostic importance of nuclei morphometry in these cells.

The candidate's contributions to these studies were:

i. obtaining ethical approval
ii. slide retrieval and digital dataset preparation
iii. clinical data collection
iv. training and testing of the ML models
v. digital quantification and data extraction
vi. statistical analyses
vii. writing up the manuscript

The study was conceptualised with guidance from supervisors.

7.1 Application of machine learning and digital image analysis for assessment and quantification of dysplastic, immune and stromal cells in oral epithelial dysplasia

Authors: **H. Mahmood[1],** A. Shephard[2], N. Rajpoot[2], S. A. Khurram[3]

[1] Academic Unit of Oral & Maxillofacial Surgery, School of Clinical Dentistry, University of Sheffield

[2] Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, UK

[3] Unit of Oral & Maxillofacial Pathology, School of Clinical Dentistry, University of Sheffield

Abstract

Background: Histopathological analysis of oral epithelial dysplasia (OED) to predict malignant transformation remains problematic due to poorly constructed grading systems with unreliable predictive value. Modern digital image analysis tools and machine learning (ML) algorithms can be utilised for automated detection and cell quantification, in addition to providing valuable spatial information about cell arrangement and distribution. This study explores ML models for automated classification of dysplastic, immune, and stromal cells in OED and conducts a spatial analysis and digital quantification of stromal cellularity and nuclei morphometry.

Methods: ML models were trained to classify dysplastic (n=185) and immune cells in OED (n=215). The training datasets comprised whole-slide images of OED (n=135), non-dysplastic oral tissue (n=50) in addition to a lymph node section (n=1) and oropharyngeal carcinoma (n=29) cases acquired from The Cancer Genome Atlas for immune cell training. Algorithm performance was tested on unseen images (n=25-32) and the top-performing classifier used to extract spatial data for dysplastic cells. A digital quantitative analysis of stromal cellularity and nuclei morphometry was also conducted (n=100) and features correlated to clinical outcomes.

Results: Artificial Neural Network-Multilayer Perceptron model achieved high predictive accuracies for detection of dysplastic, immune and stromal cells (F1 scores 0.78-0.94). The distance and area between OED cells were lower than non-dysplastic cells ($p<0.03$) and reduced with grade severity. Stromal cellularity in OED was higher than control ($p<0.0001$) and increased with grade severity. Analysis of the nuclei in stromal cells of OED indicated that ‘nuclear circularity’, ‘nuclear eccentricity’, ‘nuclear to cell area ratio’ and ‘cytoplasm eosin OD’ were potentially associated with an increased risk of malignant transformation ($p<0.05$).

Conclusion: This study demonstrates the promising potential for traditional ML algorithms to reliably detect dysplastic, immune, and stromal cells in OED, and reveals novel findings pertaining to the spatial arrangements of cells and prognostic importance of stromal features in OED.

Keywords: Oral epithelial dysplasia; malignant transformation; histological grading; prognosis; machine learning; artificial intelligence.

7.1.1 Background

The study of oral epithelial dysplasia (OED) is gaining more attention as oral cancer incidence rates rise[53]. OED, often diagnosed in oral potentially malignant disorders (OPMD), is associated with an increased prevalence of oral squamous cell carcinoma (OSCC)[54] which is amongst the most common cancers worldwide, and a leading cause of death[53,23]. Histopathological analysis of haematoxylin and eosin (H&E) stained tissue is still the accepted practice for diagnosis of OED, and histological grading using the WHO (2017) system remains the main measure of malignant transformation prediction[53]. However, this system lacks objectivity, reproducibility, and evidence to support its prognostic value[57]. Recent iterations of the WHO system further complicates OED grading by expansion of the histological criteria[54], resulting in the potential for greater irreproducibility and observer disconcordance[28,42,54]. Notably, OED cases are referred to specialist centres for second opinions due to the risk of misdiagnosis[123,124].

With recent advancements in artificial Intelligence (AI) and digitisation of histology imaging data into whole slide images (WSI), there are increasing efforts to generate models towards computer-aided diagnosis. These methods have the potential to increase accuracy and efficiency in diagnosis in addition to providing quantifiable outputs which can increase our understanding of tissue architecture in complex histopathological conditions and aid discovery of novel cancer predictors[8].

Machine learning (ML) is a branch of AI, in which computational algorithms are trained to learn patterns directly from data. In the context of predictive modelling, ‘classification’ refers to a supervised learning approach that requires the use of ML algorithms to predict a class labels. With high throughput data, ML algorithms can adaptively improve predictive performance to overcome many of the limitations associated with manual approaches, such as reducing subjectivity, bias and misclassification. An increasing number of studies have applied ML methods for analysis of OPMD using H&E WSI[125,126,127,128,129]. These studies have mostly focussed on analysis of geometric and morphological characteristics of tissue components, such as cell or nuclei number, shape, staining intensity or surface texture. Beyond this, there has been little exploration of structural and tissue organisational characteristics, both in OPMD[125,130] and other types of precancers[131,132,133].

This study had three main aims. First, to explore supervised ML techniques for classification of dysplastic, immune, and stromal cells in OED; second, to explore spatial arrangements of cells in OED using geometric principles of Delaunay Triangulation and spatial clustering; third, to conduct a digital quantification of stromal cellularity in the subepithelial layer of OED and evaluate the prognostic importance of nuclei morphometry in these cells.

7.1.2 Methods

7.1.2.1 *Study Design & Clinical Samples*

Various traditional ML models were trained to detect and classify dysplastic, immune and stromal cells in OED. Specifically, two object classifiers were built:

1. *Classifier D-ND:* for binary classification of dysplastic *(D)* and non-dysplastic cells *(ND)* (n=185)
2. *Classifier S-I:* for binary classification of stromal cells *(S)* (including fibroblasts and endothelial cell types) and immune cells *(I)* in the subepithelial layer of OED (n=215)

The classifiers were trained on high resolution digital histopathology images retrieved from the Unit of Oral Pathology, School of Clinical Dentistry, Sheffield, UK during the period of 2008 to 2013. Approval of the institutional ethical committee was obtained (reference: 18/WM/0335) and all experiments were conducted in accordance with the Declaration of Helsinki.

Due to the different classification tasks, slightly different training datasets were prepared for the two classifiers. The training dataset for *Classifier D-ND* comprised 185 cases total including histologically confirmed cases of OED (n=135) with equal numbers of mild, moderate and severe WHO grades in addition to non-dysplastic oral tissue (n=50). The same dataset was also used for training of *Classifier S-I* with the addition of a head and neck lymph node section (n=1) and cases of oropharyngeal squamous cell carcinoma (OPSCC) (n=29) acquired from The Cancer Genome Atlas (TCGA) dataset generating a total of 215 cases.

For the spatial and quantitative stromal analysis, the same sample as that described in the study by Mahmood *et al.* (2023)[10] presented in Chapter 5 was used. This retrospective sample comprised 75 OED cases (one representative H&E slide per case) in addition to 25 non-dysplastic oral tissue sections (including hyperplasia and traumatic hyperkeratosis) for comparative analyses.

Prior to the inclusion of slides, each H&E tissue section was re-evaluated by a certified Oral and Maxillofacial Pathologist, blinded to clinical variables, to confirm the original diagnosis and ensure: 1) a histologically representative sample with sufficient tissue for reliable classification and analysis; 2) availability of minimum five-year clinical follow up data and 3) absence of tissue artefacts and staining defects. Cases with features suggestive of human papilloma virus (HPV) or lesions with a verrucous morphology were excluded from this study.

Demographic and clinic-pathological data was acquired for each case from patient records. Collected information included: i) sex; ii) age at time of histological diagnosis; iii) WHO 2017 OED grade (mild, moderate or severe); iv) binary OED grade (low or high grade); v) clinical outcome of lesion (malignant transformation and recurrence of OED).

7.1.2.2 *Preparation of Digital Datasets*

New 4µm H&E sections were obtained for each case and converted to high resolution whole slide images (WSI) at 20x (0.4952 microns-per pixel) using Aperio CS2 (Leica Biosystems, Germany) and Hamamatsu NanoZoomer 360 (Japan) digital slide scanners. Each image was assigned a unique anonymised-linked identifier and saved as SVS/NDPI files compatible for uploading in QuPath® software (version 0.4.1)[134]. This powerful platform was chosen due to its flexible annotation tools and in-built functions for automated nuclei detection and cell classification.

7.1.2.3 *Training of Machine Learning Models*

Two projects were created within QuPath® (one for each classifier) and the histology images were uploaded as brightfield H&E images. To overcome stain variability, the automatic "estimate stain vectors" feature within QuPath was used. This feature digitally separates the stains in the image using a method of colour deconvolution[135]. Representative regions containing relatively clear examples of the stains were selected to set the estimate stain vector and when the most optimal results were obtained, this was applied to the training sample.

The training slides were meticulously annotated and labelled into classes to highlight representative regions of interest (ROI) corresponding to dysplastic and non-dysplastic epithelial tissue, and immune and stromal cells in the subepithelial layer. Following the annotation phase, the cell detection algorithm in QuPath was run across all ROI using standardised nuclear (sigma 1.5 µm, minimum area 10 µm^2, maximum area 400 µm^2) and intensity parameters (threshold 0.04) at 10x magnification.

The following traditional ML classification algorithms were trained, and performances between the learned classifiers compared. Details of each model type are described below:

1) Artificial neural network-multilayer perceptron (ANN-MLP) – ANN mimic the human brain architectures. Each ANN consists of units (neurons) which are arranged in layers. Each unit takes an input, applies a function (often nonlinear) to it and then passes the output on to the next layer. Generally, outputs are fed forwards to the next layer, with no feedback to the previous layer. Hidden layers help solve more complex computational tasks, when a model includes more than 1 hidden layer, it is called a deep ANN. In this study, the number of hidden layers was optimised to achieve the best model accuracy. The final model had five hidden layers, 0 nodes, maximum iteration of 1000, and Epsilon of 0.01.

2) Random Forest (RF) – RF or random decision forests operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of over fitting to their training set. In this study, the model included 50 trees at a depth of 10 with a sample count of 10.

3) K nearest neighbour– this algorithm uses labelled points as a means of learning how to label new observations. In this way, the model learns the class label of the closest point or 'nearest neighbour'. Closeness is typically expressed in terms of a dissimilarity function. Once it checks with 'k' number of nearest neighbours, it assigns a label based on whichever label most of the neighbours have. In this study, the K value was adjusted to reach the most optimal results (K value of 10).

Using the top-performing classifier, analysis of nuclear features pertaining to the detected stromal and immune cells were extracted and evaluated. Grade-wise comparisons were made, and prognostic association of individual features analysed.

7.1.2.4 *Classification Performance*

The accuracy of the learned models was tested on representative ROI on unseen slides obtained from the School of Clinical Dentistry, Sheffield, UK. For reliable assessment, the number of cells in each ROI was kept consistent (300,000 +/- 5000 µm^2) across all test images. A breakdown of the test samples for each classifier is as follows:

- *Classifier D-ND* - 25 WSI; 35 ROI including 10 mild OED, 12 moderate OED, 9 severe OED and 4 non-dysplastic.
- *Classifier S-I:* 32 WSI; 67 ROI including 27 mild OED, 16 moderate OED and 24 severe OED.

Standard performance metrics including sensitivity, specificity, accuracy, precision, recall and F1 score were calculated to assess the model's ability to correctly identify the different classes. Individual class performance was measured for each ROI to generate an average score determined per image. The mean scores were aggregated to produce an overall accuracy, and comparisons made between models. Several of the classification performance metrics are mathematically defined below.

Precision: Ability of model to distinguish what portion of positive were actually positive.

$$\text{Precision} = \frac{True\ Positives}{True\ Positives\ +\ False\ Positives}$$

Recall: Determines what proportion of actual positives are identified correctly.

$$\text{Recall} = \frac{True\ Positives}{True\ Positives\ +\ False\ Negatives}$$

F1-score: A weighted mean between the precision and recall:

$$\text{F1} = \frac{2 * (precision * recall)}{(precision + recall)}$$

7.1.2.5 *Quantitative Spatial & Stromal Analysis*

Quantitative spatial and stromal analysis was conducted on the same sample as that described in the study by Mahmood *et al.* (2023)[10] presented in Chapter 5. Measurements corresponding to the distribution and density of cells were extracted using the inbuilt functions in QuPath® which measures two-dimensional distances to cell annotations and distances between cell centroids using principles of Delaunay Triangulation (µm) (Figure 20). Spatial cluster data was also extracted which included the number of neighbouring cells.

Cellularity quantification was conducted in the subepithelial ROI in OED area by counting the number of nuclei per unit area (in mm^3). Associated nuclear data (circularity, eccentricity, haematoxylin optical density, cytoplasm eosin optical density and nuclear to cell area ratio) were extracted and correlated to clinical outcomes. Similar stromal features, in addition to nuclear area and perimeter, were also analysed at the WSI-level using a deep learning-based pipeline in Python. OED segmentations were generated for each WSI using a Transformer model[136] and a Convolutional Neural Network-based method (HoVer-Net+[137]) used for the segmentation of epithelium and nuclei in the WSIs. Individual feature metrics were generated, and an average obtained for all nuclei in the stromal area to produce an overall WSI-level value.

7.1.2.6 *Statistical Methods*

Statistical comparisons between OED grades were conducted, and feature-specific associations with malignant transformation and recurrence determined. All statistical tests were conducted using GraphPad Prism® (version 9.3.1). Continuous data was tested for normality using the Shapiro-Wilk or D'Agostino & Pearson tests. Where normal distribution was assumed, an unpaired Student T test or one-way ANOVA with an applicable post hoc analysis (Tukey's or Dunnett's) was conducted for pairwise comparisons. Otherwise, an equivalent non-parametric test (Mann-Whitney or Kruskal-Wallis) was conducted with Dunn's pairwise comparisons based on rank sums. Feature-specific correlation to malignant transformation and OED recurrence was conducted using binary logistic regression. All tests were two-tailed, and p values < 0.05 considered significant. An overview of the study design is demonstrated in Figure 21.

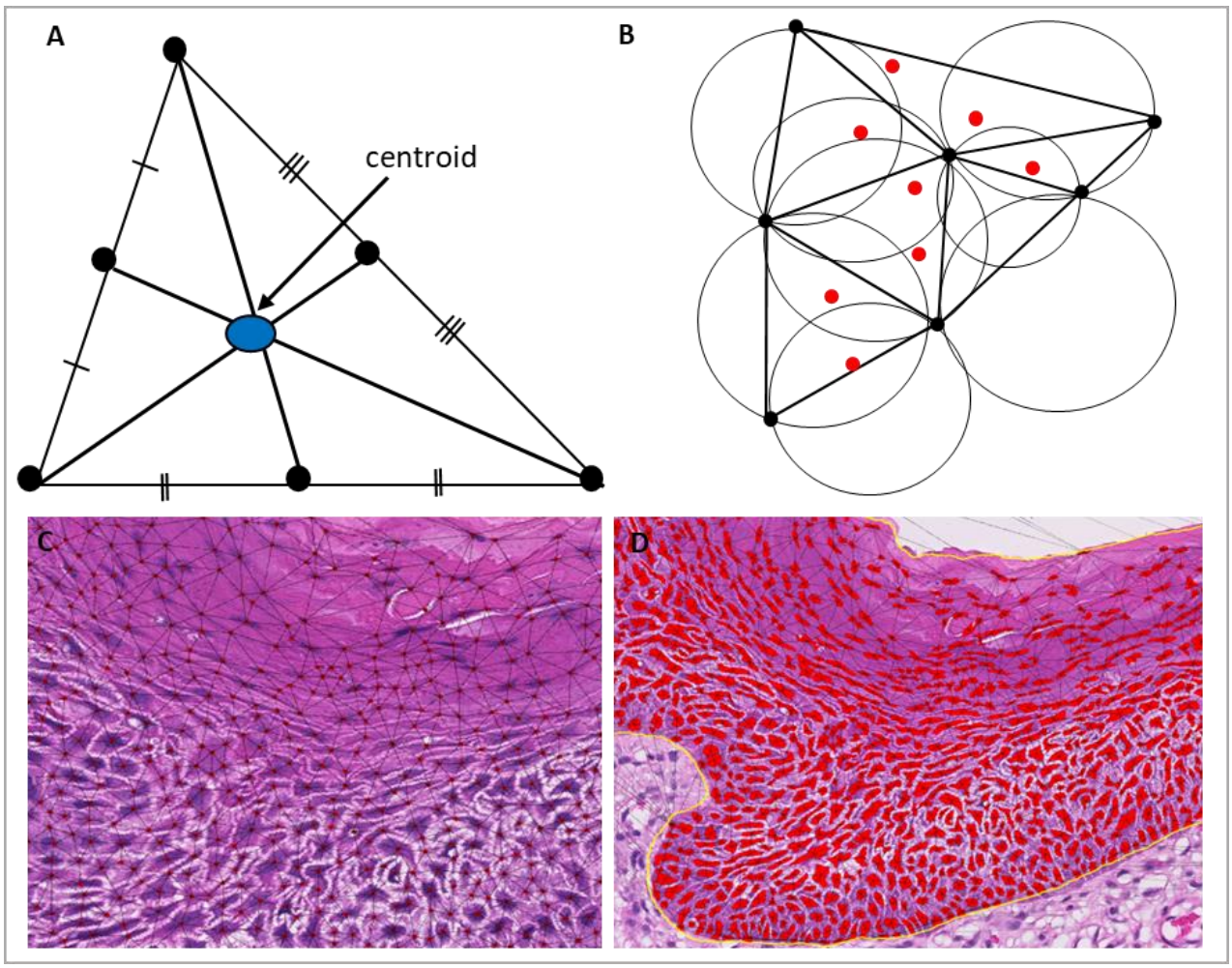

***Figure 20**. Geometric and mathematical principles for spatial analysis*

A) centroid of a two-dimensional triangle; B) Delaunay-based triangle connections; C-D) Delaunay spatial cluster analysis based on cell centroid detections. A Delaunay triangulation is a point-wise structure consisting of non-overlapping triangles. Where there is a discrete set of points (P) in Delaunay triangulation DT(P) no point in P is inside the circumcircle of any triangle in DT(P).

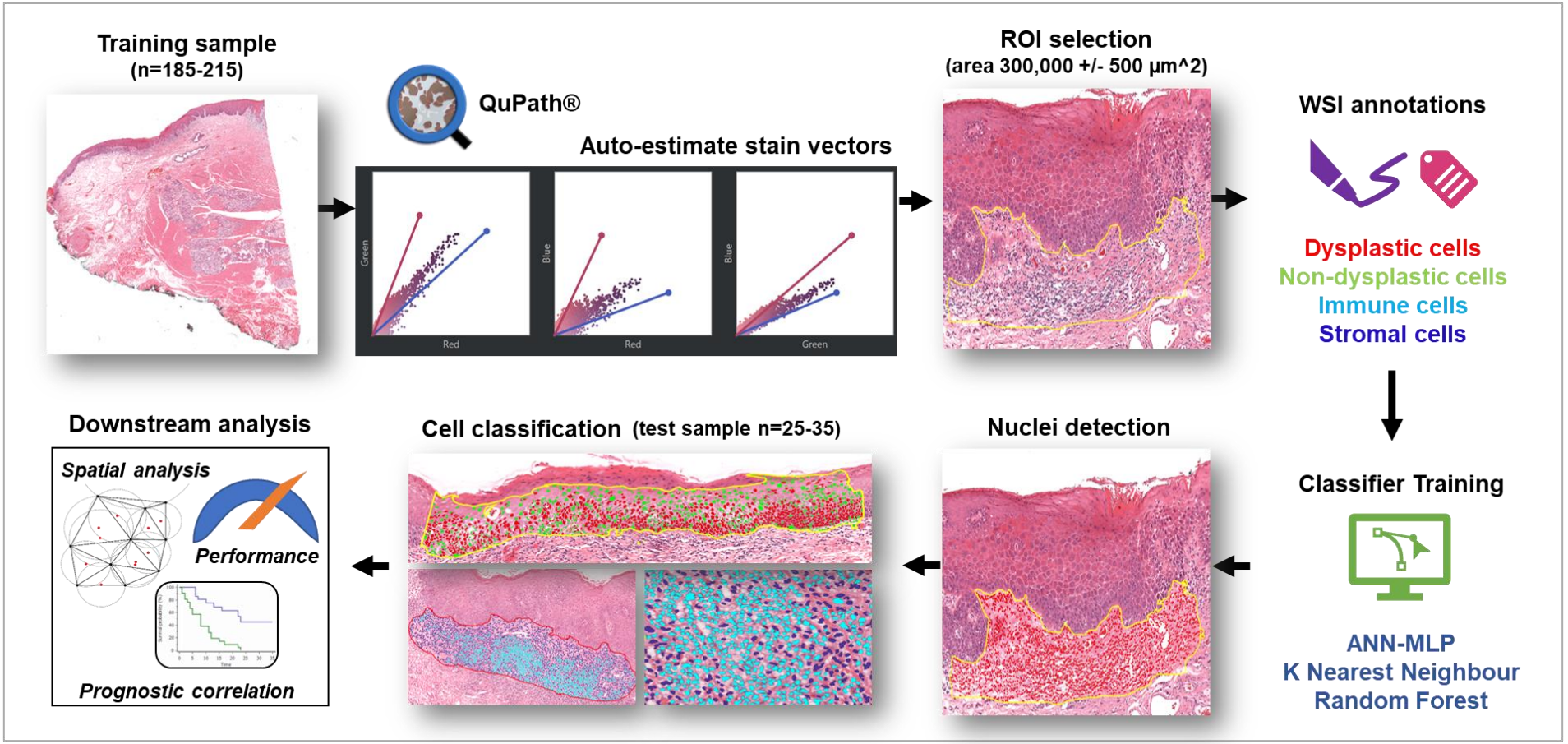

Figure 21. *Overview of study design and methods.*

7.1.3 Results

7.1.3.1 *Performance of ML classifiers*

The performance of the trained models for classification of dysplastic oral epithelial cells (n=25; 35 ROI) is presented in Table 9. The ANN-MLP classifier demonstrated the highest overall F1 score (0.89) followed by K nearest neighbour (0.77) and Random Forest (0.70) (Figure 22). demonstrates performance of the top performing ANN-MLP classifier on a range of cases including moderate OED (A), mild OED (B) and non-dysplastic control tissue (C).

Table 9. *Classification accuracy of dysplastic cells (n=25 WSI; 35 ROI)*

	Classification of dysplastic epithelial cells				
Model	**F1-score**	**Recall**	**Precision**	**Specificity**	**Sensitivity**
ANN-MLP	0.897	0.978	0.844	0.844	0.978
K Nearest Neighbour	0.771	0.735	0.862	0.762	0.735
Random Forest	0.701	0.806	0.744	0.618	0.806

The ANN-MLP model was then trained to detect immune and stromal cells in the subepithelial layer of OED. The F1 score for classification of immune cells was 0.78, with good recall and precision scores of 0.97 and 0.71, respectively. Similarly, classification of stromal cells (including endothelial and fibroblasts) based on the ANN-MLP model was good, with an F1 score of 0.94, recall of 0.92 and precision of 0.96 (n=32 WSI; 67 ROI) (Table 10). Examples of cell classifications are shown in Figure 22.

Digital nuclear feature analysis demonstrated significant differences in ‘nuclear eccentricity’ in immune cells between OED grades (mild vs moderate vs severe OED; p=0.0203) and between mild vs moderate OED (p=0.0256) (Table 11). There were no significant prognostic associations between nuclear features in either immune or stromal cells in OED using *Classifier S-I.*

Table 10. *Classification accuracy of immune and stromal cells (n=32 WSI; 67 ROI)*

	Classification Performance				
ANN-MLP Model	**F1-score**	**Recall**	**Precision**	**Specificity**	**Sensitivity**
Immune cells	0.785	0.971	0.718	0.904	0.971
Stromal cells	0.941	0.921	0.967	0.953	0.921

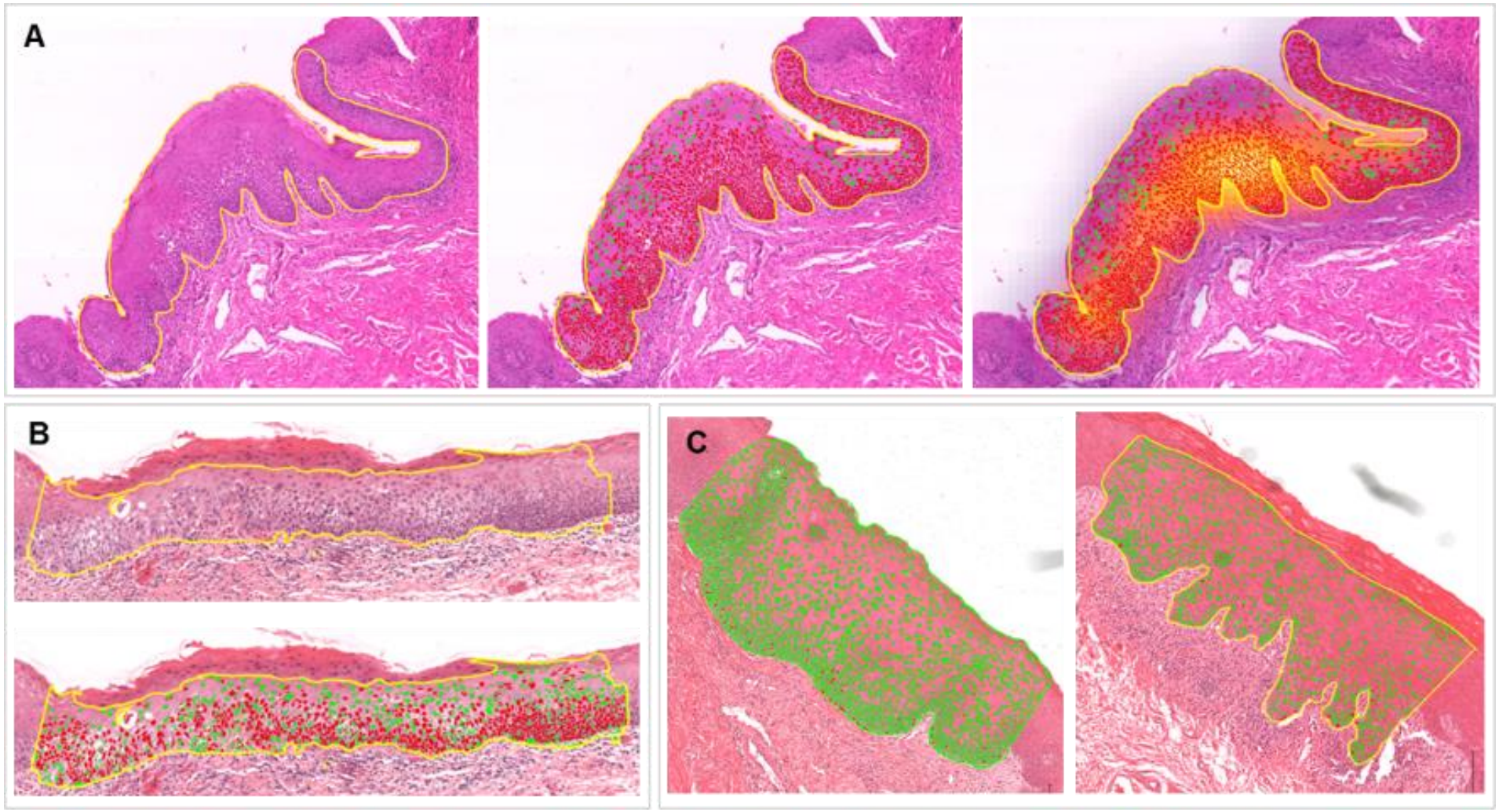

Figure 22. *Classification of dysplastic cells*

Performance of ANN-MLP model (n=210: training n=185; testing n=25). (A-B) dysplastic vs non-dysplastic cells with heat map highlighting region of greatest dysplastic activity; (C) non-dysplastic tissue including oral lichen planus (left) and hyperkeratosis/inflammation (right). Red = dysplastic cells; green = non-dysplastic cell.

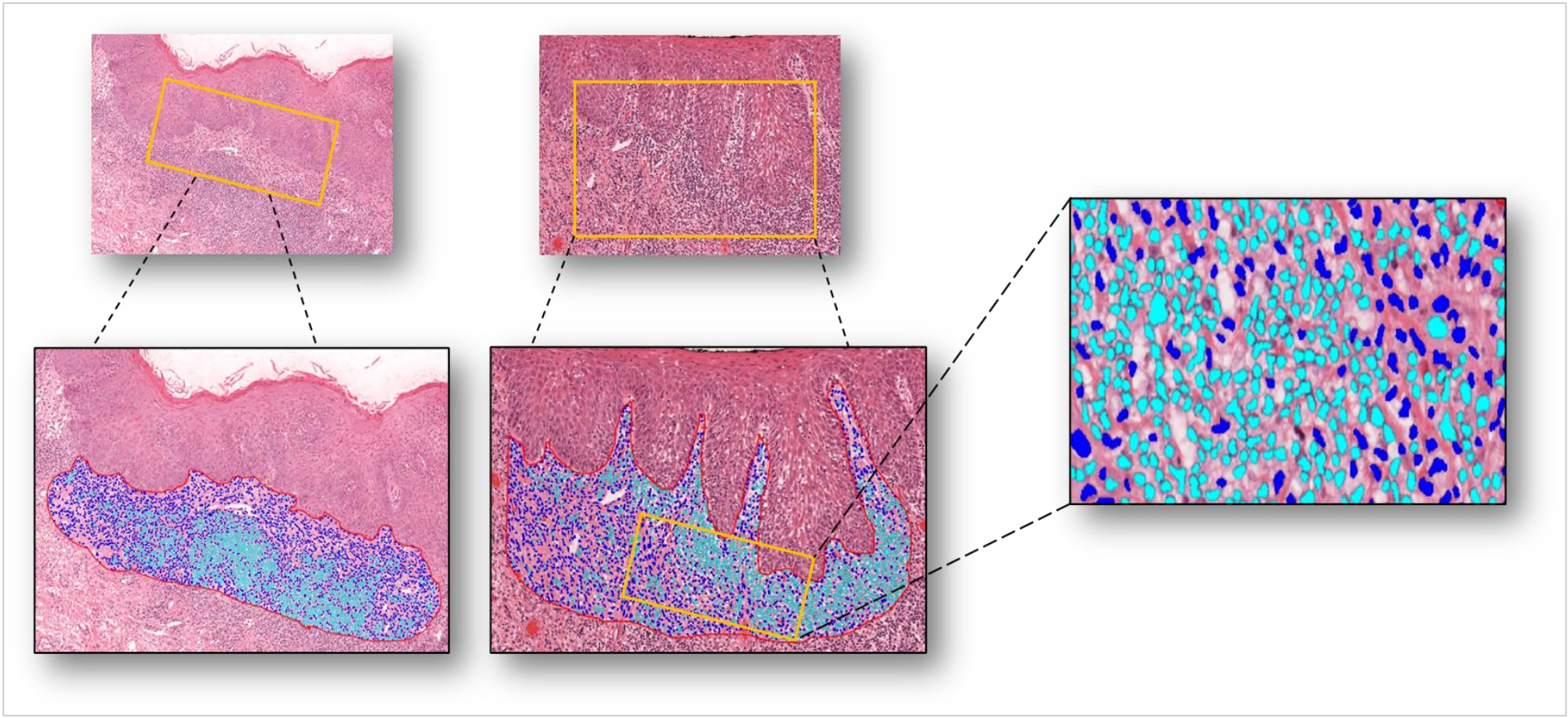

Figure 23. *Classification of immune and stromal cells in OED.*

Performance of ANN-MLP model (n=250: training n=215; testing n=32). Cyan = immune cells; navy blue = other stromal cells (fibroblasts, endothelial cells).

Table 11. *Nuclear feature analysis using ANN-MLP classifier.*

*Classifier S-I: training n=215; testing n=32 WSI; 67 ROI. OD = optical density. Asterisk highlights a statistical significance (*p<0.05) and displayed p values are corrected for multiple pairwise comparisons.*

	Nuclear area		Nuclear Perimeter		Nuclear circularity		Nuclear Eccentricity		Nuclear haematoxylin OD		Cytoplasm eosin OD		Nucleus to cell area ratio	
	Immune cells	**Stromal cells**	**Immune cells**	**Stromal cells**	**Immune cells**	**Stromal cells**	**Immune cells**	**Stromal cells**	**Immune cells**	**Stromal cells**	**Immune cells**	**Stromal cells**	**Immune cells**	**Stromal cells**
WHO grading	0.288	0.5828	0.9044	0.184	0.0655	0.3504	**0.0203***	0.2104	0.9533	0.337	0.3278	0.2653	0.0961	0.0854
Mild vs Moderate OED	0.2785	>0.9999	0.9385	0.9686	0.064	0.9988	**0.0256***	0.9177	0.9921	>0.9999	0.9976	0.4747	0.0893	0.1486
Mild vs Severe OED	0.5029	0.6385	0.9059	0.2049	0.1998	0.4065	0.0616	0.2111	0.9808	0.4029	0.7851	0.9023	0.2825	0.1125
Moderate vs Severe OED	0.9031	0.6403	0.9961	0.3018	0.8231	0.4322	0.9161	0.3867	0.9495	0.4082	0.7469	0.2575	0.7962	0.9879
Binary grading	0.4695	0.8918	0.503	0.2721	0.2995	0.2103	0.1464	0.1406	0.6957	0.2225	0.2566	0.3229	0.0988	0.1404

7.1.3.2 *Spatial Analysis in OED*

Distance between cells

There was a significantly reduced distance between cells in OED (mild, moderate and severe grades combined) compared with non-dysplastic cells (p=0.026). The average distance between cells reduced as OED grade severity increased, and statistical differences were seen on pairwise comparisons between moderate OED vs control (p=0.0475), severe OED vs control (p=0.0169) and high-grade OED vs control (p= 0.0141) (Figure 24 and **Error! Reference source not found.**).

Area between cells

Similar trends were seen for the area between cells, with a significantly lower mean triangle area between cells in OED (mild, moderate and severe grades combined) compared with cells in the control group (p= 0.0289). The area between OED cells also decreased as the grade worsened. Pairwise comparisons demonstrated significant differences in triangle area between moderate OED vs control (p= 0.0396), severe OED vs control (p= 0.0178) and high-grade OED vs control (p= 0.0074) (Figure 24 and **Error! Reference source not found.**).

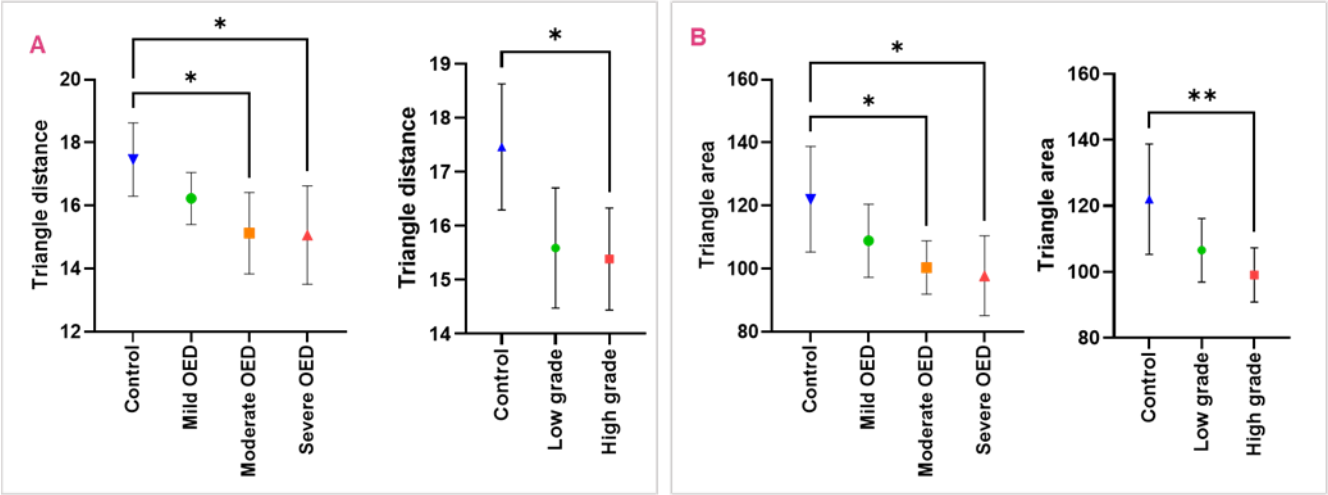

Figure 24. *Delaunay spatial analysis*

Grade-based analysis (WHO and binary) demonstrating triangle distance (A) and triangle area (B) between cell nuclei (OED n=75, control n=25).

*Data expressed as mean values with 95% CI error bars. Asterisk highlights a statistical significance (*p<0.05).*

Number of neighbouring cells

The number of neighbouring cells in OED was marginally higher in moderate and severe OED groups (mean 5.965 and 5.97 respectively) compared mild OED (mean 5.95, SD 0.07951), though this difference was not statistically significant. There were no significant differences when comparing neighbouring cell number between OED groups and control group (p=0.33) (Figure 25).

Individually, none of the examined Delaunay features were statistically associated with either malignant transformation or OED recurrence within the studied cohort (**Error! Reference source not found.**).

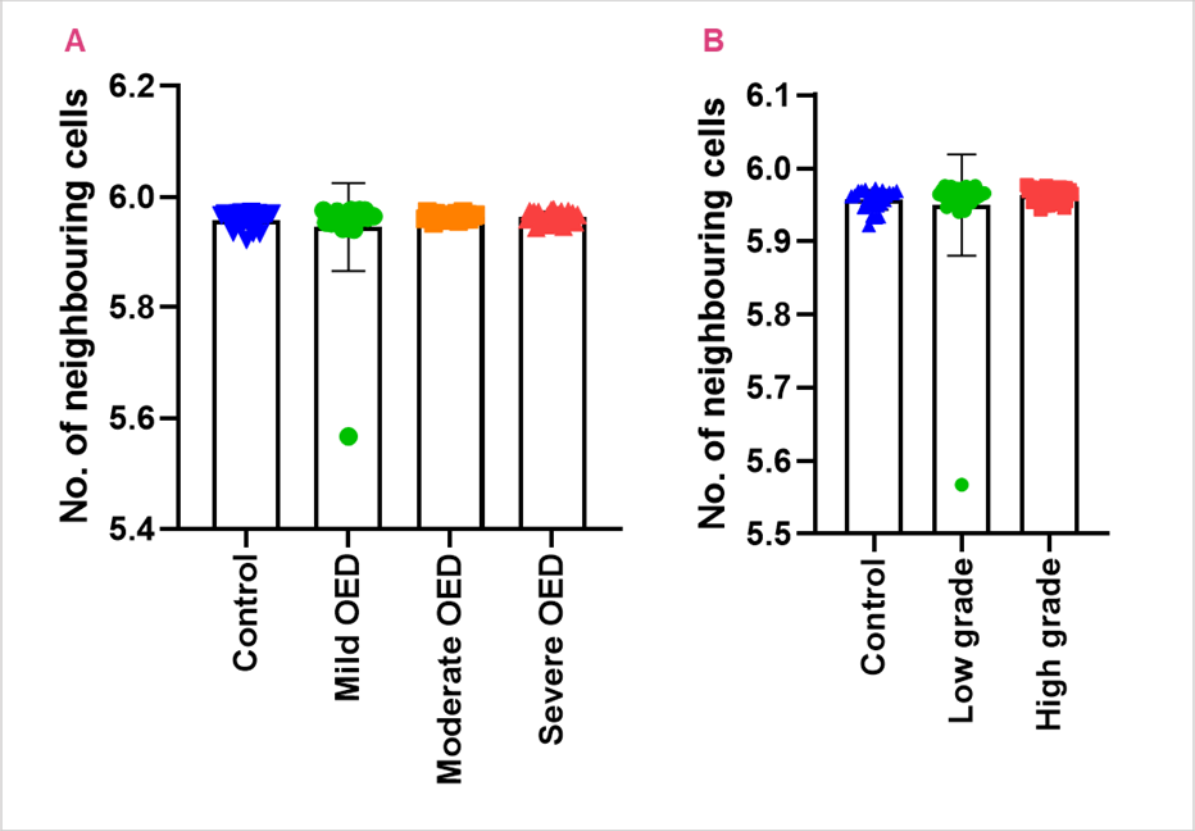

Figure 25. *Number of neighbouring cells in OED.*

Grade-based analysis (WHO and binary) of number of neighbouring cells (OED n=75, control n=25).

Data expressed as mean values with standard deviation error bars.

Table 12. *Spatial analysis in OED with comparison to control and prognostic associations. Asterisk highlights statistical significance (*p<0.05).*

	No. neighbour cells	Triangle distance	Triangle area
WHO grades	0.4936	0.3166	0.3139
Mild vs moderate OED	>0.9999	0.9436	0.5043
Mild vs severe OED	0.8299	0.4119	0.3122
Moderate vs severe OED	>0.9999	>0.9999	0.9358
Binary grades	0.2983	0.5138	0.2332
OED vs control	0.2561	**0.026***	**0.0289***
Mild OED vs control	>0.9999	0.5032	0.3079
Moderate OED vs control	0.3271	**0.0475***	**0.0396***
Severe OED vs control	0.2618	**0.0169***	**0.0178***
Low grade OED vs control	0.773	0.0836	0.1053
High grade OED vs control	0.1166	**0.0141***	**0.0074***
Transformation association	0.3423	0.387	0.9059
Recurrence association	0.3687	0.4411	0.7576

Distance between cells using Classifier D-NP

Distances between dysplastic and non-dysplastic cells were generated using the ANN-MLP model for *Classifier D-ND* (test sample n=25, 35 ROI). Measurements of relative distance were calculated based on cell centroid detections. Findings demonstrated the average distance between detected dysplastic cells was 11.4507 µm (SD 12.876). Distances between mildly dysplastic cells were the greatest (20.3374 µm, SD 30.44), followed by severely dysplastic cells (10.9326 µm, SD 6.800). Distances between moderately dysplastic cells were the lowest (3.008 µm, SD 2.721). The average distance to detection between non-dysplastic cells amongst the selected ROI was 21.5155 µm (SD 17.6729) with highest centroid distances between cells in moderate OED (30.7892 µm, SD 20.51), followed by mild OED (24.166 µm, SD 27.43) and severe OED (7.6038 µm, SD 9.434). There were no grade-wise differences for distances between either dysplastic or non-dysplastic cells in the studied test sample.

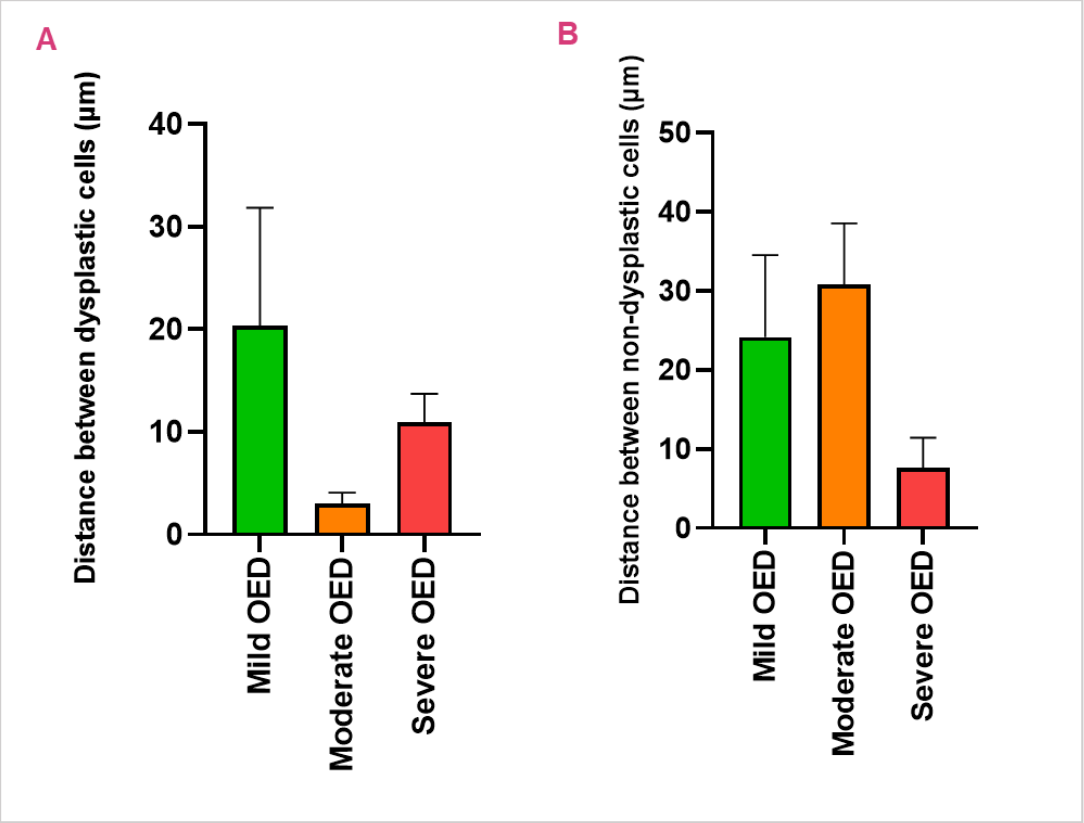

Figure 26. *Cell centroid distances in OED*

Measurements based on ANN-MLP model detection using Classifier D-NP (n=25; 35 ROI)

Values expressed as mean with error bars representing standard error.

7.1.3.3 *Quantitative stromal analysis*

Stromal cellularity

Quantitative analysis at ROI-level demonstrated a significantly higher number of stromal cells in the subepithelial layer of OED (mild, moderate and severe groups combined) compared to control ($p<0.0001$). The mean number of stromal cells was also found to increase with grade severity; mild OED 2956 (SD 1044), moderate OED 3085 (SD 986.4), severe OED 3091 (SD 1029) compared to 968.8 (SD 652.1) for control. Further pairwise comparisons demonstrated statistical differences between mild OED vs control ($p<0.0001$), moderate OED vs control ($p<0.0001$) and severe OED vs control ($p<0.0001$). Evaluation of binary grading showed similar trends of significance between low-grade OED vs control ($p<0.0001$) and high-grade OED vs control ($p<0.0001$). Similar trends were also seen at WSI-level, but these findings were not statistically significant.

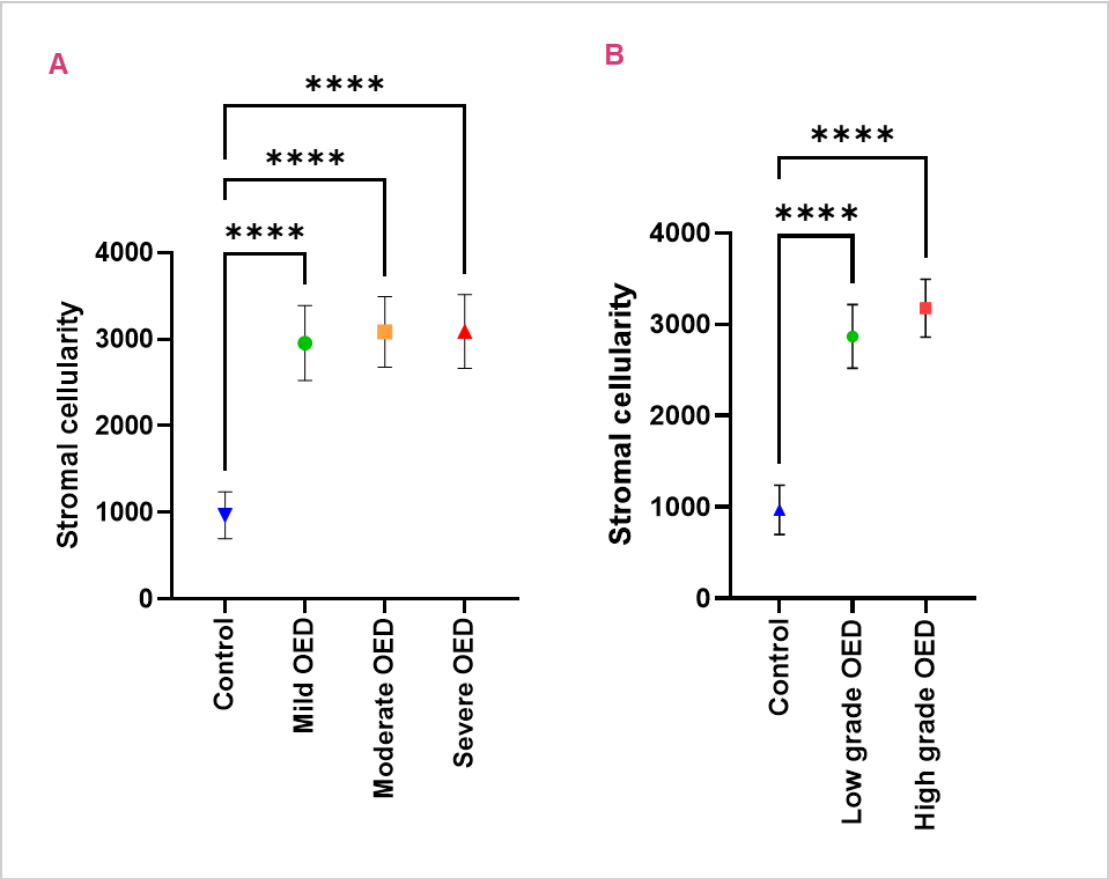

Figure 27. *Stromal cellularity in OED (at ROI-level)*

Grade-based analysis demonstrating stromal cellularity in OED compared to control (OED n=75, control n=25). (A) WHO grade comparisons; (B) Binary grade comparisons.

*Data expressed as mean values with 95% CI error bars. Asterisk highlights a statistical significance (*p ⩽ 0.05, **p ⩽ 0.01, ***p < 0.001, ****p ⩽ 0.0001).*

Nuclear feature analysis in stromal cells

At the ROI-level, two features were shown to be statistically different between OED and non-dysplastic/control groups, these being 'nuclear haematoxylin OD' and 'nuclear to cell area ratio' (Figure 28). There was a significantly higher detection of 'nuclear haematoxylin OD' in OED (mild, moderate and severe grades combined) compared to the control group (p=0.0062), with further significance when comparing control vs moderate OED (p=0.0393) and control vs severe OED (p=0.0019). Similar findings were also seen on binary grades, with statistical differences observed between high-grade OED vs control (p=0.0013) (Figure 28). The 'nuclear to cell area ratio' was also significantly different between OED groups (mild, moderate and severe grades combined) and the control group (p<0.0001). Pairwise comparisons showed further differences between moderate OED vs control (p=0.0024) and severe OED vs control (p<0.0001). A similar statistical trend was also seen

when comparing binary OED grades (low grade and high grades combined) to control (p<0.0001) as well as high-grade OED vs control (p<0.0001) and low grade vs high grade OED (p=0.0052) (Figure 28).

WSI-level analysis demonstrated a contrasting finding, with a statistically lower detection of 'nuclear haematoxylin OD' in OED compared to control (p<0.0001). The nuclear 'contour area' and 'perimeter' were also significantly smaller in OED (all grades combined) compared to control (p<0.0001).

Prognostic relationships of stromal features

Stromal cellularity in OED was not shown to be statistically associated with either malignant transformation or recurrence within this cohort (both at ROI and WSI-level). However, several nuclear features of stromal cells in OED were shown to be statistically correlated to clinical outcomes at ROI level (Table 13). Specifically, 'nuclear circularity' (p=0.001), 'nuclear eccentricity' (p=0.003), 'cytoplasm eosin OD' (p=0.045) and 'nuclear to cell area ratio' (p=0.003) were associated with an increased incidence of malignant transformation. These same features, with the exception of 'cytoplasm eosin OD', were also associated with an increased risk of lesion recurrence (p<0.05) (Table 13).

WSI-level analysis also showed that 'nuclear circularity' and 'nuclear eccentricity' were associated with an increased risk of malignant transformation (p<0.05).

Table 13. *Prognostic relationships of nuclear features in stromal cells (at ROI-level)*

*Analysis conducted on OED n=75. Asterisk highlights statistical significance (*p<0.05). OD = optical density.*

Nuclear feature	**Malignant transformation** (*p* value)	**OED Recurrence** (*p* value)
Nuclear circularity	**0.001***	**0.040***
Nuclear eccentricity	**0.003***	**0.022***
Nuclear haematoxylin OD	0.092	0.580
Cytoplasm eosin OD	**0.045***	0.089
Nuclear to cell area ratio	**0.003***	**0.009***

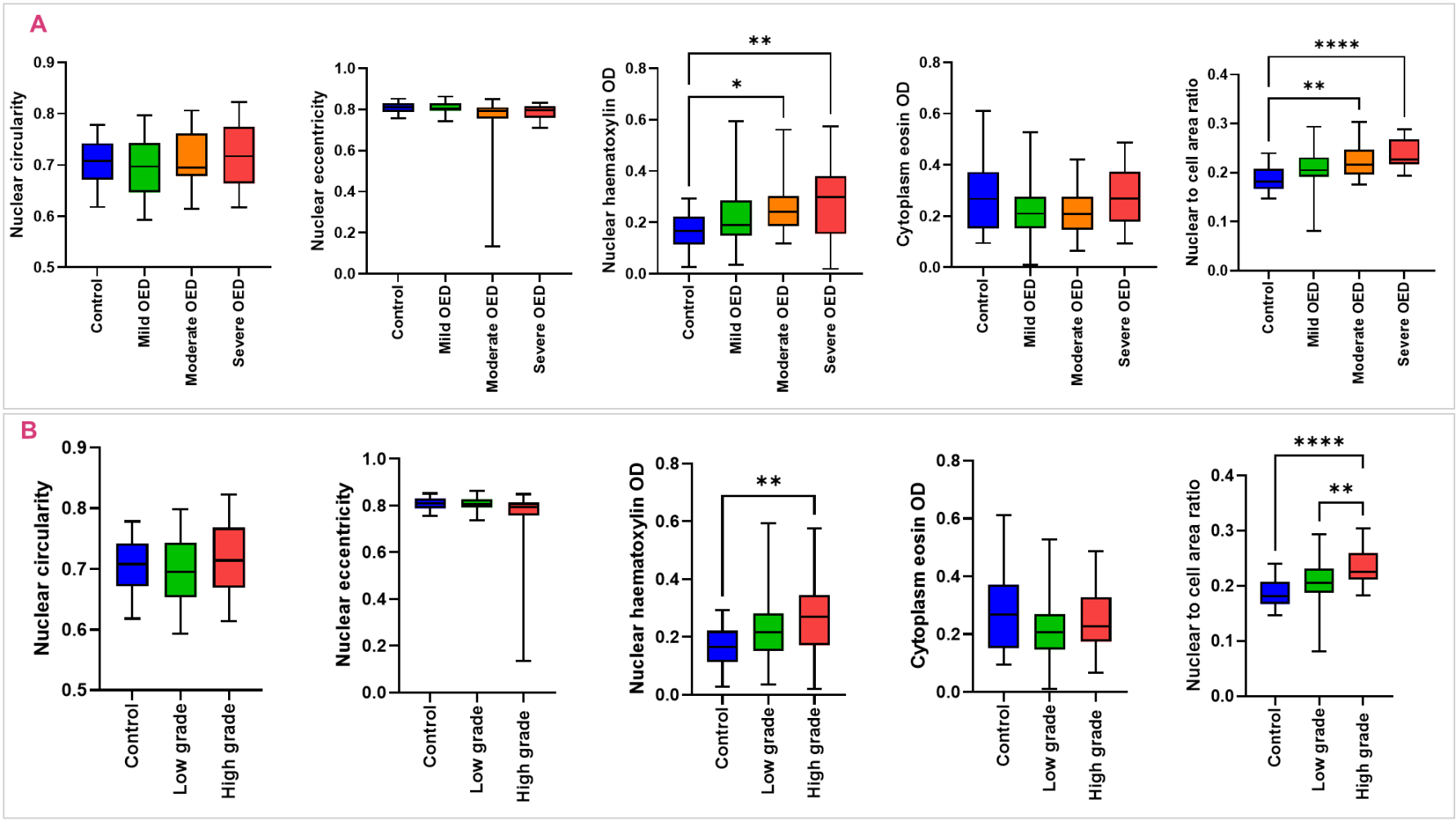

Figure 28. *Nuclear features analysis of stromal cells in OED (at ROI-level).*

*Grade-based analysis of nuclear features pertaining to stromal cells in subepithelial layer of OED (n=75) compared to control (n=25). (A) WHO grading; (B) binary grading. Data expressed as mean values with error bars representing minimum to maximum ranges. Asterisk highlights statistical significance (*p ⩽ 0.05, **p ⩽ 0.01, ***p < 0.001, ****p ⩽ 0.0001).*

7.1.4 Discussion

A large proportion of OSCC diagnoses are made at a late stage of disease, impacting on patient mortality and morbidity[138]. It is therefore important that OED, the precursor disorder, is detected early and accurately so that optimal intervention can be provided[139]. Digital pathology represented by whole slide imaging has gained attention for its potential to overcome subjectivity in diagnosis and reduce misclassification by providing quantifiable outputs for more objective risk prediction. This study demonstrates the promising potential for ML algorithms to reliably classify dysplastic, immune, and stromal cells in OED, and conducts a quantitative analysis to reveal novel findings pertaining to the spatial arrangements of cells and prognostic relationships between nuclei features in stromal cells of OED.

Among the ML models explored in this study, the ANN-MLP algorithm demonstrated the best detection accuracy for classification of dysplastic cells in OED epithelium. This model was therefore also trained to classify immune cells and other stromal cells in OED (i.e. fibroblasts and endothelial cells). ANN-MLP is a traditional supervised ML model which mimics the operation of a human brain, featuring three component layers, namely an input layer (like human dendrites), a hidden layer, and an output layer (like human axons)[49]. The hidden layers are designed to solve more complex computational tasks, and where a model includes more than 1 hidden layer, it is called a deep ANN. In this study, five hidden layers were used to achieve the best classification results yielding F1 scores >0.78 for classification of dysplastic, immune and stromal cells in OED (Table 9, Table 10). The optimised classifiers were then used to extract quantitative data related to nuclei morphometry for both immune and stromal cells in the test sample. Findings demonstrated grade-based differences (mild vs moderate vs severe OED) for ‘nuclear eccentricity’ in immune cells ($p=0.0203$) (Table 11) but there were no prognostic associations for these features with malignancy in the explored dataset.

The authors were also interested in understanding the spatial patterns of cells in OED tissue. One of the perceived challenges in evaluating the spatial arrangements of cells in a three-dimensional space is how to reliably capture true biological relationships from two-dimensional tissue sections. Several studies have shown that certain aspects of spatial organisation can be recognised and conceptualised using mathematical models[131, 140]. Delaunay-based cluster spatial analysis is a computational geometric technique that applies an algorithm to detect cells based on their centroid location (Figure 20). This helps to identify clusters of cells neighbouring one another and allows the potential to extract measurements between cells based on centroid locations[141]. This approach can help measure cellular density and organisation based on cell centroid detection in tissue. In this study, quantitative spatial analysis demonstrated a reducing Delaunay triangle area and distance as the OED grade severity increased. Unsurprisingly, these differences were notably more prominent between extreme groups (control and severe OED; mild and severe OED) and less prominent between groups where greater ambiguity in grading is likely to occur (moderate and severe OED). These findings could also be explained by the increased level of cellular pleomorphism and disorganisation in more extensively

dysplastic lesions[54,42]. If the number and size of cells is increased per unit area, then it may explain why the relative triangular area and distance between dysplastic cells to be relatively less. Furthermore, cell crowding (particularly in the basal layer) is commonly seen in dysplasia[142], so the more organised and densely packed dysplastic cells are, the relatively smaller the triangle area/distance between cells is likely to be (in comparison to normal keratinocytes). Significant differences were also seen in the triangle distance and area between cells in the control group in comparison with cells in OED groups (p= 0.026 – 0.028). Pairwise comparisons demonstrated further statistical differences between moderate OED vs control (p= 0.039 – 0.047), severe OED vs control (p= 0.016 – 0.017) and high-grade OED vs control (p= 0.0074 - 0.0141). There were no statistical associations with malignant transformation or OED recurrence with the studied Delaunay features though further exploration on larger datasets is needed to establish this. Spatial clustering analysis demonstrated the number of neighbouring cells in moderate and severe OED to be marginally greater than in mild OED, however, this difference was not statistically significant.

Quantitative analysis of stromal cells in the subepithelial layer of OED demonstrated a significantly higher number of cells compared to control, which was also consistently seen when comparing individual grades (using WHO and binary systems) to the control group as well ($p<0.0001$). Nuclear feature analysis in stromal cells showed that 'nuclear haematoxylin OD' and 'nuclear to cell area ratio' were statistically different between OED and non-dysplastic/control groups (p=0.0001-0.0062). Whilst stromal cellularity itself was not shown to be a prognostic feature, several nuclear features pertaining to stromal cells in OED were associated malignant transformation, including 'nuclear circularity' (p=0.001), 'nuclear eccentricity' (p=0.003), 'cytoplasm eosin OD' (p=0.045) and 'nuclear to cell area ratio' (p=0.003). These features (except for cytoplasm eosin OD) were also associated with an increased risk of lesion recurrence ($p<0.05$).

The authors acknowledge a few limitations of this study. First, the analysis of nuclear features in stromal and immune cells (using *Classifier S-I) and* spatial clustering (using *Classifier D-ND)* was conducted at the ROI-level on relatively small test sets. The main benefit of ROI-based analysis in this work was that it provided a focussed analysis of dysplastic epithelium and the subepithelial stromal region in OED. This was particularly useful since the study included a varying cohort of OED lesions in which dysplasia was not necessarily present throughout the tissue section, and it also allowed us to avoid regions which could affect model training, such as regions of ulceration or inflammation. As ROIs were selected manually, to minimise the risk of bias, the ROI were verified by a Consultant Oral & Maxillofacial Pathologist (SAK) and Senior Specialist Registrar (PH). With regards to the sample size, it is important to also highlight the application of ML in this study was mainly focussed on evaluating the performance of models to classify cells in OED, and the downstream statistical analysis was largely exploratory work. Future work should expand the test set to further assess model generalisability and evaluate the studied features, particularly since the latter part of our study (in

which quantitative stromal analysis was conducted) identified potentially important nuclear features in stromal cells for prediction of malignancy transformation risk in OED.

This work provides a novel insight into the local spatial architecture of cells in OED tissue, based on quantitative digital WSI analysis. Further investigation using such approaches on larger independent datasets with inclusion of OSCC tissue will allow a greater understanding of the relative importance of the examined features for diagnostic and prognostic purposes. This study demonstrates the promising potential for ML models to accurately detect and classify dysplastic, immune and stromal cells in OED, and we learn that general spatial arrangements and more local interrelations between cells may have potential importance in OED progression. Cell organisation in tissue is critical to understanding complex disease pathology, and this work highlights the need for further research in this regard.

DATA AVAILABILITY

All data generated or analysed during this study are included in this published article.

ACKNOWLEDGEMENTS

Paul Hankinson (Academic Clinical Fellow in Oral Pathology, University of Sheffield) is acknowledged for his contribution to histological grade assessment.

AUTHOR CONTRIBUTIONS

HM and SAK conceived and designed the study. HM prepared the digital cohort, collected clinical data, conducted the digital experiments, quantitative analyses, classifier training and testing. PH and SAK conducted histological grade analysis. AS performed the deep learning-based stromal analysis. HM, AS, NMR and SAK were involved in analysis of results. HM led on writing of the manuscript with contributions from all authors.

FUNDING & COMPETING INTERESTS

HM is funded by the National Institute for Health Research (Award ID: NIHR300904). NR is the Director and CSO of Histofy, an AI start-up in the area of computational pathology. SAK is a shareholder of Histofy. The remaining authors do not declare any competing interests.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval was granted by the West Midlands - Edgbaston Committee (reference: 18/WM/0335).

Chapter 8 – Deep Learning Models

Chapter 8 – Deep Learning Models

This chapter is formed by a publication-ready manuscript for which the candidate is joint first author with Dr Adam Shephard, a Data Scientist at the University of Warwick. This study develops a novel AI pipeline for OED segmentation, classification and transformation prediction using H&E-stained whole slide images.

The candidate's contributions to these studies were:

i. obtaining ethical approval
ii. slide retrieval and digital dataset preparation
iii. clinical data collection
iv. ground-truth slide annotations
v. co-development, training and testing of DL models
vi. downstream statistical analyses
vii. writing up and editing the manuscript

The computational aspects of the study (including writing of the code) were led by Dr Adam Shephard under the guidance of Dr Shan Raza and Professor Nasir Rajpoot at the Tissue Image Analytics Centre, University of Warwick.

Datasets for model validation were acquired from external national and international centres including: 1) Precision Medicine Centre, Patrick G. Johnston Centre for Cancer Research, Queen's University Belfast, UK; 2) Institute of Head and Neck Studies and Education (InHANSE), Birmingham, UK and 3) Piracicaba Dental School, UNICAMP, Brazil.

8.1 Development and Validation of an Artificial Intelligence-based Pipeline for the Prediction of Malignant Transformation in Oral Epithelial Dysplasia: A Retrospective Multi-Centric Study

Adam J Shephard[1†], **Hanya Mahmood[2†],** Shan E Ahmed Raza[1], Anna Luiza Damaceno Araujo[3], Alan Roger Santos-Silva[3], Marcio Ajudarte Lopes[3], Pablo Agustin Vargas[3], Kris D. McCombe[4], Stephanie G. Craig[4], Jacqueline James[4], Jill Brooks[5], Paul Nankivell[5], Hisham Mehanna[5], Syed Ali Khurram[2⊥], Nasir M Rajpoot[1⊥]

[1]Tissue Image Analytics Centre, Department of Computer Science, University of Warwick, UK

[2]School of Clinical Dentistry, University of Sheffield, Sheffield, UK

[3]Oral Diagnosis Department, Piracicaba Dental School University of Campinas, São Paulo, Brazil

[4]Patrick G. Johnston Centre for Cancer Research, Queen's University Belfast, UK

[5]Institute of Head and Neck Studies and Education, University of Birmingham, UK

† Joint first authorship

⊥ Joint co-senior authorship

Abstract

Background: Oral epithelial dysplasia (OED) is a premalignant histopathological diagnosis given to lesions of the oral cavity that have an increased risk of progression to malignancy. The grading of OED is subject to wide inter/intra-rater variability, and does not reliably predict prognosis, potentially resulting in sub-optimal treatment decisions. To address this, we developed an artificial intelligence (AI) pipeline for OED segmentation, classification and transformation prediction in haematoxylin and eosin (H&E) stained whole slide images (WSIs).

Methods: We have developed a novel transformer-based pipeline (ODYN) that can both classify OED and assign a predictive score (ODYN-score) to quantify the risk of malignant transformation. We use a shallow neural network to assign slide-level ODYN-scores, based on patch-level nuclear features in dysplastic tissue. The model was trained on a large digital dataset using three different scanners (Sheffield, 358 OED WSIs, 105 control WSIs) and externally validated on cases from three independent centres (Birmingham and Belfast, UK, and São Paulo, Brazil; 108 OED WSIs).

Findings: Model testing yielded an F1-score of 0.71 for OED segmentation, and 0.96 for classification of dysplastic vs non-dysplastic tissue. Our AI pipeline gained an AUROC of 0.75 for malignancy prediction, outperforming other state-of-the-art methods, and existing clinical grading systems.

Interpretation: We present a new Transformer-based model for the semantic segmentation and classification of OED using the largest multi-centric OED dataset to date (496 WSIs). This is the first study to use transformers for semantic segmentation in oral histology WSI, demonstrating reliable and promising results to aid OED diagnosis and prognosis prediction. This study demonstrates the promise of computational pathology for the automatic detection, diagnosis and prognosis of OED.

Funding: Cancer Research UK Early Detection Project Grant (grant no. C63489/A29674) and National Institute for Health Research (NIHR300904).

Keywords: Oral epithelial dysplasia; transformer; semantic segmentation; dysplasia detection; malignant transformation; artificial intelligence; deep learning

8.1.1 Introduction

Oral epithelial dysplasia (OED) presents a significant challenge in the realm of oral pathology, where accurate diagnosis and early detection are paramount for effective intervention and the prevention of malignant progression[54]. OED is a premalignant histopathological diagnosis encompassing various lesions of the oral mucosa, typically manifesting as white (leukoplakia), red (erythroplakia) or mixed red-white (erythroleukoplakia) lesions[42,46].

Histopathological grading of haematoxylin and eosin (H&E) stained tissue using the World Health Organisation (WHO, 2017) classification system remains the current accepted practice for diagnosis and risk stratification of OED lesions[52]. This is a three-tier system for grading OED into mild, moderate and severe groups based on the presence, severity and location of a wide range of cytological and architectural histological features (28 in total) [28,54]. This approach suffers from significant intra- and inter-observer variability and has unreliable predictive strength for malignant transformation risk, subsequently impacting on patient treatment. An alternate binary grading system, categorising lesions as low- or high-risk, based on the number of cytological and architectural features (as listed in the WHO criteria) aimed to improve the reproducibility of grading[57,60]. However, studies have shown significant variability in grading using both systems, highlighting the need for a more objective and reproducible method that can better predict malignant transformation in OED[54,61].

To address challenges in subjectivity and misclassification of precancerous and cancerous lesions, there is a growing interest in leveraging advanced technologies, particularly deep learning (DL), which has seen extensive use in medical image analysis over the past decade[143,144,145]. Several state-of-the-art models, such as U-Net[146] and DeepLab[147] have been developed to perform image classification and segmentation. These models typically use convolutional neural networks (CNN) as feature extractors, such as ResNet[148]. However, despite their success, these models often suffer from limitations such as high computational overhead and difficulty in capturing long-range dependencies in images.

Transformers have gained widespread attention in recent years as they have been successfully applied in several natural language processing and computer vision tasks such as classification[149,150,151]. A typical Transformer encoder consists of a multi-head self-attention (MSA) layer, a multi-layer perceptron (MLP), and a layer normalisation (LN). This MSA layer empowers Transformers to capture long-range dependencies, making them a strong candidate for semantic segmentation in medical images[152153154]. Transformers therefore have the potential to overcome some of the limitations of traditional CNNs. However, only a handful of methods have applied Transformers for semantic segmentation in medical images[152,155]. Their application in histology has primarily been constrained to classification tasks[156,157], with semantic segmentation left relatively unexplored. This raises the question of whether Transformers can be harnessed for semantic segmentation of histological images.

In this study, we aimed to develop a novel, weakly supervised, DL pipeline that could reliably and objectively segment and classify OED, whilst predicting the risk of malignant transformation in OED, using H&E-stained WSIs. Specifically, we aimed to do this using interpretable nuclear features, from dysplastic regions on the WSI. Moreover, we aimed to rigorously evaluate the performance of our pipeline by comparing it to other state-of-the-art methods. To demonstrate the robustness and generalisability of our approach, we aimed to develop our model based on a large cohort and extend our validation on unseen datasets acquired from three other national and international centres (Birmingham and Belfast, UK, and Brazil).

8.1.2 Methods

8.1.2.1 *Study Design*

In this retrospective multicentre study, we have developed an innovative weakly supervised method for predicting the progression of OED lesions to malignancy. We additionally aimed to produce a model that classifies oral tissue slides as being dysplastic vs non-dysplastic. We achieved this by analysing H&E-stained WSIs obtained from oral tissue biopsies, using a CNN, a Transformer and a MLP, in what we have called our Oral DYsplasia Network, “ODYN”.

In many cases of OED, histological atypia is not present across the entire tissue section, and thus, the first step of this work was to identify only the regions where dysplastic changes were present. A fine-tuned Transformer (based on Trans-UNet[152]) was used to detect and segment the different dysplastic areas in a given WSI. Next, we used a pretrained CNN (HoVer-Net+[137]) to segment the epithelium and the nuclei in the WSI.

For OED classification, we calculated the proportion of the epithelium mask that was segmented as dysplastic and used an empirically determined threshold to classify slides as being dysplastic vs. non-dysplastic. The remainder of the pipeline was then used for ODYN-scoring, with OED cases alone. We generated patch-level morphological features in the dysplastic regions, which were used as input

to a MLP to calculate a risk-score for malignancy progression (named the ODYN-score). Both the Transformer and MLP were trained during this experiment and then used for inference in the ODYN -scoring pipeline. However, HoVer-Net+ was used for inference alone, as it is a state-of-the-art model for epithelium and nuclear segmentation and classification, that has been extensively pre-trained on OED data.

8.1.2.2 *Study Cohorts*

Development and Internal Validation Cohort

The training cohort consisted of a retrospective sample of histology tissue sections (dating 2008 to 2016 with minimum five-year follow-up data) collected from the Oral and Maxillofacial Pathology archive at the School of Clinical Dentistry, University of Sheffield, UK (referred to as the internal centre, hereafter). After microscopic inspection of the tissue sections by a Consultant Oral & Maxillofacial Pathologist (SAK), newly cut 4 µm sections of the selected cases were obtained from formalin fixed paraffin embedded blocks and stained with H&E for analysis. Ethical approval was obtained by the NHS Health Research Authority West Midlands (Ref: 18/WM/0335) and experiments were conducted in compliance with the Declaration of Helsinki.

In total, 509 slides were collected from 406 patients. The slides were digitised to high-resolution WSI at 40× objective power using one of three scanners: NanoZoomer S360 (Hamamatsu Photonics, Japan; 0.2258 mpp), Aperio CS2 (Leica Biosystems, Germany; 0.2520 mpp), Pannoramic 1000 (P1000, 3DHISTECH Ltd, Hungary; 0.2426 mpp). Further inspection of the WSIs excluded cases with poor staining quality, artefacts, distortions or blurring. The resulting cohort comprised 358 WSIs (n=277 patients) with a confirmed histological diagnosis of OED and 105 WSIs (n=81 patients) confirmed as non-dysplastic (controls). Due to incomplete follow-up data for five patients with OED (7 WSIs), these cases were only used for training and excluded from clinical outcome analysis. Thus, the final cohort included 351 WSIs (n=272 patients) amongst which 64 patients (79 WSIs) exhibited malignant transformation. Slides from the same patients were assigned to the same fold during algorithm training/testing. An overview of the dataset is provided in Table 14, and a CONSORT diagram given in Figure 29.

Clinical follow-up data for the OED cohort included patient age (at time of diagnosis), sex, intraoral site, OED grade (using binary and WHO 2017 systems) and transformation status. Transformation was defined as the progression of a dysplastic lesion to OSCC at the same clinical site within the follow-up period, and transformation time was measured in months. To ensure diagnostic consistency, all cases were initially evaluated for diagnosis by a certified pathologist (PMS, PMF, DJB, KH), who provided an initial diagnosis based on the WHO grading system (between 2008-2016). To confirm the WHO (2017) grade and assign binary grades, the cases were blindly re-evaluated by SAK and a clinician with a specialist interest and expertise in OED analysis (HM).

HM exhaustively delineated regions of interest (ROI) representative of dysplasia in the 260 OED WSIs, in addition to the entire epithelium in the 96 control WSIs, using in-built annotation tools in the QuPath® software[134].

Independent Validation Cohorts

The ODYN model was tested on three external datasets acquired from:

i. Precision Medicine Centre, Patrick G. Johnston Centre for Cancer Research, Queen's University Belfast, UK (47 WSIs)
ii. Institute of Head and Neck Studies and Education, Institute of Cancer and Genomic Sciences, University of Birmingham, UK (42 WSIs)
iii. Oral Diagnosis Department, Semiology and Oral Pathology Areas, Piracicaba Dental School University of Campinas (UNICAMP), São Paulo, Brazil (19 WSIs)

Owing to the limited size of these datasets we combined them into a single multi-institutional external test set. Prior to the inclusion of external cases in the study, all WSIs were checked for suitability. Slides of poor quality, insufficient epithelium and cases positive for Candida Albicans or suggestive of Human Papilloma Virus infection were excluded. The WSI cohorts from Birmingham and Belfast were scanned at 40× objective power using a Panoramic 250 (P250, 3DHISTECH Ltd., Hungary; 0.1394 mpp) and Aperio AT2 (Leica Biosystems, Germany; 0.2529 mpp) whole-slide scanner, respectively, to obtain digital WSIs. The Brazil cases were scanned at 20× objective power, by an Aperio CS (Leica Biosystems, Germany; 0.4928 mpp) scanner. The same clinical follow-up information was collected as that for the development/internal cohort. Amongst the OED cases, 44 lesions transformed to malignancy. The external dataset did not include any control cases. A summary of this cohort is provided in Table 14, and a CONSORT diagram in Figure 29. HM exhaustively delineated ROIs of dysplasia in 30 cases each from both Birmingham and Belfast, and an additional 18 cases from Brazil, using the QuPath® software.

Table 14. *Overview of OED samples included in this study.*

	Sheffield	Birmingham	Belfast	Brazil
OED Cases, *n*	277	47	42	19
OED Slides, *n*	358	47	42	19
Median Age[a] (IQR)	63 (54 – 72)	61 (50 – 70)	62 (52 – 71)	53 (45 – 58)
Sex, *n* (%)				
Female	167 (48)	24 (51)	21 (50)	13 (68)
Male	183 (52)	23 (49)	21 (50)	6 (32)
Site, *n* (%)				
Buccal Mucosa	45 (13)	6 (12)	0 (0)	7 (37)
Tongue	158 (45)	30 (64)	29 (69)	8 (42)
Floor of Mouth	68 (19)	3 (6)	9 (21)	0 (0)
Other	80 (23)	8 (17)	4 (10)	3 (16)
WHO grade, *n* (%)				
Mild	118 (33)	24 (51)	6 (14)	11 (58)
Moderate	134 (38)	18 (38)	25 (60)	4 (21)
Severe	99 (28)	5 (11)	11 (26)	4 (21)
Binary grade, *n* (%)				
Low-risk	218 (62)	28 (60)	7 (17)	11 (58)
High-risk	133 (38)	19 (40)	25 (83)	8 (42)
Transformation[b], *n* (%)	79 (23)	10 (21)	30 (71)	4 (25)
Scanner, *n* (%)				
Aperio CS2	180 (50)	0 (0)	0 (0)	0 (0)
NanoZoomer S360	98 (27)	0 (0)	0 (0)	0 (0)
P1000	80 (22)	0 (0)	0 (0)	0 (0)
Aperio CS	0 (0)	0 (0)	0 (0)	19 (100)
P250	0 (0)	47 (100)	0 (0)	0 (0)
Aperio AT2	0 (0)	0 (0)	42 (100)	0 (0)

[a] *Median age at OED diagnosis*

[b] *5 Sheffield cases (7 slides) and 3 Brazil cases (3 slides) had no follow-up data and are excluded from the transformation figures.*

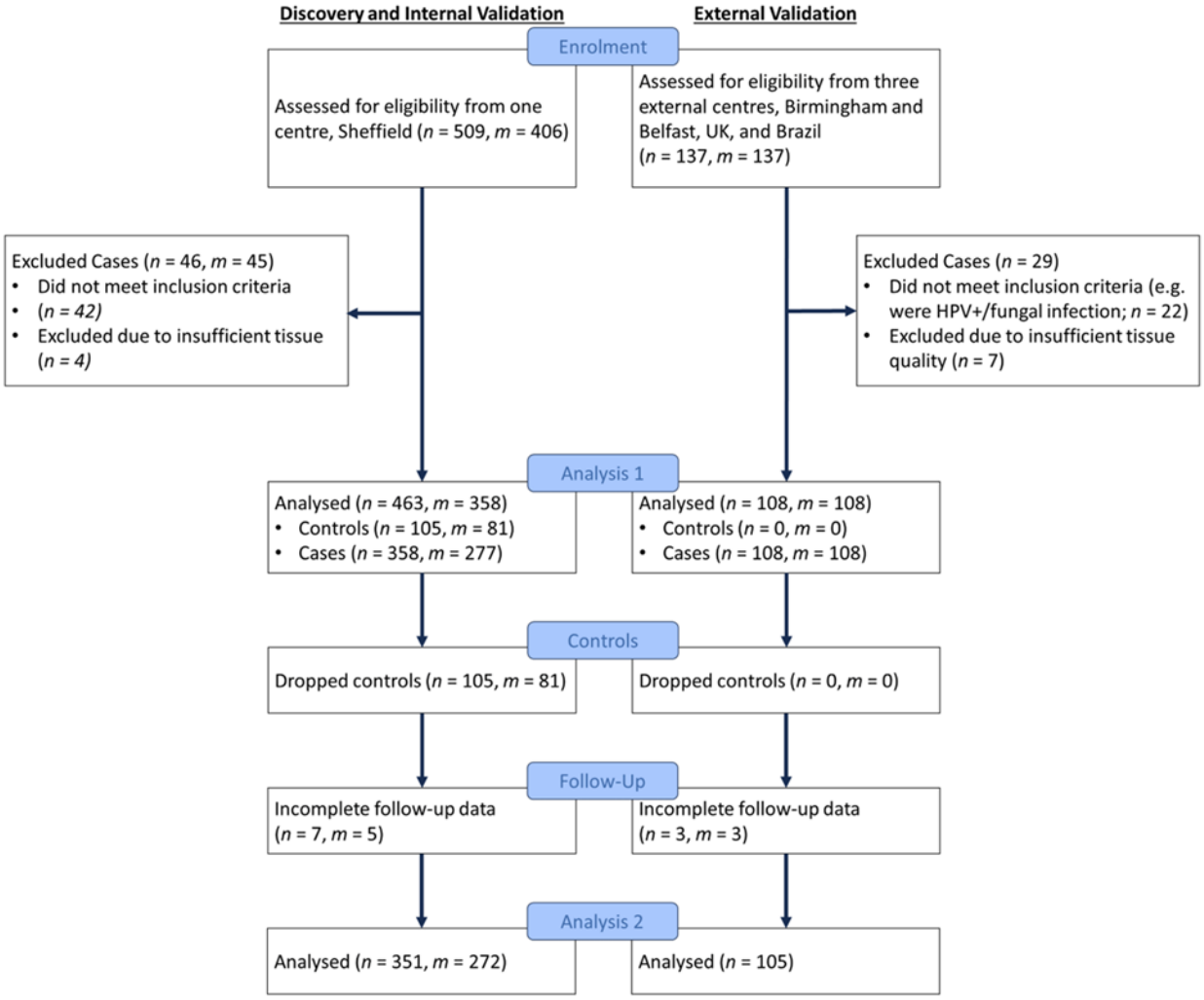

Figure 29. *CONSORT flowchart illustrating samples.*

8.1.2.3 *Deep Learning Pipeline*

Dysplasia segmentation

Since dysplastic changes may not be widespread across the entire tissue section in a slide, the first step of developing the DL pipeline involved identification and localisation of the dysplastic tissue regions for semantic segmentation. To achieve this, a fine-tuned Transformer, based on Trans-UNet[152] was used to automatically detect and segment the different dysplastic regions in each WSI across the training dataset. The model takes an input image of size 512 × 512 (at 1.0 micron per pixel, mpp, resolution) and outputs a dysplasia segmentation map.

For internal model testing, the dataset was split at 80/20, and controlled for both scanner type and OED grade. This resulted in 206 OED and 75 control WSIs in the training set, and 54 OED and 21 controls WSIs in the internal testing set, with ground truth annotations. Note, a higher proportion of controls were kept in the test set to ensure high specificity of OED segmentation in the control sample. After tessellating the WSIs and region masks into smaller patches (512 x 512 pixels, 184 pixels overlap, 10X magnification, 1.0 mpp), a total of 19,063 OED and 11,756 non-dysplastic patches were generated for model training/validation on the internal discovery cohort. This totalled 6,341 patches with ground truth annotations from the 78 WSIs in the external cohort. Model testing has been reported at the ROI level. Various stain augmentation algorithms were tested during the development of the final model, using the TIAToolbox[158].

OED classification

A pretrained CNN-based HoVer-Net+[137] model was used to segment the epithelium and the individual nuclei across each WSI. To classify OED, the proportion of the epithelium mask that was segmented as dysplastic was calculated and an empirically determined threshold used to classify slides as being dysplastic vs. non-dysplastic (dysplasia-epithelium ratio, R_{Epith}). We found thresholds for this ratio based on all the WSIs used for training the dysplasia segmentation model (281 WSIs). We therefore tested the model internally on the remaining 182 WSIs, and externally on all 108 WSIs.

Malignant transformation prediction (ODYN-scoring)

The WSIs were tessellated into smaller patches (512 x 512 pixels, 256 pixel overlap, 0.5 mpp) using tissue in the dysplastic regions alone. The nuclear segmentations from HoVer-Net+ were used to generate a total of 168 nuclear-based morphological and spatial features for each (dysplastic) patch. These patch-level features were used as input to an MLP to calculate a risk-score for malignant transformation (ODYN-score). The MLP model had three layers with 168 nodes in the input layer, 64 nodes in the hidden layer, and 2 nodes in the output layer. It used a leaky ReLU activation function and dropout (0·2) after the hidden layer. The MLP was trained by Monte Carlo iterative-draw-and-rank sampling IDARS[159], using a symmetric cross-entropy loss function and the Adam optimiser.

This loss function was chosen as it has been shown previously to help overcome errors associated with weak labels[159,160]. IDaRS sampling was performed with parameter values of k = 5 for top predictive patches and r = 45 random patches, using a batch size of 256. On inference, the trained MLP calculated a prediction score for each patch in the dysplastic regions of the WSI, which can be considered the likelihood of a tile belonging to the positive class in the classification task (i.e. transformation). Slide-level scores were then obtained by taking the average prediction score across the top 50% ranked tiles. We used interpretable nuclear features with the aim of making the model interpretable. However, we additionally provided comparison to a ResNet classifier (trained with

Macenko stain augmentation), using deep features, to show the impact on performance. We additionally include the OMTscore[137] as part of our ODYN model. The OMTscore uses the epithelium segmentation in place of the dysplasia segmentation for generating features patches. An overview of the ODYN-scoring pipeline is presented in Figure 30.

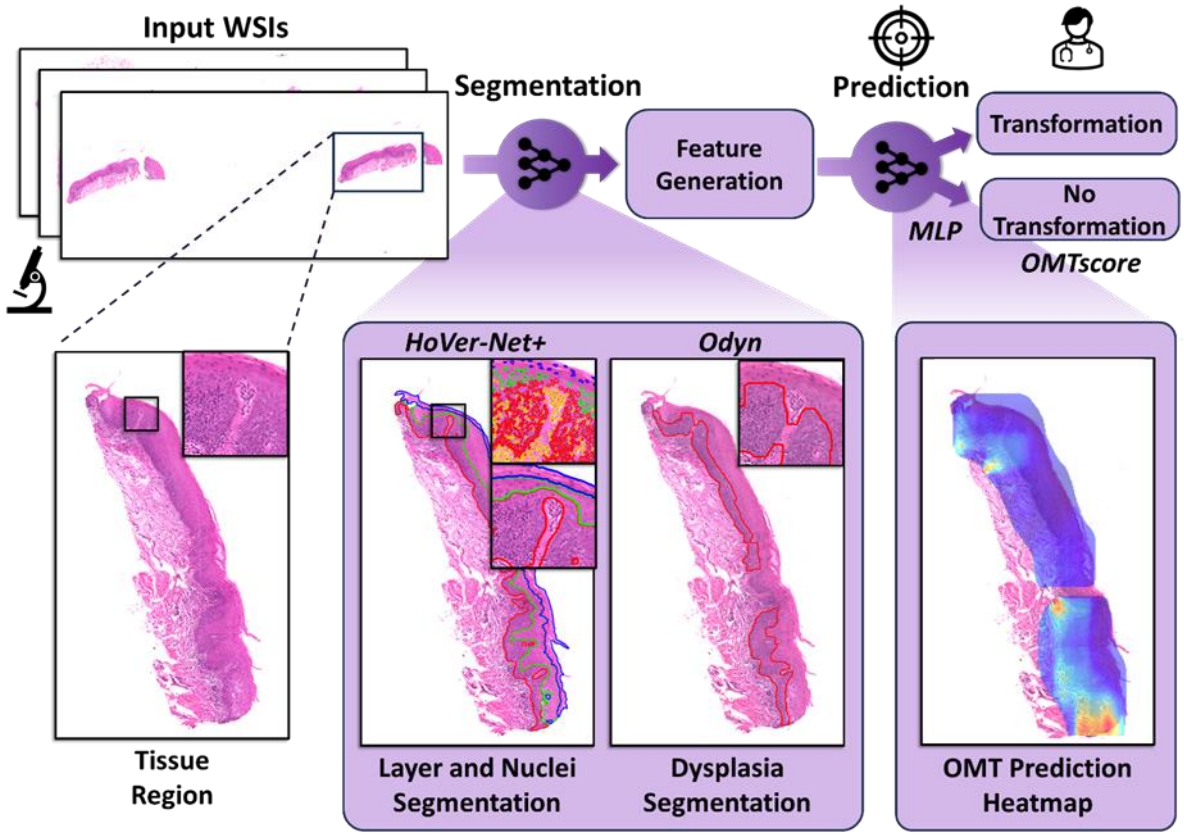

Figure 30. *Overview of ODYN-scoring pipeline*

An input WSI is used with HoVer-Net+ to generate an epithelial mask and nuclear segmentations. ODYN is then used to generate the dysplasia segmentation. These patches are then used within the OMTscoring pipeline to predict malignant transformation. This includes the WSIs being tessellated into smaller patches and morphological/spatial features being generated for each patch within dysplastic regions. These patch-level features are then used within a MLP to predict whether the lesion transformed to malignancy.

8.1.2.4 *Statistical analysis*

For the evaluation of OED segmentation, large ROIs centred on the annotated tissue section were generated, and the remaining tissue was regarded redundant and excluded. Both internal and external testing was performed on these large ROIs, which in the majority of cases, encapsulated all tissue on the WSI. Dysplasia segmentation performance (aggregated across all ROIs) was measured by calculating the F1-score, recall and precision. For internal testing of controls, a single measure of specificity for OED segmentation was reported, since a single incorrectly predicted pixel (e.g. incorrectly predicted as OED), would result in an F1-score, Recall, and Precision values of 0; thus not giving an accurate representation of the model performance. For the evaluation of OED classification (dysplastic vs non-dysplastic) the F1-score, Recall, and Precision across all ROIs (and slides) were measured. Area under the receiving operating characteristic (AUROC) score was calculated for internal testing across all ROIs.

We used five-fold cross-validation in our ODYN-scoring internal experiments based on the internal cohort. We then tested our model externally, by training across the entire internal cohort and validation on the external cohort. For the evaluation of the ODYN-scoring pipeline, we provide an AUROC score across all slides. Survival analyses were additionally conducted to assess the prognostic significance of the ODYN-score in predicting transformation-free survival. The ODYN-score indicated whether the algorithm predicted the case to have transformed (high-risk) or not transformed (low-risk). Kaplan-Meier curves were generated, and long-rank tests were used to determine the statistical significance of grading (for ODYN-score, WHO and binary grades). A multivariate Cox proportional hazards model was employed, incorporating sex, age, lesion site, binary and WHO grade, to predict transformation-free survival. Transformations were right censored at eight years across these analyses to ensure consistency between internal and external cohorts.

8.1.2.5 *Role of the funding source*

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or in the submission or writing of the report.

8.1.3 Results

8.1.3.1 *Dysplasia segmentation*

Internal testing of the ODYN dysplasia segmentation model demonstrated an F1-score of 0.807 (Recall = 0.845, Precision = 0.773) on OED cases and a specificity of 0.998 on controls (Table 15). On external testing, the ODYN model achieved a F1-score of 0.708 (Recall = 0.764, Precision = 0.660) (Table 16). The results of the ODYN model were superior (F1 = 0.81) to that of other state-of-the-art methods including U-Net[146], HoVer-Net+[137], DeepLabV3+[161], Efficient-UNet[162], and Swin-UNet[155] with good generalisability. Examples of dysplasia segmentation heatmaps are shown in Figure 31.

Table 15. *Internal testing of ODYN model*

	OED cases			Controls
Model	F1-score	Recall	Precision	Specificity
U-Net	0.775	0.796	0.755	0.996
HoVer-Net+	0.789	0.827	0.754	0.996
DeepLabV3+	0.802	0.817	**0.788**	0.998
Efficient-UNet	0.790	0.834	0.751	0.998
Swin-UNet	0.795	0.845	0.750	0.997
ODYN	**0.807**	**0.845**	0.773	**0.998**

Table 16*. External testing of ODYN model.*

	Combined (n = 78)		
Model	F1-score	Recall	Precision
U-Net	0.685	0.694	0.676
HoVer-Net+	0.668	0.719	0.623
DeepLabV3+	0.704	0.704	**0.705**
Efficient-UNet	0.700	**0.777**	0.638
Swin-UNet	0.680	0.728	0.638
ODYN	**0.708**	0.764	0.660

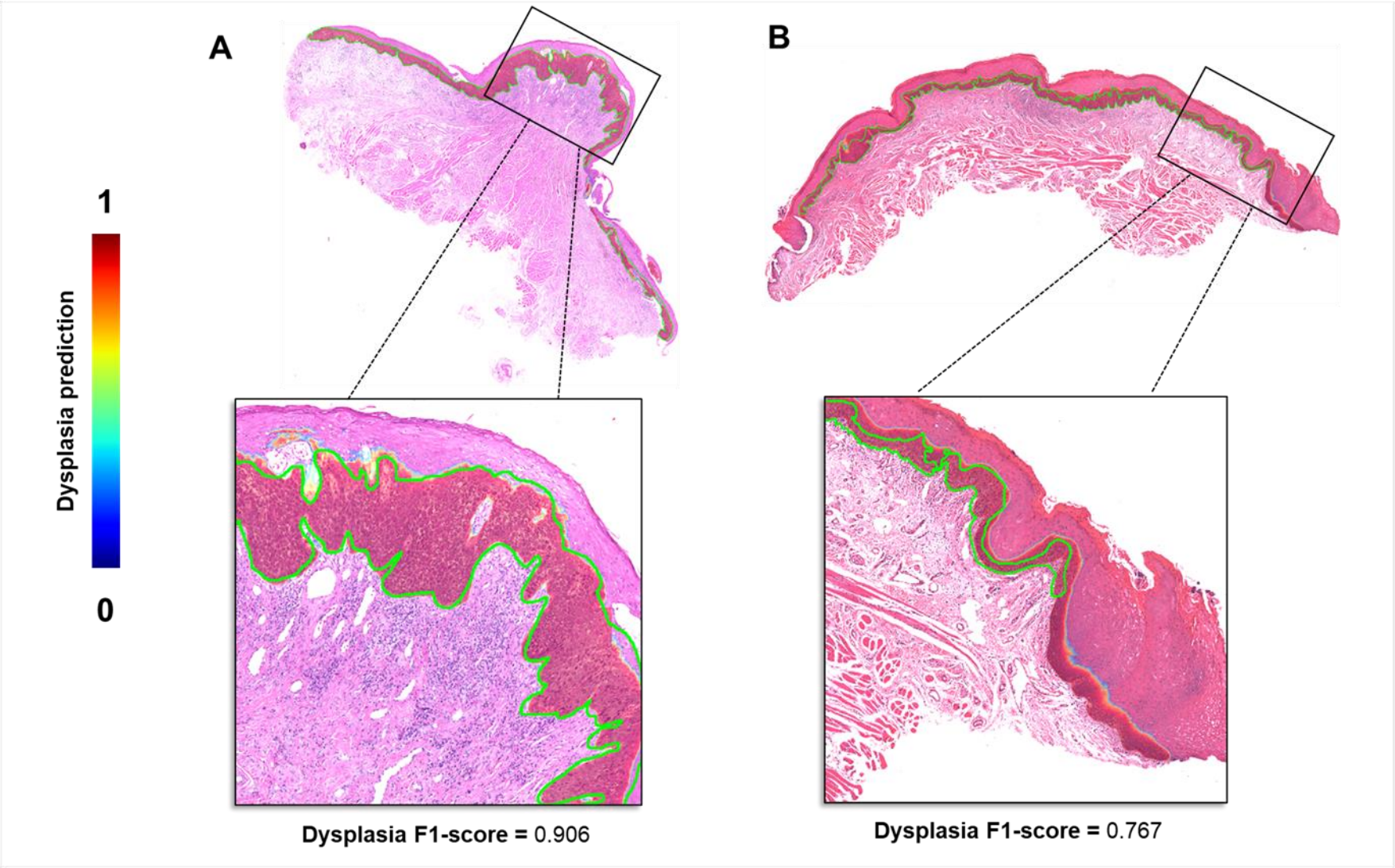

Figure 31. *Dysplasia segmentation heatmap using ODYN model.*

A) Severe OED (binary grade: high-risk) which transformed; B) Mild OED (binary grade: low risk) which did not transform. The green line depicts the ground truth dysplasia segmentation.

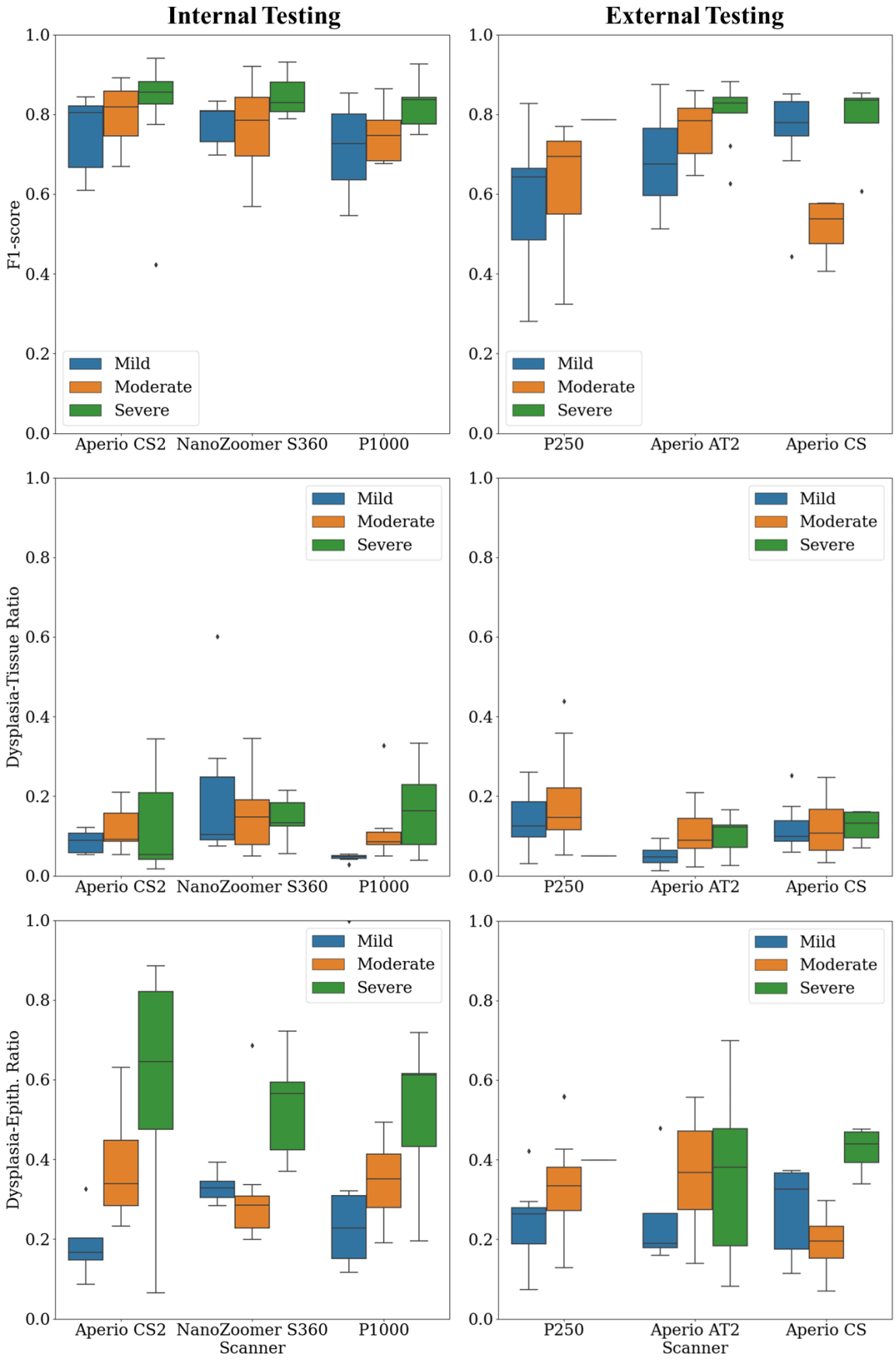

Figure 32. *ODYN model performance by scanner and histological grade.*

8.1.3.2 *OED classification*

We calculated the classification accuracy of ODYN, given a H&E-stained oral tissue WSI, based on the proportion of the epithelial mask that was dysplastic. On internal testing, we achieved an F1-score of 0.98 (AUROC = 0.96, Recall = 0.97, Precision = 0.97). The performance remained high on external testing, gaining an F1-score = 0.96 (Recall = 0.93, Precision = 1.00), showing the robustness and generalisability of the proposed method. ROI-level classification performance is shown in Table 17, and WSI level classification in Table 18.

Table 17. *ROI-level classification of OED based on dysplasia scores.*

**AUROC values not determined for external testing as this dataset did not include control*

	Internal Validation (*n* = 89 ROIs)				External Validation (*n* = 87 ROIs)			
Score	F1-score	AUROC	Recall	Precision	F1-score	AUROC*	Recall	Precision
Dysplasia-Tissue	0.96	0.95	**0.98**	0.93	**0.98**	-	**0.96**	**1.00**
Dysplasia-Epithelium	0.98	0.99	0.95	**1.00**	0.94	-	0.89	**1.00**
Dysplasia-Layer:								
Basal	**0.98**	**1.00**	0.97	**1.00**	0.91	-	0.84	**1.00**
Epithelial	0.91	0.97	0.83	**1.00**	0.94	-	0.89	**1.00**
Keratin	0.74	0.78	0.59	0.98	0.89	-	0.80	**1.00**

Table 18. *WSI-level classification of OED, based on dysplasia scores.*

	Internal Validation (*n* = 463 WSIs)				External Validation (*n* = 108 WSIs)			
Score	F1-score	AUROC	Recall	Precision	F1-score	AUROC	Recall	Precision
Dysplasia-Tissue	0.97	0.96	0.95	0.98	0.96	**-**	0.93	1.00
Dysplasia-Epithelium	**0.98**	0.96	**0.97**	**0.97**	0.96	**-**	0.93	1.00
Dysplasia-Layer:								
Basal	0.97	**0.98**	0.96	0.96	0.96	**-**	0.93	1.00
Epithelial	0.97	0.96	**0.97**	**0.97**	**0.97**	**-**	**0.94**	1.00
Keratin	0.75	0.78	0.62	0.62	0.86	**-**	0.76	1.00

8.1.3.3 *Malignant transformation prediction*

For the prediction of malignant transformation (our ODYN-score), we performed both internal cross-validation on the development cohort and external validation. On internal cross-validation we attained an AUROC of 0.74, which increased to 0.75 on external validation (Table 19). These scores are significantly higher than existing clinical grading systems including WHO (2017) and binary (Figure 33). We additionally compared them to other recently published methods such as IDaRS[159] (trained with stain augmentation, over dysplastic patches) and the OMTscore[137].

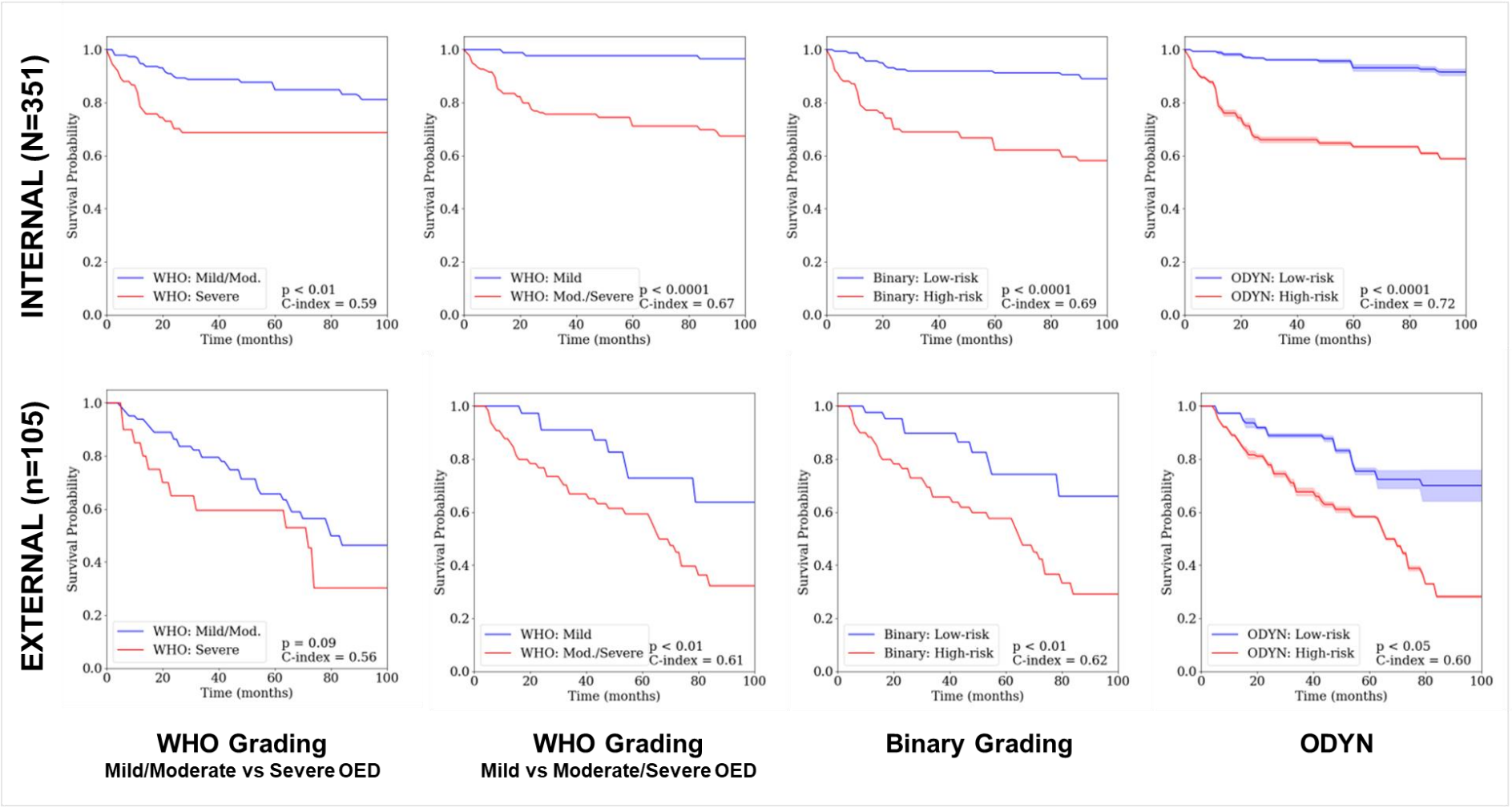

Figure 33. *Kaplan Meier curves for ODYN compared to other grading system*

Table 19. *Slide-level results for transformation prediction.*

	Internal Validation		External Validation	
Model	AUROC	C-Index	AUROC	C-Index
WHO Grade 1	0.61 (0.06)	0.60	0.56	0.56
WHO Grade 2	0.67 (0.03)	0.67	0.65	0.61
Binary Grade	0.73 (0.05)	0.69	0.68	0.62
ODYN: ResNet34	0.75 (0.09)	0.73	0.71 (0.01)	0.59
ODYN: OMTscore	0.73 (0.09)	0.70	0.71 (0.02)	0.58
ODYN-score	0.74 (0.07)	0.72	0.75 (0.01)	0.60

8.1.4 Discussion

Several studies have explored the application of machine learning, including DL, to study OED. The general focus of these methods has been to segment the epithelium (and the nuclei), either manually or via DL models[137,163,164]. These segmentations have then been used in further DL models to predict grade or transformation[10,137,163] or for pathologist-curated features based on digital images[9]. However, there has been little focus on segmenting dysplastic regions only for downstream prediction of malignant transformation.

In this study, we introduce ODYN, a novel transformer-based pipeline built on the Trans-UNet architecture[152] for OED segmentation, classification and malignant transformation prediction. This pipeline has been developed using the largest and most diverse multicentric OED dataset to date, digitised using six different scanners. The results obtained through rigorous testing and validation demonstrate the effectiveness of our models in the diagnosis and prognostication of OED. The ODYN dysplasia segmentation performance has consistently outperformed other state-of-the-art DL models, as evidenced in both internal and external testing. Furthermore, the ODYN model has demonstrated good generalisability across external unseen datasets, indicating its robustness for diverse clinical setting. This highlights the potential of transformer-based architectures in accurately delineating regions of dysplasia in H&E stained WSIs of OED tissue. This novel ground-breaking approach has the potential to redefine the landscape of OED diagnosis by providing more precise and consistent results that may also help reform treatment guidelines and standardise patient care.

ODYN has also demonstrated promising results for OED classification. In this study, we used the predicted dysplastic proportion of the epithelium in a WSI to determine a diagnosis of OED. We chose this method to classify a WSI as dysplastic, rather than classifying a WSI as dysplastic solely based on the presence of any predicted dysplasia. We made this choice because our model predictions often included small areas of false positives. This decision to define a threshold, proved to be successful on both internal and external testing. The high precision and recall achieved in classifying OED indicates the potential for automated diagnosis, which has the potential to increase diagnostic efficiency.

The application of ODYN-produced segmentation maps in predicting malignant transformation represents a significant advancement in computational pathology. Notably, this approach outperforms the OMTscoring pipeline proposed by Shephard *et al.* (2023)[165] with a substantial improvement in AUROC score.

We believe that the application of cutting-edge DL techniques, such as the ODYN pipeline, has huge translational potential which could help improve the accuracy and objectivity of OED diagnosis and grading. In addition to this, AI-based pipelines can improve prognostic reliability for prediction of cancer risk to improve patient outcomes. Future research should explore the scalability of the ODYN

model to accommodate a broader range of oral conditions (such as those which can mimic OED) and tissue variations to assess whether it can accurately discriminate OED from other similar appearing conditions whilst still accurately predicting malignancy risk. This will enhance the clinical utility of the model and ultimately help provide more personalised patient care.

The authors acknowledge some challenges and opportunities for future research based on this study. A potential challenge highlighted by this work is the need to address the interpretability of DL models in clinical practice. We have therefore used an interpretable model for transformation prediction, that takes into account known histological features (e.g. shape and size variations of nuclei) to generate predictions from dysplastic ROIs. We provide heatmaps for each slide to help explain model decisions. We believe such approaches can enhance trust and acceptance amongst healthcare professionals. The authors additionally acknowledge limitations associated with a retrospective study. It would have been of interest to further explore the model performance for predicting OED recurrence. However, as there is no standardised treatment protocol for OED, there may have been variations in patient management between centres, and it can be difficult to reliably distinguish between a lesion that truly recurred from a lesion that has developed at the same clinical site due to field change. We would have additionally liked to incorporate social risk factors (e.g. smoking, alcohol consumption) in the multivariable modelling, however, it was not possible to acquire consistent information between the different centres retrospectively. These issues could be addressed by a future prospective validation study. Despite this, the external validation of our models across multiple centres and scanners is a notable strength of this study. Future research could explore the application of ODYN in even more diverse clinical settings and expand its utility to other histopathological tasks beyond OED analysis. We suggest testing the method on other head and neck precancerous lesions, such as laryngeal dysplasia, as an interesting future direction of research.

8.1.5 Conclusion

In conclusion, our study signifies a substantial leap forward in the field of digital oral and head and neck pathology, offering a powerful tool in ODYN for the detection, segmentation, and classification of OED, which we have made publicly available. This technology, underpinned by DL and transformer-based architectures, showcases the potential of computational pathology to revolutionise the diagnosis and management of OED. The model's exceptional performance in both internal and external testing, along with its ability to improve transformation prediction, underscores its potential to impact clinical practice positively. By addressing challenges and continuing to refine the model, we envision ODYN playing an important role in improving the diagnosis and management of OED and potentially other head and neck precancerous lesions in the future.

Acknowledgements

This study was supported by a Cancer Research UK Early Detection Project Grant, as part of the ANTICIPATE study (grant no. C63489/A29674) in addition to funding from the National Institute for Health Research (award no. NIHR300904). The authors express their gratitude to Professor Paul Speight (PMS), Professor Paula Farthing (PMF), Dr Daniel Brierley (DJB), and Professor Keith Hunter (KDH) for their valuable contribution in providing the initial histological diagnosis.

Disclosure/Conflict of Interest

NMR is the co-founder, Director and a shareholder of Histofy Ltd. SAK is a shareholder of Histofy Ltd. All other authors have no competing interests to declare.

Ethics Approval and Consent to Participate

Ethical approval for the study was obtained from the NHS Health Research Authority West Midlands (18/WM/0335).

Chapter 9 – General Thesis Discussion

9.1 Discussion

This thesis has presented a comprehensive body of work comprising two literature reviews and five studies. Each of the studies has used a different digital approach to inform the next component, culminating in the development of prognostic models and identification of novel digital markers. As the key findings, points of discussion, study limitations, challenges and areas for future research have been discussed at the end of each study, this chapter will provide a general summary of the entire work and the salient messages arising from it.

9.1.1 General thesis summary

Prior to the commencement of the research studies, it was imperative to evaluate the existing body of literature to gain an insight into the application of AI/ML/DL for detection, diagnosis and grading of head and neck precancerous and cancerous lesions. Two comprehensive literature reviews were undertaken to examine the published literature between 2009-2020.

The first, a scoping review, which highlighted a wide breadth of computational methods and imaging modalities (i.e. multimodal optical, micro-endoscopic, hyperspectral, infrared thermal, clinical) used for algorithm training in the last decade, amongst which histology WSI and radiological imaging were the most frequently used modalities for detection of OPMD and cancerous oral lesions. This finding was consistent with the increasing ubiquity of digital slide scanners in pathology departments, and the emergence of radiomics that has broadened the scope of medical imaging in clinical oncology. Due to the heterogeneity of the datasets included in the review, a detailed statistical analysis could not be undertaken, though most studies demonstrated high degrees of accuracy and precision of the presented ML models which were shown to exceed the abilities of standard statistical techniques and human judgement in making predictions about data.

Since the scoping review provided a general overview of AI application with no restrictions on the type of imaging modalities used, the systematic review aimed to focus on the use of histology WSI only. This review additionally conducted a formal risk of bias assessment using a tailored QUADAS-2 tool to evaluate the quality of evidence supporting the use of AI. Most studies evaluated histological features in OPMD and head and neck cancers, with only one study focussing on the detection of OED specifically. A combination of heuristics, supervised and unsupervised learning methods were employed, including more than 10 different classification and segmentation techniques. Overall, the quality of evidence was low, with most studies showing a high risk of bias, likely to have overestimated accuracy rates. Most studies used uni-centric datasets (range 40-270 images) comprising small sub-images within WSI with accuracy between 79-100%.

Since most studies included in the reviews were published prior to 2015, they largely employed classical ML methods, despite there being significant advancements in the medical image analysis field. The evolution of the computational pathology has since seen a rise in the use of DL algorithms

for tackling complex computer vision tasks; with many state-of-the-art methods already being applied for early detection of breast, prostate, lung and skin malignancies. However, these reviews highlight that these latest methods have not been explored to the same extent in head and neck cancers and pre-cancers. The findings from these review studies helped inform the design of the subsequent set of studies presented in this thesis, and particularly the work described in Chapters 8 which develops a novel DL pipeline for automated detection, diagnosis, and transformation prediction in OED, for the first time, using a Transformer model.

Analysis of individual histological features

The next stage of work aimed to evaluate individual histological features in WSIs of OED, including nine established features listed in the WHO (2017) criteria, in addition to other commonly observed features, such as 'abrupt orthokeratosis', 'lymphocytic band' and 'verrucous surface architecture'. Basal cell hyperplasia (crowding) and irregular surface keratin were shown to be the most prevalent features in OED, despite them not being formally part of the WHO criteria. Interestingly, these features were not strongly linked to clinical outcomes and therefore did not form part of the proposed predictive models. Six other histological features were significantly associated with an increased risk of malignant transformation ($p<0.036$) and OED recurrence ($p<0.015$) including 'bulbous rete pegs', 'hyperchromatism', 'loss of epithelial cohesion', 'loss of stratification', 'suprabasal mitoses' and 'nuclear pleomorphism'. These features led to the development of two prognostic scoring systems including a 'six-point model' using all six of the significant features and a simplified 'two-point model' in which a single point would be given to each of the features which had the highest inter-pathologist agreement amongst the six features ('loss of epithelial cohesion' and 'bulbous rete pegs'). Both the 'six-point' and 'two-point' models showed good predictive ability (AUROC > 0.774 for transformation and 0.726 for recurrence) exceeding conventional grading systems. Validity testing of the 'six-point' and 'two-point' models was conducted to assess model performance on unseen independent datasets including OED cases from three national UK centres and a centre in Brazil. Findings supported the prognostic strength of both models, particularly the 'six-point' model which maintained its predictive ability on the validation cohorts (AUROC 0.81 compared to AUROC 0.68-73 for other models). Further improvements were seen with the addition of patient age, sex and histological grade. Future modifications of these models may consider giving different weights to the individual features, based on our findings from univariate feature analysis, which showed that different features carry different levels of association with malignant transformation. Validation on a larger scale is needed to assess the incremental benefit of doing so, since re-fitting the model to allow the different weights made a small difference in our validation study (AUROC of 0.82, compared to the 0.81 when equal weighting was given to the six features).

Following on from this study which analysed 'conventional' OED features, the next stage of work aimed to evaluate a range of other features seen in OED, which have not been previously analysed,

quantified, or correlated to clinical outcomes before. This study conducted a digital quantitative analysis of epithelial cellularity and thickness, nuclei geometry and staining intensity of cell cytoplasm. The prognostic relationships of these features were explored, and multivariable digital models developed and validated for prediction of malignant transformation and OED recurrence. Key findings demonstrated that epithelial cellularity increased with grade severity. In contrast, basal epithelial cellularity was lower in OED compared to control. Increased cellularity is a common feature observed in many types of cancers (such as gliomas[166], meningiomas[167], lymphomas[168], hepatocellular carcinomas[169]) and forms an important component of some diagnostic criterion as it can indicate tumour growth, progression and metastatic potential. Despite increased cellularity being a common feature in OED, it is not routinely quantified for diagnosis, mainly due to the difficulty in doing so using conventional microscopy. However, understanding cell number and density in OED may be important in risk stratification and also the surveillance of lesions, particularly to help inform treatment decisions and the need for re-biopsies. A possible explanation for the reduced basal cellularity in OED, is that it may correspond to the increasing level of cellular pleomorphism (and larger size) seen in OED, which in turn, may have resulted in fewer cells being detected by the software algorithm. This may also explain the finding of a reduced triangle area and distance between cells in OED as grade worsens, as noticed in the quantitative spatial analysis described in Chapter 6.

Another relevant finding of this work was that OED epithelial thickness was associated with an increased risk of malignant transformation ($p=0.03$). Whilst there was no significant difference with regards to epithelial thickness between OED grades, the keratin thickness/perimeter was statistically increased in OED compared to control ($p < 0.05$) (Figure 34). These results suggest that epithelial and surface thickness (or curvature/architecture) may be important histological features in risk stratification of OED lesions. Keratin is a structural protein that is a natural component of the epithelial lining of the oral mucosa. Hyperkeratosis can be observed in benign oral lesions in response to various stimuli, including chronic irritation or inflammation such as from tobacco use or chronic trauma but it can also be seen in leukoplakia that can result in the development of OPMD. Further analysis of keratin pattern and morphology in OED would be interesting to explore, particularly since verrucous surface morphology is commonly seen in some OPMD.

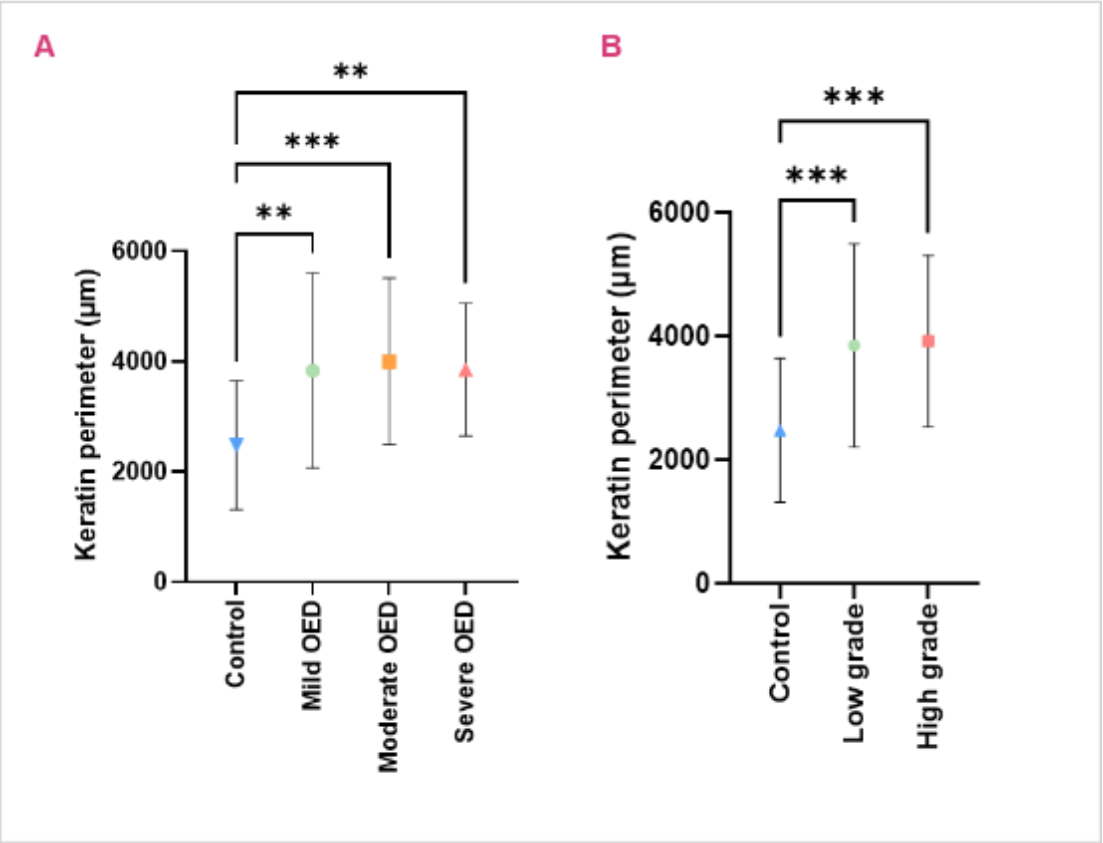

Figure 34. *Quantitative analysis of keratin thickness (perimeter) in OED.*

*Data expressed as mean values with 95% CI error bars. Asterisk highlights a statistical significance (*p ⩽ 0.05, **p ⩽ 0.01, ***p < 0.001).*

Analysis of nuclear and cytoplasmic features highlighted significant differences in 'cytoplasmic eosin', 'nuclear eccentricity' and 'circularity' in basal epithelial cells of OED ($p<0.05$). 'Nucleus circularity' was associated with OED recurrence ($p=0.018$). The altered eosin concentration in cytoplasm may be explained by the altered nuclear morphometry in dysplastic cells, which in turn, could affect eosin representation. The presence of dyskeratosis and premature/individual cell keratinisation may also contribute, giving the cytoplasm a more eosinophilic cytoplasm. The unique aspect of this study was the development of a multivariate model for outcome prediction using the digitally quantified histological data which produced a superior predictive model (Model 6: AUROC of 0.77 for malignant transformation and AUROC of 0.74 for recurrence) that outperformed conventional clinical grading systems (WHO grading for malignant transformation prediction AUROC 0.69). External validation supported the strength of the superior 'Model 6' (AUROC 0.76 for malignant transformation; AUROC 0.93 for OED recurrence).

Mitotic activity

The next study conducted a quantitative analysis of mitotic activity in OED (including number, type, and intraepithelial location of mitoses) using digitised H&E sections and IHC stained tissue with PHH3. Mitotic figure counting has been used for diagnosis and prognostication of various malignancies including brain[167], breast[170] and neuroendocrine carcinomas[171]. However, this was the first study to explore its prognostic role in OED. Key findings from this study demonstrated the total number of mitoses (TNOM) increased with OED grade severity both on H&E and PHH3-IHC analysis, which may be explained by the increased stem cell turnover and quantity of abnormal mutations as dysplasia progresses. The PHH3 mitotic count was greater than H&E, likely due to the inclusion of early prophase stage, which cannot be identified on H&E-stained sections. TNOM, mitosis type and intra-epithelial location were important prognostic features, with greater predictive strength when combined. Multivariable models incorporating multiple mitotic features demonstrated better predictive strength for prediction of malignant transformation (H&E model: AUROC 0.8113; PHH3-IHC model AUROC 0.7714) and OED recurrence (H&E model: AUROC 0.7895; PHH3-IHC model: AUROC 0.7783) than WHO grading alone (AUROC 0.65). In the case of PHH3-IHC models, the most superior models utilised fewer mitotic features compared to the H&E models indicating that PHH3-IHC may be important for prognostication of OED, complementing H&E analysis.

PHH3 was chosen as the single most reliable marker of mitotic activity. Alternative markers such as Ki-67 were considered, but not used in this study. This is because Ki67 is a general cell cycle marker of proliferation which limited sensitivity since the proliferative activity of cells may not directly correlate with the clinical behaviour of the lesion. Interpretation of Ki67 staining can also be subjective, leading to inter-observer variability, which might affect its reliability as a standalone diagnostic or prognostic tool. Furthermore, a number of studies comparing PHH3 to Ki67 in OSCC progression have shown conflicting results.

Multivariable predictive models

The findings from Chapters 3, 4 and 5 highlight the benefits of multivariable models for prognosis prediction of oral precancers. Firstly, these models have been developed by integrating multiple histological and digital features which have been generated through objective quantification, and secondly, the incorporated variables have been statistically analysed and individually correlated to clinical outcomes. Whilst some features included in the models have greater levels of prognostic strength than others, overall, the integration of multiple features has strengthened the predictive accuracy and value of the models. Another benefit of these models is that they are easy to adapt, and are conducive to ongoing research and development as they can be regularly updated and refined as new data and insights emerge to maintain their relevance and effectiveness in clinical practice. In the context of precancerous lesions, these models aid in more precise risk assessment and decision-

making, potentially leading to improved patient outcomes through early detection, appropriate monitoring, and timely intervention when necessary.

Application of ML to OED

The next two arms of work (Chapters 7 and 8) explored the application of AI algorithms for OED detection, diagnosis, and discovery of novel digital markers of transformation. First, traditional ML models were trained to segment and classify OED epithelium, immune and stromal cells demonstrating good accuracy (F1 scores: 0.78-0.89, at ROI level). The optimised classifiers were used to study nuclei morphometry in the detected sub-epithelial immune cells and stromal cells in OED. Findings demonstrated significant grade-based differences for 'nuclear eccentricity' in subepithelial immune cells ($p=0.0203$); none of the other nuclear features were shown to be significantly associated with clinical outcomes of interest. A potential limitation of this analysis is that it was performed at the ROI-level on a relatively small test sample, with a modest proportion of transformed cases. As such, the sample may have lacked sufficient power to draw out subtle or meaningful differences in the studied features, which lends to the need to further explore immune cell quantification and architecture on larger samples, ideally at the WSI-level.

The main reason for investigating immune infiltration in OED was because emerging evidence has indicated a role for the immune system in OPMD disease progression. A number of studies have shown that the presence and behaviour of certain types of immune cells may indicate a higher risk of dysplastic lesions progressing to OSCC. In one study, there was an objective increase in the infiltration of CD4, CD8, CD14, CD19+20, and HLA/DR positive cells in the epithelium of tongue lesions that progressed from hyperkeratosis to dysplasia and carcinoma, with the most pronounced increase in B lymphocytes, that was in accordance with the transformation level ($p<0.001$)[172]. In another study, moderate and severe OED exhibited a high lymphocyte infiltration and upregulation of genes involved in immune surveillance (major histocompatibility complex-I, T cells, B cells and cytolytic activity) and immune suppression (immune checkpoints, T regulatory cells, and tumour-associated macrophages). The study found three distinct subtypes of moderate and severe OED: immune cytotoxic, non-cytotoxic and non-immune reactive[173]. In a study by Bashir *et al.* (2023)[163] (for which the candidate was a co-author) a weakly supervised DL model was trained on OED cases to reveal certain nuclei features in the epithelium and per-epithelial lymphocytes (PELs) to be significant prognostic factors for prediction of malignant transformation, including the PELs count ($p<0.05$). Progression-free survival using the PELs count ($p < 0.05$, C-index = 0.73) showed a strong association with malignant progression of OED on univariate analysis. These findings highlight the need to better understand the underlying immune mechanisms in OED to aid discovery of specific immune cell profiles important in malignant transformation. This may help develop more targeted treatment and the role of immunotherapy in high-risk lesions, which would be particularly useful for patients not suitable for surgery.

In the same arm of work, a quantitative analysis of stromal cells in OED was also conducted. Findings demonstrated increased stromal cellularity the subepithelial layer in OED compared to control, which was also consistently seen when comparing individual grades to control ($p<0.0001$). Analysis of nuclear features in the stromal cells showed that 'nuclear haematoxylin OD' and 'nuclear to cell ratio' were also statistically different between OED and control ($p=0.0001$-0.0062). Whilst stromal cellularity was not shown to be a prognostically significant feature for malignant transformation prediction, several nuclear features pertaining to the stromal cells in OED were statistically associated with an increased risk of malignancy, including 'nuclear circularity' ($p=0.001$), 'nuclear eccentricity' ($p=0.003$), 'cytoplasm eosin OD' ($p=0.045$) and 'nuclear to cell area ratio' ($p=0.003$). With the exception of 'cytoplasm eosin OD', these same features were also associated with an increased risk of OED recurrence ($p<0.05$).

Classical ML methods, whilst powerful and versatile, are subject to some limitations. Whilst the *D-ND and S, I classifiers* demonstrated good reliability and cell classification accuracy on internal testing, model performance worsened somewhat on external testing. It is likely that differences in scanner types and slide staining would have contributed to this reduced performance, and attempts were made to address this by adding external cases to the training dataset with application of stain augmentation. However, this had little effect on the model performance overall. While classical ML models can perform well on training data, these methods can struggle to generalise to unseen images, especially when faced with images that differ significantly from the training dataset. Another limitation with classical ML models is that they can be highly sensitive to data quality. Due to variations in slide quality between the internal and external cases, background "noise" may have negatively impacted on model performance resulting in less accurate predictions. A significant increase in the volume and diversity of training data would be an obvious way to overcome such issues, and to reduce bias in the sample. However, during experimentation, the ML models were struggling to handle larger datasets, which required greater computational demand and memory, potentially reducing scalability and external interpretability of the models.

Application of DL to OED

Newer AI approaches and DL models can overcome many of the challenges associated with traditional ML methods and have been shown to handle complex data more effectively. The final arm of the work, described in Chapter 8, focussed on the application of DL methods, and developed a novel transformer-based pipeline (ODYN) that can both classify OED and assign a predictive score (ODYN-score) to quantify the risk of malignant transformation. Model testing yielded an F1-score of 0.71 for OED segmentation, 0.96 for classification of dysplastic vs non-dysplastic tissue and an AUROC of 0.75 for malignancy prediction, outperforming other state-of-the-art methods, and existing clinical grading systems. The ODYN model was developed using the largest and most diverse multicentric OED dataset to date, incorporating digitised slides from six different scanners. These

factors contributed to good generalisability across external unseen datasets, indicating model robustness and applicability in the clinical setting.

9.2 Limitations

Specific limitations related to the different components of work presented in this thesis have been discussed in the respective chapters. Here, some general limitations are acknowledged and discussed.

One of the main limitations across the studies relates to the use of retrospective data. However, in order to develop and test the statistical, digital and computational models, it was necessary to have clinical follow-up data to assess the prognostic abilities of the models and make comparisons with existing clinical grading systems. It would not have been practical to collect data prospectively to achieve these aims. Other limitations of retrospective data relate to the potential for incomplete and inconsistent information due to variations in the quality of clinical record keeping between clinicians and centres. Gaps in data meant that it was not possible to acquire comprehensive social history information (tobacco and alcohol use) which would have been interesting to evaluate and incorporate into the multivariable models, particularly as they are established risk factors for OED and OSCC development. It was also difficult to acquire accurate information regarding OED recurrence, partly due to the variations in treatment protocols between surgeons in different centres, and since it is difficult to reliably know which lesions truly recurred and which developed secondary to oral field change. Without controlling the clinical, genetic, molecular and histological variables that can influence the development of dysplasia, it would not be possible to establish the cause of an OED lesion arising at the same clinical site after surgical removal. Whilst OED recurrence was evaluated in some parts of the work, the sole focus in the later studies was to identify predictors of malignant transformation.

Another limitation relates to the clinical follow-up period, which was set at a minimum of five years. This was chosen on balance between availability and access of clinical data and reported mean malignant transformation rates. Whilst malignant transformation of OED can occur five years after diagnosis, the literature indicates the greatest risk of progression to occur within the first five years[40,46,174].

Another limitation relates to the varying cohort sizes between the studies. Typically, this would make it difficult to make direct comparisons between studies, but in this case, we do not feel it has affected the reliability or generalisability of the presented results since the datasets were developed in accordance with the different tasks, goals and approaches for each individual studies. For the AI-based studies (Chapters 7 and 8) a power calculation was performed with support from a statistician in the NIHR Research Design Service team. A sensitivity of 90% was set for the prediction of malignancy risk, resulting in the need for a sample of at least 330 WSI for training (95% CI 0.068) and 200 WSI for validation and testing (95% CI 0.088). The digital datasets used for developing the DL

pipeline exceeded these numbers and is the largest and most diverse OED dataset to date, bigger than samples used in some landmark studies, such as CAMELYON[110].

A potential limitation of the ML work (Chapter 6) is that the analysis was performed at the ROI-level, as appose to WSI-level. Whilst both are useful methods of examining tissue samples, the benefit of ROI-level analysis is that it provides a focussed analysis on a targeted region of interest which in the case of my work was the dysplastic epithelium and subepithelial stromal region in OED. This method requires less computational resources, memory and can save time. It is also particularly useful for studying OED of varying severity, particularly as dysplasia is not always present throughout the entire tissue section and there may be areas important to avoid – such as ulcerated or inflamed regions. However, this method may introduce some level of subjectivity especially if the ROI selection is not automated. To overcome the risk of bias, the candidate received extensive training in analysis and diagnosis of OED and all ROIs were verified by a Consultant Oral & Maxillofacial Pathologist or a Senior Registrar in Oral and Maxillofacial Pathology. The DL analysis (Chapter 7) using the ODYN pipeline was performed at WSI level for prediction of malignant transformation which means the entire tissue section was represented at high-resolution, capturing all histological details to inform the outcome. This method of analysis has the added benefit of uncovering hidden predictive histological patterns in the tissue, would otherwise be difficult to visualise, examine and objectively explore.

9.3 Ongoing & Future Research

9.3.1 Model validation

Ultimately, it is intended that the proposed prognostic model will be validated prospectively in the clinical setting. Prior to doing so, it would first be necessary to refine the developed models, and potentially combine variables of greatest prognostic weight to formulate a single multivariable 'super' model for transformation prediction. In doing so, redundant, or irrelevant variables, that do not contribute much to the model's overall predictive power may be removed. Feature selection for the final model will be achieved using statistical methods or AI techniques to identify the most clinically important features in OED progression. The resulting model would benefit from further testing, initially using multicentric retrospective OED datasets that are representative of the general population affected by this condition. Wider validation will be achieved through collaboration with clinical experts both nationally and internationally, to allow a better understanding of the clinical application, reliability, and generalisability of the model's predictions. For the AI-assisted models, the use of large datasets from different geographic regions will increase biological diversity and model training, further enhancing the quality of predictions.

Any future trials to validate the proposed prognostic model will involve critical stages to ensure patient safety, clinical efficacy, regulatory and ethical compliance. Support will be obtained from the NIHR Research Design Service and Sheffield Clinical Trials Research Unit to help develop future research

projects. Clinical experts, statisticians, regulatory specialists, PPI groups and other relevant stakeholders would be engaged with throughout the development and implementation of future trials. Health economists at the University will also be consulted to integrate a cost-benefit analysis of the proposed tool to ensure it is economically viable within the NHS. Broadly speaking, it is anticipated a future clinical trial will comprise three arms for patient recruitment: 1) *control group:* pathologist led OED analysis and transformation prediction using conventional grading systems; 2) *test group:* digital OED analysis and transformation prediction using the proposed model; 3) *augmented group:* digital/AI analysis with pathologist supported prediction. To undertake the above work, the candidate intends to apply for post-doctoral grants/fellowships such as an NIHR Advanced Fellowship, an Academy of Medical Sciences Starter Grant or an MRC Early Investigator Grant.

9.3.2 Multiplex imaging and spatial biomarker analysis

The work presented in this thesis has been focused purely on histopathological predictors in OED. This is because, despite advancements in omic research, there are still no proven biological or molecular markers predictive of OED progression to malignancy. Nonetheless, the role of biomarkers in OSCC development is important to explore, however most studies have been largely mechanistic and do not consider the spatial architecture of the tissue. Spatial biomarkers can help deepen our understanding of the tissue microenvironment by giving an insight into the spatial distribution and organisation of cells, structures and gene expressions that may be important in cancer development.

To explore the role of spatial markers in OED, spatial transcriptomic analysis was conducted on formalin-fixed paraffin-embedded tissue (FFPE) tissue samples of severe OED (n=1) and OSCC (n=1) using 10x Genomics® Visium technology. The laboratory processes involved in this work were conducted by Dr Amir Zaki Abdullah Zubir (Research Associate, University of Sheffield) whom the candidate shadowed to gain technical insight. Preliminary findings demonstrated more than 200 gene expressions in OED, including the upregulation of cytokeratin in epithelial cells and downregulation of TGFβ in stromal cells (Figure 35). These findings sparked interest to further explore top relevant candidate markers in OED.

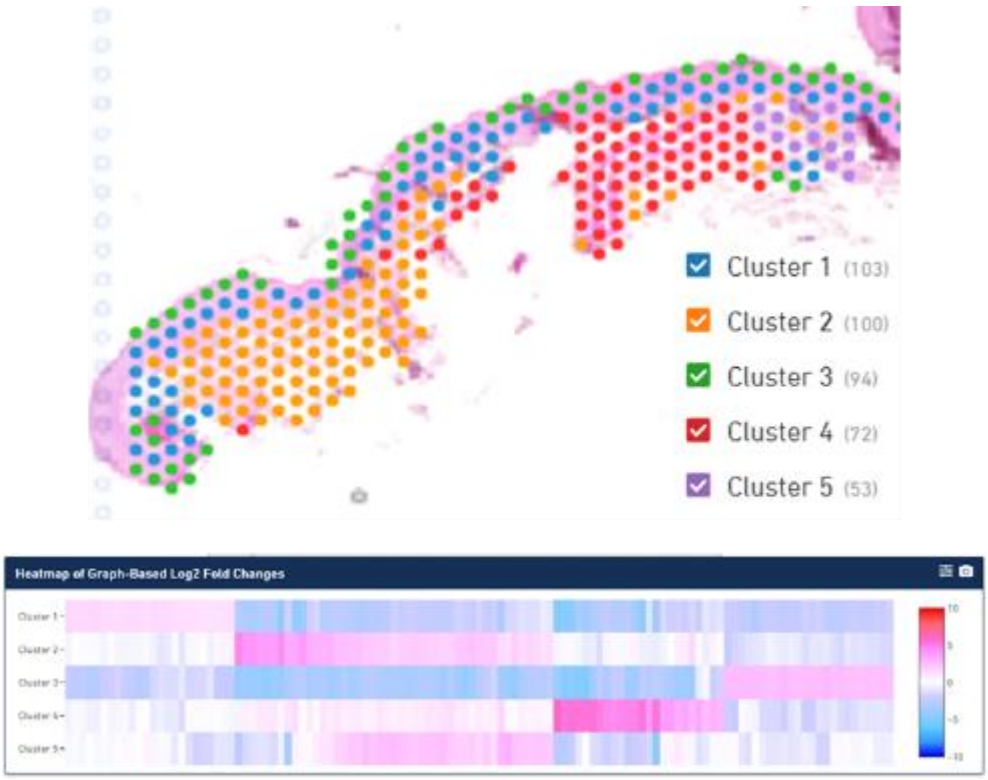

Figure 35. *Preliminary spatial transcriptomic analysis.*

Spatial transcriptomic heatmap visualisation on a histological section of OED tissue indicating the number of altered genes. Clusters 1 and 3- Epithelial (blue, green), Cluster 2, 4 and 5- Stroma (Orange, Red and Purple). Graph showing fold change in gene expression. Unpublished data (2023).

Whilst spatial transcriptomics is a highly powerful technique with brilliant potential to map the whole transcriptome and analyse gene profiling information in the morphological context of tissue[175,176]. It is an extremely expensive and time-consuming technique, presenting unique analytical challenges that require complex computational processes. Cell DIVE™ (Leica, Microsystems) technology provides an alternative multiplex imaging solution for spatial biomarker analysis, integrating the powers of cell biology, software algorithms and precision imaging (Figure 36). The candidate has been successfully awarded a Fellowship jointly funded by the Faculty of Dental Surgery and Association of British Academic Oral and Maxillofacial Surgeons to study spatial biomarkers in OED and explore multiplex imaging using this innovative technique. As part of this work, markers identified through preliminary work as being potentially relevant (such as cytokeratin 5/6, TGFβ, CD34, CD4/8/20/45, p53, Ki67, PD-L1) will be visualised, quantified and correlated to clinical outcomes. This work will form the foundation of future grant applications for further multicentre analysis to improve clinical treatment decisions, by assisting with risk stratification of OED lesions, helping improve patient outcomes.

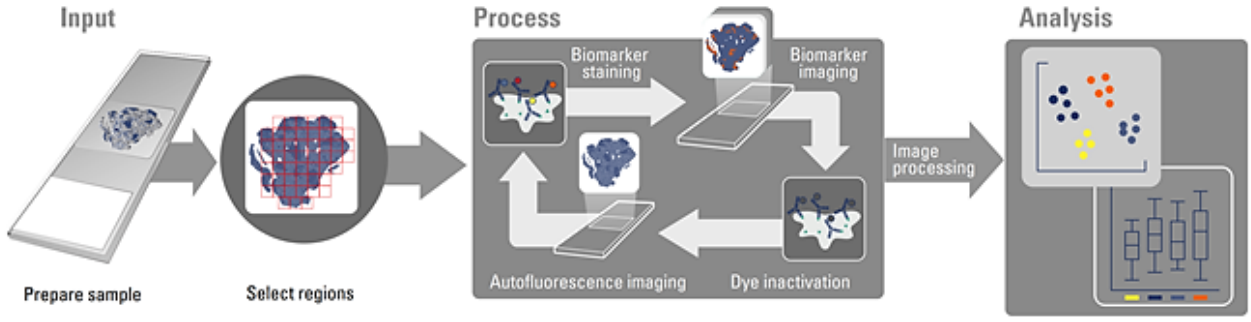

Figure 36. *Cell DIVE™ workflow pipeline.*

Image Source: Cell DIVE™ Open Multiplexed Imaging Solutions, Leica Microsystems.

9.4 Research Impact & Conclusions

Research presented in this thesis shows correlations between individual OED histological features, digital morphometric features, and prognosis for the first time on the largest digital multicentre cohort to date. The developed models have demonstrated better predictive strength for malignant transformation risk in OED (AUROC: 0.74-0.81) compared to 'gold-standard' histological grading (AUROC: 0.60-69) with superior performance maintained on unseen external datasets. Trained AI models segmented and classified OED epithelium, immune and stromal cells with good accuracy (F1 scores: 0.78-0.89). A novel transformer-based pipeline (ODYN) was developed for reliable detection and diagnosis of OED (F1-score of 0.71 for OED segmentation, 0.96 for classification of dysplastic vs non-dysplastic tissue) and prediction of malignant transformation risk (ODYN-score) demonstrating good accuracy (AUROC of 0.75), outperforming other state-of-the-art methods, and existing clinical grading systems. The thesis findings are strongly translational with a significant potential to aid OED diagnosis, predict the risk of malignant transformation to allow early detection of 'at risk' lesions and inform treatment decisions. Prospective validation on larger multicentre cohorts is required to further explore and establish the clinical utility of our developed tools.

References

[1] NICE, Evidence standards framework for digital health technologies. December 2018. www.nice.org.uk/ Media/Default/About/what-we-do/our-programmes/ evidence-standards-framework/digital-evidence-standardsframework.pdf (Accessed 1st October 2023).

[2] M. Hancock, 'My vision for a more tech-driven NHS. Secretary of State for Health and Social Care Matt Hancock's speech at NHS Expo 2018', 6 Sept 2018, www.gov.uk/government/speeches/my-vision-for-a-moretech-driven-nhs, (Accessed 1st October 2023).

[3] Health Education England, The Topol Review: Preparing the healthcare workforce to deliver the digital future, Interim Report. June 2018. www.hee.nhs.uk/sites/ default/files/documents/Topol%20Review%20interim%20 report_0.pdf (Accessed 1st October 2023)

[4] NHS England, Developing the long-term plan for the NHS Briefing from the Long-Term Plan Engagement Team. August 2018. www.engage.england.nhs.uk/consultation/ developing-the-long-term-plan-for-the-nhs/user_uploads/ developing-the-long-term-plan-for-the-nhs-v2.pdf (Accessed 1st October 2023)

[5] Oral and dental health top 10 priorities. James Lind Alliance Priority Setting Partnerships. 2018. National Institute for Health Research. http://www.jla.nihr.ac.uk/priority-setting-partnerships/oral-and-dental-health/top-10-priorities.htm (Accessed 1st October 2023).

[6] Testing times to come? An evaluation of pathology capacity across the UK. Cancer research UK. Nov 2016. https://www.cancerresearchuk.org/sites/default/files/testing_times_to_come_nov_16_cruk.pdf (Accessed 1st October 2023).

[7] Mahmood H, Shaban M, Rajpoot N, Khurram SA. Artificial Intelligence-based methods in head and neck cancer diagnosis: an overview. Brit J Cancer. 2021;124(12):1934-40.

[8] Mahmood H, Shaban M, Indave BI, Santos-Silva AR, Rajpoot N, Khurram SA. Use of artificial intelligence in diagnosis of head and neck precancerous and cancerous lesions: a systematic review. Oral Oncology. 2020 Nov 1;110:104885.

[9] Mahmood H, Bradburn M, Rajpoot N, Islam NM, Kujan O, Khurram SA. Prediction of malignant transformation and recurrence of oral epithelial dysplasia using architectural and cytological feature specific prognostic models. Modern Pathology. 2022 Sep 1;35(9):1151-9.

[10] Mahmood H, Shephard A, Hankinson P, Bradburn M, Araujo AL, Santos-Silva AR, Lopes MA, Vargas PA, McCombe KD, Craig SG, James J. Development and validation of a multivariable model for prediction of malignant transformation and recurrence of oral epithelial dysplasia. British Journal of Cancer. 2023 Sep 27:1-9.

[11] Sathyamoorthy H, Mahmood H, Zubir AZ, Hankinson P, Khurram SA. Prognostic importance of mitosis quantification and PHH3 expression in oral epithelial dysplasia. Virchows Archiv. 2023 Oct 26:1-3.

[12] Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians. 2018 Nov;68(6):394-424.

[13] Pai SI, Westra WH. Molecular pathology of head and neck cancer: implications for diagnosis, prognosis, and treatment. Annual Review of Pathology: Mechanisms of Disease. 2009 Feb 28;4:49-70.

[14] Cancer Research UK (2018). Head and neck cancers incidence statistics. Available at URL: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/head-and-neck-cancers (accessed 20 October 2023).

[15] Warnakulasuriya S. Global epidemiology of oral and oropharyngeal cancer. Oral oncology. 2009 Apr 1;45(4-5):309-16.

[16] Shaw R, Beasley N. Aetiology and risk factors for head and neck cancer: United Kingdom National Multidisciplinary Guidelines. The Journal of Laryngology & Otology. 2016 May;130(S2):S9-12.

[17] 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.

[18] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Alcohol consumption and ethyl carbamate. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. 2010;96:3.

[19] Lumerman H, Freedman P, Kerpel S. Oral epithelial dysplasia and the development of invasive squamous cell carcinoma. Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology and Endodontics. 1995 Mar 1;79(3):321-9.

[20] Sankaranarayanan R, Ramadas K, Amarasinghe H, Subramanian S, Johnson N. Oral cancer: prevention, early detection, and treatment. Cancer: disease control priorities, third edition (volume 3). 2015 Nov 1.

[21] Chandu A, Smith AC, Rogers SN. Health-related quality of life in oral cancer: a review. Journal of oral and maxillofacial surgery. 2006 Mar 1;64(3):495-502.

[22] Office for National Statistics (2020). Information request for Cancer Registration Statistics for England. Request made by Oral Health Foundation.

[23] State of Mouth Cancer UK Report, 2022. Oral Health Foundation. Available via URL: https://www.dentalhealth.org/thestateofmouthcancer (accessed 29 September 2023).

[24] Kalavrezos N, Scully C. Mouth cancer for clinicians Part 2: epidemiology. Dental update. 2015 May 2;42(4):354-9.

[25] Public Health England. Oral Cancer in England. A report on incidence, survival and mortality rates of oral cancer in England, 2012 to 2016. Available at URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/891699/Oral_cancer_report_170420.pdf (accessed 20 October 2023)

[26] Sankaranarayanan R, Ramadas K, Thomas G, Muwonge R, Thara S, Mathew B, Rajan B, Trivandrum Oral Cancer Screening Study Group. Effect of screening on oral cancer mortality in Kerala, India: a cluster-randomised controlled trial. The Lancet. 2005 Jun 4;365(9475):1927-33.

[27] White DA, Tsakos G, Pitts NB, Fuller E, Douglas GV, Murray JJ, Steele JG. Adult Dental Health Survey 2009: common oral health conditions and their impact on the population. British dental journal. 2012 Dec;213(11):567-72.

[28] Muller P, Belot A, Morris M, Rachet B, Cancer Research UK Cancer Survival Group, London School of Hygiene and Tropical Medicine. Net survival and the probability of cancer death from rare cancers. Available from http://csg.lshtm.ac.uk/rare-cancers/ (accessed 1 October 2023).

[29] Van der Waal I, de Bree R, Brakenhoff R, Coebergh JW. Early diagnosis in primary oral cancer: is it possible? Med Oral Patol Oral Cir Bucal. 2011;16(3):e300–5. doi: 10.4317/medoral.16.e300.

[30] IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Human immunodeficiency viruses and human T-cell lymphotropic viruses. IARC monographs on the evaluation of carcinogenic risks to humans. 1996.

[31] Papillomaviruses H. IARC monographs on the evaluation of carcinogenic risks to humans. Lyon, France: IARC. 2011.

[32] 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. Clinical Cancer Research. 2000 Mar 1;6(3):1046-51.

[33] Gupta PC, Mehta FS, Daftary DK, Pindborg JJ, Bhonsle RB, Jalnawalla PN, Sinor PN, Pitkar VK, Murti PR, Irani RR, Shah HT. Incidence rates of oral cancer and natural history of oral precancerous lesions in a 10-year follow-up study of Indian villagers. Community dentistry and oral epidemiology. 1980;8(6):283-333.

[34] Ranganathan K, Kavitha L. Oral epithelial dysplasia: Classifications and clinical relevance in risk assessment of oral potentially malignant disorders. Journal of oral and maxillofacial pathology: JOMFP. 2019 Jan;23(1):19.

[35] International Head and Neck Epidemiology (INHANCE) consortium. University of Utah: Utah, 2018. Available at https://www.inhance.utah.edu/ (accessed July 2018).

[36] Conway D I, Hashibe M, Boffetta P et al. Enhancing epidemiologic research on head and neck cancer: INHANCE The international head and neck cancer epidemiology consortium. Oral Oncol 2009; 45: 743–746.

[37] Lubin JH, Purdue M, Kelsey K et al. Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: a pooled analysis of case-control studies. Am J Epidemiol 2009; 170: 937–947.

[38] Bagnardi V, et al (2015) 'Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis', Br J Cancer. 2015 Feb 3;112(3):580-93.

[39] Llewellyn CD, Johnson NW, Warnakulasuriya KA. Risk factors for oral cancer in newly diagnosed patients aged 45 years and younger: a case–control study in Southern England. Journal of oral pathology & medicine. 2004 Oct;33(9):525-32.

[40] Mehanna HM, Rattay T, Smith J, McConkey CC. Treatment and Follow-up of Oral Dysplasia - a Systematic Review and Meta-Analysis. Head Neck-J Sci Spec. 2009;31(12):1600-9.

[41] Warnakulasuriya S, Ariyawardana A. Malignant transformation of oral leukoplakia: a systematic review of observational studies. J Oral Pathol Med. 2016;45(3):155-66.

[42] Speight PM. Update on oral epithelial dysplasia and progression to cancer. Head and neck pathology. 2007 Sep 1;1(1):61-6.

[43] Locca O, Sollecito TP, Alawi F, Weinstein GS, Newman JG, De Virgilio A, Di Maio P, Spriano G, Pardinas Lopez S, Shanti RM. Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype. Head & neck. 2020 Mar;42(3):539-55.

[44] Warnakulasuriya S, Kujan O, Aguirre-Urizar JM, Bagan JV, González-Moles MÁ, Kerr AR, Lodi G, Mello FW, Monteiro L, Ogden GR, Sloan P. Oral potentially malignant disorders: A consensus report from an international seminar on nomenclature and classification, convened by the WHO Collaborating Centre for Oral Cancer. Oral diseases. 2020 Oct 31.

[45] Reibel J. Prognosis of oral pre-malignant lesions: significance of clinical, histopathological, and molecular biological characteristics. Critical Reviews in Oral Biology & Medicine. 2003 Jan;14(1):47-62.

[46] Speight PM, Khurram SA, Kujan O. Oral potentially malignant disorders: risk of progression to malignancy. Oral surgery, oral medicine, oral pathology and oral radiology. 2018 Jun 1;125(6):612-27.

[47] Field EA, McCarthy CE, Ho MW, Rajlawat BP, Holt D, Rogers SN, et al. The management of oral epithelial dysplasia: The Liverpool algorithm. Oral Oncol. 2015;51(10):883-7.

[48] Ho MW, Risk JM, Woolgar JA, Field EA, Field JK, Steele JC, et al. The clinical determinants of malignant transformation in oral epithelial dysplasia. Oral Oncol. 2012;48(10):969-76.

[49] Sultan AS, Elgharib MA, Tavares T, Jessri M, Basile JR. The use of artificial intelligence, machine learning and deep learning in oncologic histopathology. J Oral Pathol Med. 2020;49(9):849-56.

[50] Mashberg A, Boffetta P,Winkelman R, et al. Tobacco smoking, alcohol drinking, and cancer of the oral cavity and oropharynx among U.S. veterans. Cancer 1993;72:1369-1375.

[51] Lewin F, Norell SE, Johansson H, et al. Smoking tobacco, oral snuff, and alcohol in the aetiology of squamous cell carcinoma of the head and neck. A population-based case-referent study in Sweden. Cancer 1998;82:1367-1375.

[52] WHO, Classification of tumours editorial board. Head and neck tumours [Internet; beta version ahead of print]. 5th ed. Lyon, France: International Agency for Research on Cancer; 2023.

[53] El-Naggar AK, Chan JK, Grandis JR. WHO classification of head and neck tumours. 2017.

[54] Odell E, Kujan O, Warnakulasuriya S, Sloan P. Oral epithelial dysplasia: recognition, grading and clinical significance. Oral Diseases. 2021 Aug 21.

[55] Barnes L, Eveson JW, Sidransky D, Reichart P, editors. Pathology and genetics of head and neck tumours. IARC; 2005.

[56] Warnakulasuriya S. Histological grading of oral epithelial dysplasia: revisited. The Journal of pathology. 2001 Jul;194(3):294-7.

[57] Kujan O, Khattab A, Oliver RJ, Roberts SA, 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 oncology. 2007 Mar 1;43(3):224-31.

[58] Leemans CR, Snijders PJ, Brakenhoff RH. The molecular landscape of head and neck cancer. Nature Reviews Cancer. 2018 May;18(5):269-82.

[59] Califano J, Van Der Riet P, Westra W, Nawroz H, Clayman G, Piantadosi S, Corio R, Lee D, Greenberg B, Koch W, Sidransky D. Genetic progression model for head and neck cancer: implications for field cancerization. Cancer research. 1996 Jun 1;56(11):2488-92.

[60] Kujan O, Oliver RJ, Khattab A, Roberts SA, Thakker N, Sloan P. Evaluation of a new binary system of grading oral epithelial dysplasia for prediction of malignant transformation. Oral oncology. 2006 Nov 1;42(10):987-93.

[61] Nankivell P, Williams H, Matthews P, Suortamo S, Snead D, McConkey C, Mehanna H. The binary oral dysplasia grading system: validity testing and suggested improvement. Oral surgery, oral medicine, oral pathology and oral radiology. 2013 Jan 1;115(1):87-94.

[62] Silva BSD, Batista DCR, Roriz CFD, Silva LR, Normando AGC, Silva ARD, et al. Binary and WHO dysplasia grading systems for the prediction of malignant transformation of oral leukoplakia and erythroplakia: a systematic review and meta-analysis. Clinical Oral Investigations. 2021;25(7):4329-40.

[63] Hankinson P, Mahmood H, Walsh H, Speight PM, Khurram SA. Demystifying oral epithelial dysplasia: a histological guide. Pathology. 2023 Nov 16.

[64] Smith J, Rattay T, McConkey C, Helliwell T, Mehanna H. Biomarkers in dysplasia of the oral cavity: a systematic review. Oral oncology. 2009 Aug 1;45(8):647-53.

[65] Rivera C, Gallegos R, Figueroa C. Biomarkers of progression to oral cancer in patients with dysplasia: A systematic review. Molecular and Clinical Oncology. 2020 Nov 1;13(5):1-.

[66] Kaur J, Matta A, Kak I, Srivastava G, Assi J, Leong I, Witterick I, Colgan TJ, MacMillan C, Siu KM, Walfish PG. S100A7 overexpression is a predictive marker for high risk of malignant transformation in oral dysplasia. International journal of cancer. 2014 Mar 15;134(6):1379-88.

[67] Li X, Wan L, Geng J, Wu CL, Bai X. Aldehyde dehydrogenase 1A1 possesses stem-like properties and predicts lung cancer patient outcome. Journal of Thoracic Oncology. 2012 Aug 1;7(8):1235-45.

[68] Yang L, Ren Y, Yu X, Qian F, Bian BS, Xiao HL, Wang WG, Xu SL, Yang J, Cui W, Liu Q. ALDH1A1 defines invasive cancer stem-like cells and predicts poor prognosis in patients with esophageal squamous cell carcinoma. Modern Pathology. 2014 May;27(5):775-83.

[69] Morimoto K, Kim SJ, Tanei T, Shimazu K, Tanji Y, Taguchi T, Tamaki Y, Terada N, Noguchi S. Stem cell marker aldehyde dehydrogenase 1-positive breast cancers are characterized by negative estrogen receptor, positive human epidermal growth factor receptor type 2, and high Ki67 expression. Cancer science. 2009 Jun;100(6):1062-8.

[70] Barzegar Behrooz A, Syahir A, Ahmad S. CD133: beyond a cancer stem cell biomarker. Journal of drug targeting. 2019 Mar 16;27(3):257-69.

[71] Krishnan H, Rayes J, Miyashita T, Ishii G, Retzbach EP, Sheehan SA, Takemoto A, Chang YW, Yoneda K, Asai J, Jensen L. Podoplanin: an emerging cancer biomarker and therapeutic target. Cancer science. 2018 May;109(5):1292-9.

[72] Shirata NK, Longatto A, Roteli-Martins C, Espoladore LMW, Pittoli JE, Syrjanen K. Applicability of liquid-based cytology to the assessment of DNA content in cervical lesions using static cytometry. Analytical and Quantitative Cytology and Histology. 2003;25(4):210-4.

[73] Partridge M, Pateromichelakis S, Phillips E, Emilion GG, A'Hern RP, Langdon JD. A case-control study confirms that microsatellite assay can identify patients at risk of developing oral squamous cell carcinoma within a field of cancerization. Cancer Research. 2000;60(14):3893-8.

[74] Abdelsalam M, Mayall BH, Chew K, Silverman S, Greenspan JS. Which oral white lesions will become malignant - an image cytometric study. Oral Surgery Oral Medicine Oral Pathology Oral Radiology and Endodontics. 1990;69(3):345-50.

[75] Abu Eid R, Landini G. Oral Epithelial Dysplasia: Can Quantifiable Morphological Features Help in the Grading Dilemma? First ImageJ User and Developer Conference Proceedings; Luxembourg 2006.

[76] Masthan KM, Babu NA, Dash KC, Elumalai M. Advanced diagnostic aids in oral cancer. Asian Pacific Journal of Cancer Prevention. 2012;13(8):3573-6.

[77] Messadi DV. Diagnostic aids for detection of oral precancerous conditions. International journal of oral science. 2013 Jun;5(2):59-65.

[78] Patton LL, Epstein JB, Kerr AR. Adjunctive techniques for oral cancer examination and lesion diagnosis: a systematic review of the literature. The Journal of the American Dental Association. 2008 Jul 1;139(7):896-905.

[79] Sciubba JJ . Improving detection of precancerous and cancerous oral lesions. Computer-assisted analysis of the oral brush biopsy. U.S. Collaborative OralCDx Study Group. J Am Dent Assoc 1999; 130(10): 1445–1457.

[80] Rick GM . Oral brush biopsy: the problem of false positives. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 2003; 96(3): 252–258.

[81] Li Y, St John MA, Zhou X et al. Salivary transcriptome diagnostics for oral cancer detection. Clin Cancer Res 2004; 10(24): 8442–8450.

[82] Cheng YS, Rees T, Wright J. A review of research on salivary biomarkers for oral cancer detection. Clinical and translational medicine. 2014 Dec;3:1-0.

[83] Williams BR, Amon A. Aneuploidy: cancer's fatal flaw?. Cancer research. 2009 Jul 1;69(13):5289-91.

[84] Garnis C, Chari R, Buys TP, Zhang L, Ng RT, Rosin MP, Lam WL. Genomic imbalances in precancerous tissues signal oral cancer risk. Molecular cancer. 2009 Dec;8(1):1-7.

[85] Sperandio M, Brown AL, Lock C, Morgan PR, Coupland VH, Madden PB, Warnakulasuriya S, Møller H, Odell EW. Predictive value of dysplasia grading and DNA ploidy in malignant transformation of oral potentially malignant disorders. Cancer Prevention Research. 2013 Aug 1;6(8):822-31.

[86] Seoane J, Asenjo JA, Bascones A, Varela-Centelles PI, Romero MA. Flow cytometric DNA ploidy analysis of oral cancer comparison with histologic grading. Oral oncology. 1999 May 1;35(3):266-72.

[87] Sieroń A, Sieroń-Stołtny K, Kawczyk-Krupka A, Latos W, Kwiatek S, Straszak D, et al. The role of fluorescence diagnosis in clinical practice. Onco Targets Ther. 2013;6:977–82.

[88] Schantz SP, Kolli V, Savage HE, Yu G, Shah JP, Harris DE, et al. In vivo native cellular fluorescence and histological characteristics of head and neck cancer. Clin Cancer Res. 1998;4:1177–82

[89] Hung J, Lam S, LeRiche JC, Palcic B. Autofluorescence of normal and malignant bronchial tissue. Lasers Surg Med. 1991;11:99–105.

[90] Kennedy JC, Pottier RH . Endogenous protoporphyrin IX, a clinically useful photosensitizer for photodynamic therapy. J Photochem Photobiol B 1992; 14(4): 275–292.

[91] Ebihara A, Liaw LH, Krasieva TB . Detection and diagnosis of oral cancer by light-induced fluorescence. Lasers Surg Med 2003; 32(1): 17–24.

[92] Heath JP, Hunter KD, Murdoch C, Walker DC. Computational Modelling for Electrical Impedance Spectroscopy-Based Diagnosis of Oral Potential Malignant Disorders (OPMD). Sensors. 2022 Aug 8;22(15):5913.

[93] González-Correa, C.A.; Brown, B.H.; Smallwood, R.H.; Kalia, N.; Stoddard, C.J.; Stephenson, T.J.; Haggie, S.J.; Slater, D.N.; Bardhan, K.D. Virtual biopsies in Barrett's esophagus using an impedance probe. Ann. N. Y. Acad. Sci. 1999, 873, 313–321.

[94] Brown, B.H.; Tidy, J.A.; Boston, K.; Blackett, A.D.; Smallwood, R.H.; Sharp, F. Relation between tissue structure and imposed electrical current flow in cervical neoplasia. Lancet 2000, 355, 892–895.

[95] Mohr, P.; Birgersson, U.; Berking, C.; Henderson, C.; Trefzer, U.; Kemeny, L.; Sunderkotter, C.; Dirschka, T.; Motley, R.; Frohm-Nilsson, M.; et al. Electrical impedance spectroscopy as a potential adjunct diagnostic tool for cutaneous melanoma. Ski. Res. Technol. 2013, 19, 75–83

[96] Braun, R.P.; Mangana, J.; Goldinger, S.; French, L.; Dummer, R.; Marghoob, A.A. Electrical Impedance Spectroscopy in Skin Cancer Diagnosis. Dermatol. Clin. 2017, 35, 489–493.

[97] Halter, R.J.; Schned, A.; Heaney, J.; Hartov, A.; Schutz, S.; Paulsen, K.D. Electrical impedance spectroscopy of benign and malignant prostatic tissues. J. Urol. 2008, 179, 1580–1586.

[98] Wilkinson, B.A.; Smallwood, R.H.; Keshtar, A.; Lee, J.A.; Hamdy, F.C. Electrical impedance spectroscopy and the diagnosis of bladder pathology: A pilot study. J. Urol. 2002, 168, 1563–1567.

[99] Murdoch C, Brown BH, Hearnden V, Speight PM, D'Apice K, Hegarty AM, Tidy JA, Healey TJ, Highfield PE, Thornhill MH. Use of electrical impedance spectroscopy to detect malignant and potentially malignant oral lesions. International journal of nanomedicine. 2014 Sep 23:4521-32.

[100] Ciompi F, Veta M, Van Der Laak J, Rajpoot N. Editorial Computational Pathology. IEEE Journal of Biomedical and Health Informatics. 2021 Feb;25(2):303-6.

[101] Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Medical image analysis. 2016 Oct 1;33:170-5.

[102] Awan R, Sirinukunwattana K, Epstein D, Jefferyes S, Qidwai U, Aftab Z et al. Glandular morphometrics for objective grading of colorectal adenocarcinoma histology images. Scientific reports. 2017 Dec 4;7(1):16852.

[103] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. Artificial intelligence in digital pathology—new tools for diagnosis and precision oncology. Nature Reviews Clinical Oncology. 2019 Nov;16(11):703-15.

[104] Zormpas-Petridis K, Failmezger H, Raza SE, Roxanis I, Jamin Y, Yuan Y. Superpixel-based Conditional Random Fields (SuperCRF): Incorporating global and local context for enhanced deep learning in melanoma histopathology. Frontiers in oncology. 2019;9:1045.

[105] Wang S, Yang DM, Rong R, Zhan X, Fujimoto J, Liu H et al. Artificial Intelligence in Lung Cancer Pathology Image Analysis. Cancers. 2019 Nov;11(11):1673.

[106] Sirinukunwattana K, Shan e Ahmed Raza, Tsang YW, Snead DR, Cree IA, Rajpoot NM. Locality sensitive deep learning for detection and classification of nuclei in routine colon cancer histology images. IEEE Trans. Med. Imaging. 2016 May 1;35(5):1196-206.

[107] Veta M, Van Diest PJ, Willems SM, Wang H, Madabhushi A, Cruz-Roa A et al.. Assessment of algorithms for mitosis detection in breast cancer histopathology images. Medical image analysis. 2015 Feb 1;20(1):237-48.

[108] Bejnordi, B. E. *et al*. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama* 318, 2199–2210 (2017).

[109] Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. Computational and structural biotechnology journal. 2015 Dec 31;13:8-17

[110] Ehteshami Bejnordi B, Veta M, Rajpoot N, et al. Diagnostic Assessment of Deep Learning Algorithms for Detection of Lymph Node Metastases in Women With Breast Cancer. JAMA. 2017 Dec 12;318(22):2199-2210. doi: 10.1001/jama.2017.14585.

[111] Qaiser, T. et al. HER2 challenge contest: a detailed assessment of automated her 2 scoring algorithms in whole slide images of breast cancer tissues. Histopathol. **72**, 227–238 (2018).

[112] Bulten W, Kartasalo K, Chen PH, Ström P, Pinckaers H, Nagpal K, Cai Y, Steiner DF, van Boven H, Vink R, Hulsbergen-van de Kaa C. Artificial intelligence for diagnosis and Gleason grading of prostate cancer: the PANDA challenge. Nature medicine. 2022 Jan;28(1):154-63.

[113] Sandbank J, Bataillon G, Nudelman A, Krasnitsky I, Mikulinsky R, Bien L, Thibault L, Albrecht Shach A, Sebag G, Clark DP, Laifenfeld D. Validation and real-world clinical application of an artificial intelligence algorithm for breast cancer detection in biopsies. NPJ Breast Cancer. 2022 Dec 6;8(1):129.

[114] Ayodele TO. Types of machine learning algorithms. New advances in machine learning. 2010 Feb 1;3:19-48.

[115] Niazi MK, Parwani AV, Gurcan MN. Digital pathology and artificial intelligence. The lancet oncology. 2019 May 1;20(5):e253-61.

[116] Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS. Deep learning for visual understanding: A review. Neurocomputing. 2016 Apr 26;187:27-48.

[117] LeCun Y, Bengio Y, Hinton G. Deep learning. nature. 2015 May;521(7553):436-44.

[118] Casado-García Á, Domínguez C, García-Domínguez M, Heras J, Inés A, Mata E, Pascual V. CLoDSA: a tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks. BMC bioinformatics. 2019 Dec;20:1-4.

[119] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement.

[120] StataCorp. 2021. Stata Statistical Software: Release 17. College Station, TX: StataCorp LLC.

[121] Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-74.

[122] Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. BMC medical research methodology. 2013 Dec;13:1-7.

[123] Jones K, Jordan RCK. Patterns of second-opinion diagnosis in oral and maxillofacial pathology. Oral Surg Oral Med O. 2010;109(6):865-9.

[124] Sunil Mutalik VSM, Keerthilatha M Pai, Venkatesh G. Naikmasur, Khoo Suan Phaik. Oral Leukoplakia – Is Biopsy at the Initial Appointment a Must? J Clin Diagn Res 2014.

[125] Lu C, Lewis JS, Dupont WD, Plummer WD, 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. Modern Pathol. 2017;30(12):1655-65.

[126] Baik J, Ye Q, Zhang LW, Poh C, Rosin M, MacAulay C, et al. Automated classification of oral premalignant lesions using image cytometry and Random Forests-based algorithms. Cell Oncol. 2014;37(3):193-202.

[127] Krishnan MMR, Pal M, Bomminayuni SK, Chakraborty C, Paul RR, Chatterjee J, et al. Automated classification of cells in sub-epithelial connective tissue of oral sub-mucous fibrosis-An SVM based approach. Comput Biol Med. 2009;39(12):1096-104.

[128] Mookiah MRK, Shah P, Chakraborty C, Ray AK. Brownian Motion Curve Based Textural Classification and Its Application in Cancer Diagnosis. Anal Quant Cytol. 2011;33(3):158-68.

[129] Krishnan MMR, Venkatraghavan V, Acharya UR, Pal M, Paul RR, Min LC, et al. Automated oral cancer identification using histopathological images: A hybrid feature extraction paradigm. Micron. 2012;43(2-3):352-64.

[130] Landini G, Othman IE. Architectural analysis of oral cancer, dysplastic, and normal epithelia. Cytometry A. 2004;61(1):45-55.

[131] Marcelpoil R, Davoine F, Robertnicoud M. Cellular Sociology - Parametrization of Spatial Relationships Based on Voronoi Diagram and Ulam Trees. Fractals in Biology and Medicine. 1994:201-9.

[132] Kayser K, Stute H, Bubenzer J, Paul J. Combined Morphometrical and Syntactic Structure-Analysis as Tools for Histomorphological Insight into Human Lung-Carcinoma Growth. Anal Cell Pathol. 1990;2(3):167-78.

[133] Kayser K, Sandau K, Bohm G, Kunze KD, Paul J. Analysis of Soft-Tissue Tumors by an Attributed Minimum Spanning Tree. Anal Quant Cytol. 1991;13(5):329-34.

[134] Bankhead P, Loughrey MB, Fernández JA, Dombrowski Y, McArt DG, Dunne PD et al. QuPath: Open-source software for digital pathology image analysis. Sci. Rep. 4;7(1):1-7 (2017).

[135] Ruifrok, A C, and D A Johnston. Quantification of Histochemical Staining by Color Deconvolution. Analytical and Quantitative Cytology and Histology, 23 (4): 291–9.

[136] Shephard AJ, Mahmood H, Raza SE, Araujo AL, Santos-Silva AR, Lopes MA, Vargas PA, McCombe K, Craig S, James J, Brooks J. Transformer-based Model for Oral Epithelial Dysplasia Segmentation. arXiv preprint arXiv:2311.05452. 2023 Nov 9.

[137] Shephard AJ, Bashir RMS, Mahmood H, et al. A Fully Automated and Explainable Algorithm for the Prediction of Malignant Transformation in Oral Epithelial Dysplasia. arXiv 2023; : 1–35.

[138] Moore SR, Johnson NW, Pierce AM, Wilson DF. The epidemiology of mouth cancer: a review of global incidence. Oral Dis. 2000;6(2):65-74.

[139] Steele TO, Meyers A. Early Detection of Premalignant Lesions and Oral Cancer. Otolaryng Clin N Am. 2011;44(1):221-+.

[140] Sudbo J, Bankfalvi A, Bryne M, Marcelpoil R, Boysen M, Piffko J, et al. Prognostic value of graph theory-based tissue architecture analysis in carcinomas of the tongue. Lab Invest. 2000;80(12):1881-9.

[141] Kim J, Cho J. Delaunay Triangulation-Based Spatial Clustering Technique for Enhanced Adjacent Boundary Detection and Segmentation of LiDAR 3D Point Clouds. Sensors-Basel. 2019;19(18).

[142] Tilakaratne WM, Jayasooriya PR, Jayasuriya NS, De Silva RK. Oral epithelial dysplasia: Causes, quantification, prognosis, and management challenges. Periodontology 2000. 2019;80(1):126-47.

[143] Litjens G, Kooi T, Bejnordi BE, et al. A survey on deep learning in medical image analysis. Med Image Anal 2017; 42: 60–88.

[144] Madabhushi A, Lee G. Image analysis and machine learning in digital pathology: Challenges and opportunities. Med Image Anal 2016; 33: 170–5.

[145] Litjens G, Sánchez CI, Timofeeva N, et al. Deep learning as a tool for increased accuracy and efficiency of histopathological diagnosis. Sci Rep 2016; 6: 1–11.

[146] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). 2015: 234–41.

[147] Chen L-C, Papandreou G, Kokkinos I, Murphy K, Yuille AL. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. In: International Conference on Learning Representations (ICLR). 2015. DOI:10.1109/TPAMI.2017.2699184.

[148] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2016; 9908 LNCS: 630–45.

[149] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. 2020. http://arxiv.org/abs/2010.11929.

[150] He K, Gan C, Li Z, et al. Transformers in medical image analysis. Intell. Med. 2023; 3: 59–78.

[151] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In: Advances in Neural Information Processing Systems. 2017: 5999–6009.

[152] Chen J, Lu Y, Yu Q, et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv 2021; : 1–13.

[153] Dai Y, Gao Y, Liu F. TransMed: Transformers Advance Multi-Modal Medical Image Classification. Diagnostics 2021; 11: 1384.

[154] Lin A, Chen B, Xu J, Zhang Z, Lu G, Zhang D. DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation. IEEE Trans Instrum Meas 2022; 71: 1–13.

[155] Cao H, Wang Y, Chen J, et al. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 2023; 13803 LNCS: 205–18.

[156] Myronenko A, Xu Z, Yang D, Roth HR, Xu D. Accounting for Dependencies in Deep Learning Based Multiple Instance Learning for Whole Slide Imaging. Springer International Publishing, 2021 DOI:10.1007/978-3-030-87237-3_32.

[157] Vu QD, Rajpoot K, Raza SEA, Rajpoot N. Handcrafted Histological Transformer (H2T): Unsupervised representation of whole slide images. Med Image Anal 2023; 85: 102743.

[158] Pocock J, Graham S, Vu QD, et al. TIAToolbox as an end-to-end library for advanced tissue image analytics. Commun Med 2022; 2: 120.

[159] Bilal M, Raza SEA, Azam A, et al. Development and validation of a weakly supervised deep learning framework to predict the status of molecular pathways and key mutations in colorectal cancer from routine histology images: a retrospective study. Lancet Digit Heal 2021; 3: e763–72.

[160] Wang Y, Ma X, Chen Z, Luo Y, Yi J, Bailey J. Symmetric Cross Entropy for Robust Learning with Noisy Labels.

[161] Chen L-C, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation. arXiv 2017. http://arxiv.org/abs/1706.05587.

[162] Baheti B, Innani S, Gajre S, Talbar S. Eff-UNet: A novel architecture for semantic segmentation in unstructured environment. IEEE Comput Soc Conf Comput Vis Pattern Recognit Work 2020; 2020-June: 1473–81.

[163] Bashir RM, Shephard AJ, Mahmood H, Azarmehr N, Raza SE, Khurram SA, Rajpoot NM. A digital score of peri-epithelial lymphocytic activity predicts malignant transformation in oral epithelial dysplasia. The Journal of Pathology. 2023 Feb 22.

[164] Liu Y, Bilodeau E, Pollack B, Batmanghelich K. Automated detection of premalignant oral lesions on whole slide images using convolutional neural networks. Oral Oncol 2022; 134: 106109.

[165] Shephard AJ, Bashir RM, Mahmood H, Jahanifar M, Minhas F, Raza SE, McCombe KD, Craig SG, James J, Brooks J, Nankivell P. A Fully Automated and Explainable Algorithm for the Prediction of Malignant Transformation in Oral Epithelial Dysplasia. arXiv preprint arXiv:2307.03757. 2023 Jul 6.

[166] Komori T. Grading of adult diffuse gliomas according to the 2021 WHO Classification of Tumors of the Central Nervous System. Laboratory Investigation. 2022 Feb 1;102(2):126-33.

[167] Zhang T, Yu JM, Wang YQ, Yin DD, Fang LJ. WHO grade I meningioma subtypes: MRI features and pathological analysis. Life sciences. 2018 Nov 15;213:50-6.

[168] Jiang M, Bennani NN, Feldman AL. Lymphoma classification update: T-cell lymphomas, Hodgkin lymphomas, and histiocytic/dendritic cell neoplasms. Expert review of hematology. 2017 Mar 4;10(3):239-49.

[169] Takayama T, Kosuge T, Yamazaki S, Hasegawa H, Okazaki N, Takayasu K, Hirohashi S, Sakamoto M, Makuuchi M, Motoo Y. Malignant transformation of adenomatous hyperplasia to hepatocellular carcinoma. The Lancet. 1990 Nov 10;336(8724):1150-3.

[170] Mirzaiian E, Ghods ZS, Tavangar SM, Emami B, Oraei M, Safyari R, Saffar H. Utility of PHH3 in evaluation of mitotic index in breast carcinoma and impact on tumor grade. Asian Pacific journal of cancer prevention: APJCP. 2020;21(1):63.

[171] Klöppel G. Tumour biology and histopathology of neuroendocrine tumours. Best practice & research Clinical endocrinology & metabolism. 2007 Mar 1;21(1):15-31.

[172] Gannot G, Gannot I, Vered H, Buchner A, Keisari Y. Increase in immune cell infiltration with progression of oral epithelium from hyperkeratosis to dysplasia and carcinoma. British journal of cancer. 2002 May;86(9):1444-8.

[173] Gan CP, Lee BK, Lau SH, Kallarakkal TG, Zaini ZM, Lye BK, Zain RB, Sathasivam HP, Yeong JP, Savelyeva N, Thomas G. Transcriptional analysis highlights three distinct immune profiles of high-risk oral epithelial dysplasia. Frontiers in Immunology. 2022 Sep 2;13:954567.

[174] Hankinson P, Frezzini C, Khurram SA, Mohammed-Ali R, Smith A. An audit of the management of patients with Oral Epithelial Dysplasia in a Major Teaching Hospital. British Journal of Oral and Maxillofacial Surgery. 2019 Dec 1;57(10):e79-80.

[175] Anderson AC, Yanai I, Yates LR, Wang L, Swarbrick A, Sorger P, et al. Spatial transcriptomics. Cancer Cell. 2022;40(9):895-900.

[176] 11. Marx V. Method of the Year: spatially resolved transcriptomics. Nat Methods. 2021;18(1):9-14.